Dissecting the Aggregate Market Elasticity*

Victor Duarte† Mahyar Kargar‡ Jiacui Li§ Dejanir Silva¶

February 18, 2025

Abstract

We examine the price elasticity of demand for the aggregate stock market in a general equilibrium framework that incorporates rich investor heterogeneity, passive demand, and financial constraints. Using global perturbation techniques, we analytically characterize market elasticity and find that it critically depends on investors' cross-price elasticity—that is, the sensitivity of demand for risky assets to changes in the interest rate. When cross-elasticity is nonzero, the market remains infinitely elastic if passive investors hold the efficient share of risky assets, regardless of how price-inelastic individual investors are. In contrast, portfolio inflows have a positive price impact when risk is misallocated in the economy.

Keywords: Aggregate market elasticity, risk missallocation, demand shocks, demand shifters, excess volatility, asset pricing.

JEL Classification: G12, G11, G21.

---

*We thank Andy Atkeson, Tim Johnson, Ralph Koijen, Wenhao Li, Tyler Muir, Stavros Panageas, Fabrice Tourre, and Pierre-Olivier Weill; and conference and seminar participants at Liquidity in Macroeconomics Workshop at St. Louis Fed, UIUC, UCLA macro-finance lunch, and University of Zurich for valuable comments and suggestions. All errors are our own.

†Gies College of Business, University of Illinois Urbana-Champaign, vduarte@illinois.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.

¶Mitchell E. Daniels, Jr. School of Business, Purdue University, dejanir@purdue.edu.

1 Introduction

Why is the stock market so volatile? This question has long been the subject of extensive research and debate. Existing explanations often invoke the so-called "dark matter of asset pricing," where unobserved factors account for the majority of price fluctuations.[1] More recent demand-based evidence highlights the role of asset price movements in response to portfolio flows as a potential explanation for the origins of the excess market volatility. Given the relatively small magnitude of these flows in the data, as shown in the left panel of Figure 1, this hypothesis requires markets to be highly *inelastic* on average, meaning that even small changes in quantities lead to large price responses (Gabaix and Koijen, 2020). Moreover, as illustrated in the right panel of Figure 1, the volatility multiplier—defined as the ratio of return volatility to flows—is time-varying and countercyclical, spiking during crisis periods.

This new evidence naturally raises several important questions: Why is the aggregate stock market so inelastic? Can standard frictionless asset pricing models generate such low levels of market elasticity? If not, what frictions are necessary to quantitatively explain the impact of portfolio flows on asset prices and match the observed dynamics of the volatility multiplier? Addressing these questions is key to understanding and disentangling the forces driving inelastic markets and, consequently, high market volatility.

This paper develops a general equilibrium model with investor heterogeneity and portfolio frictions to address the questions outlined above. The general equilibrium (GE) perspective is essential, as the key object for studying the price impact of flows is the *macro (aggregate) elasticity*—the change in the aggregate stock market value in response to a \$1 flow from bonds into stocks.[2] When deriving the expression for aggregate elasticity, we account for market-wide adjustments in investors' portfolio holdings across stocks and risk-free bonds, emphasizing the simultaneous response of both the interest rate and the risk premium to flows.[3] Understanding these general

[1]For work on excess volatility tests, see, e.g., LeRoy and Porter (1981), Shiller (1981, 1992), and Cochrane (1992). See Chen, Dou, and Kogan (2024) for the discussion of "dark matter" in asset pricing.

[2]Throughout the paper, we use the terms "aggregate elasticity" and "macro elasticity" interchangeably.

[3]This contrasts with micro elasticity, which focuses on flows between individual stocks, where changes in the money market play a smaller role.

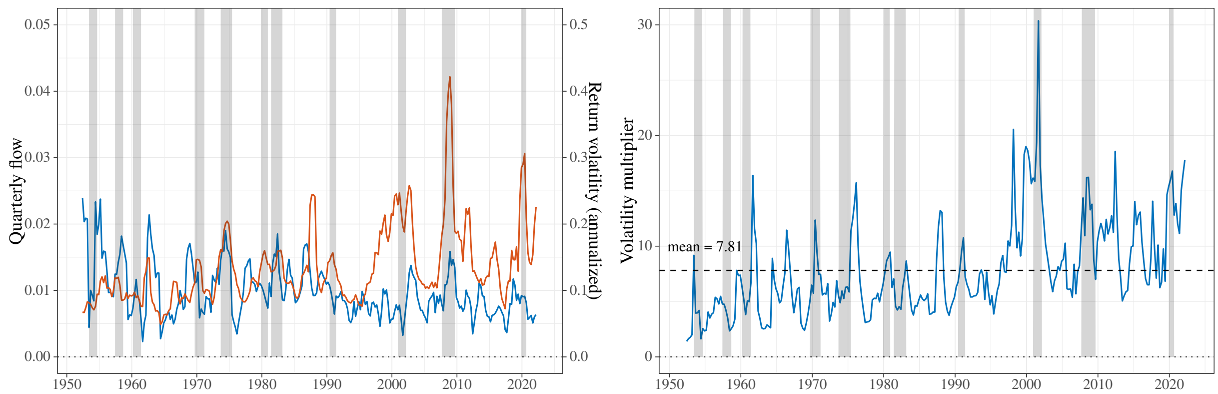

Figure 1. Quarterly flows and return volatility.
The left panel plots quarterly flows (in blue) and return volatility (in red). The right panel plots the "volatility multiplier," defined as the ratio of return volatility to flows. Source: Flow of Funds and CRSP.

equilibrium effects—particularly the impact of flows on the risk-free rate and risk premium—is crucial for characterizing the macro elasticity.

Allowing the interest rate to endogenously respond to portfolio-flow shocks plays a crucial role in determining the aggregate market elasticity.[4] If investors' cross-price elasticity is zero—meaning that demand for the risky asset is independent of the interest rate—then aggregate market elasticity is simply an average of individual investors' price elasticities, consistent with Gabaix and Koijen (2020). In this case, the market remains inelastic as long as investor demand is relatively unresponsive to price changes. Importantly, the behavioral element of their general equilibrium model leads to a constant interest rate. In contrast, we show that if cross-elasticity is nonzero, and interest rates are allowed to move, the market can become infinitely elastic, regardless of how price-inelastic individual investors are.[5]

The disconnect between individual and market elasticities arises from a general equilibrium effect. In partial equilibrium, when passive investors shift funds from the risky asset to the riskless asset, the price of the risky asset tends to decline, increasing the risk premium and incentivizing active investors to raise their exposure. However, in general equilibrium, the interest rate must adjust to clear the bond market in response to the higher demand for the riskless asset. The decline

[4]A portfolio-flow shock corresponds to an asymmetric shock to the demand of risky assets, affecting only a subset of investors and leading to a portfolio reallocation in equilibrium.

[5]Johnson (2006) studies equilibrium price changes in response to shifts in risky asset supply. While his definition of liquidity differs from ours, it also allows for interest rate adjustments when stock prices change.

in interest rates counteracts the effect of the risk premium on the price of the risky asset. We show that if passive investors initially hold an efficient share of the risky asset, the drop in interest rates fully counteracts the change in the risk premium, resulting in no impact on the price of the risky asset following a portfolio flow shock.

This result highlights that individual investors being price-inelastic, as the evidence suggests, is not sufficient to generate inelastic markets. This raises the question of what drives low aggregate market elasticity in general equilibrium. We find that the initial allocation of risk in the economy plays a crucial role. A shock to the portfolio of passive investors effectively redistributes risk between active and passive investors. If risk is initially allocated efficiently, small portfolio adjustments have no impact on aggregate savings behavior, which ultimately determines the price-dividend ratio. In equilibrium, the interest rate fully offsets the change in the risk premium. However, if risk is initially misallocated, portfolio reallocation directly impacts aggregate savings behavior, preventing the interest rate from fully offsetting changes in the risk premium. Through the lens of the model, inelastic markets serve as an indicator of inefficiencies in risk allocation.

We first derive these results within a simple two-period model and then demonstrate that the same intuition extends to a fully dynamic framework. Unlike much of the existing macro-finance literature, we go beyond a two-agent setting and incorporate rich investor heterogeneity. By considering a broader set of investors, we capture the effects of asset reallocations across different sectors, such as mutual funds, households, broker-dealers, and foreign investors.

Several factors can contribute to market inelasticity, including passive demand, institutional investment mandates, and limits to arbitrage.[6] We introduce two key frictions: passive investment and margin constraints. Households are assumed to follow a passive investment strategy rather than actively trading in financial markets, consistent with empirical evidence (e.g., Brunnermeier and Nagel, 2008). The economy also includes active investors who face margin constraints, which are widely recognized in the literature as crucial for market outcomes and the behavior of financial

[6]See Gabaix and Koijen (2020) for more details.

intermediaries.[7]

We show that, in addition to the wealth distribution among agents, market volatility is influenced by both passive flow shocks and aggregate market elasticity. To explain excess market volatility, both factors are necessary. In highly inelastic markets, flow shocks are required to generate additional volatility; without them, inelasticity alone does not increase volatility. Likewise, if markets are infinitely elastic, even large flow shocks do not create excess volatility.

Studying market elasticity in economies with frictions is challenging, as closed-form solutions are typically unavailable. To address this, we apply global perturbation techniques to derive closed-form expressions for market elasticity. Our analysis reveals that aggregate elasticity is both state-dependent and time-varying, influenced not only by the wealth distribution between active and passive investors but also by how wealth is allocated among active agents. In the absence of frictions, the market is infinitely elastic. However, the introduction of passive investors increases the price impact of flows relative to frictionless markets. Preference heterogeneity makes the aggregate market more elastic due to risk misallocation, as more risk-averse passive investors require a higher risk premium to absorb additional risky asset supply. Finally, binding leverage constraints amplify the price impact of flows. Our findings highlight the crucial role of general equilibrium effects and investor heterogeneity in driving market inelasticity.

We then decompose the aggregate elasticity into components driven by the impact of passive flows on the risk premium, the risk-free rate, and the drift of the risky asset price. This decomposition highlights the general equilibrium effects of passive flows, as the risk-free rate and risk premium influence aggregate elasticity in opposing directions.

Lastly, we conduct a quantitative assessment of the model, which involves solving and estimating a high-dimensional asset pricing framework. A key contribution of our work is demonstrating how to computationally manage this complexity. We then evaluate the role of various frictions in explaining excess market volatility.

[7]See, e.g., Brunnermeier and Pedersen (2009), Garleanu and Pedersen (2011), Chabakauri (2013), and Adrian and Shin (2014).

Related literature

Our work relates to the literature on macro demand elasticity that studies the impact of asset flows on the aggregate stock market price. The papers closest to ours are Johnson (2006) and Gabaix and Koijen (2020). In a representative agent model, Johnson (2006) conducts a different exercise than ours by perturbing the risky asset supply and finds finite elasticity even in the frictionless Lucas economy. Gabaix and Koijen (2020) present a model with a behavioral element that leads to constant interest rates to achieve a large price impact in a setting with representative households who can invest in funds subject to mandates. We present a general equilibrium model with rich investor heterogeneity and financial constraints and show that both GE effects and frictions are essential for obtaining inelastic markets.[8]

The macro elasticity literature contrasts with the much larger one on micro elasticity, which examines the change in the relative price of two stocks if one buys \$1 of one and sells \$1 of the other (e.g., Shleifer, 1986; Harris and Gurel, 1986; Chang, Hong, and Liskovich, 2015; Pavlova and Sikorskaya, 2023; Schmickler, 2020). The evidence in the literature suggests that the micro elasticity is much larger than the aggregate elasticity, which is the object of interest in our paper, given that different stocks are closer substitutes than the stock and bond market indices.

Finally, our paper is related to the theoretical microstructure literature (e.g., Kyle, 1985). However, unlike this literature, in our setting, asset flows affect the macro elasticity through their impact on the risk premium and risk-free rate. Campbell and Kyle (1993) look at the interaction of noise traders and smart money, analogous to our discussion of the interaction between passive and active investors. However, they consider a model with CARA preferences and constant interest rates, thus, abstracting from the role of wealth distribution and the GE effects we emphasize.

[8]For more on macro elasticity, see Johnson (2008, 2009), Deuskar and Johnson (2011), Li, Pearson, and Zhang (2020), Hartzmark and Solomon (2021), and other.

2 A simple model of the aggregate market elasticity

In this section, we study the determination of the aggregate market elasticity in a simple general equilibrium demand-based asset pricing model. To keep the discussion as simple as possible, we directly specify investors' demands, without deriving them from utility functions. In Section 3, we provide a micro-founded version of investors' demands in a dynamic heterogeneous-agents model.

2.1 A $2 \times 2 \times 2$ asset pricing model

Consider a two-period economy with two assets, a risky asset and a riskless asset, and two agents, a passive investor (p) and an active investor (a). A fraction ω_j of investors is of type $j \in \{a, p\}$. The risky asset pays a random dividend Y' in the second period. The price of the risky asset is denoted by P, and the price of the riskless asset is denoted by R_f^{-1}.

Investors' budget constraints in the two periods are given by $PQ_j + R_f^{-1} B_j + C_j = W_j$ and $C_j' = Y'Q_j + B_j$, where Q_j denotes the number of shares of the risky asset held by investor j, B_j is the number of riskless bonds, C_j denotes initial consumption, and C_j' denotes consumption in the second period. Initial wealth is given by $W_j = (P + Y)Q_{j,-1} > 0$.

Let $\alpha_j \equiv \frac{PQ_j}{W_j - C_j}$ denote the portfolio share of the risky asset for investor j. For the passive investor, the portfolio share is exogeneously given, that is, $\alpha_p = \overline{\alpha}_p$, where $\overline{\alpha}_p \geq 0$ is a fixed parameter. For the active investor, the portfolio share is a function of the risk premium: $\alpha_a = g_a(\pi)$, where $\pi \equiv \log \frac{1}{R_f} \mathbb{E}[\frac{Y'}{P}]$ represents the log risk premium and $g_a'(\cdot) \geq 0$.

Investor j's consumption is given by $C_j = c_j(r, \pi) W_j$, where $r \equiv \log R$ is the log risk-free rate. The consumption-wealth ratio c_j depends on both the interest rate and risk premium. If $c_j(r, \pi)$ decreased (increased) with r, then the investor saves more (less) when the interest rate is high, reflecting a standard substitution (income) effect. Note that, as an initial benchmark, we assume that $c_j(\cdot, \cdot)$ does not depend directly on passive portfolio share $\overline{\alpha}_p$.

The market clearing conditions are given by

$$\sum_{j\in\{p,a\}} \omega_j C_j = Y, \qquad \sum_{j\in\{p,a\}} \omega_j Q_j = 1, \qquad \sum_{j\in\{p,a\}} \omega_j B_j = 0,$$

and the initial endowment of the risky asset satisfies $\sum_{j\in\{p,a\}} \omega_j Q_{j,-1} = 1$.

Market for risky assets. Let $p = \log \frac{P}{Y}$ denote the log dividend-price ratio and $\mu \equiv \log \frac{\mathbb{E}[Y']}{Y}$ the log dividend growth. From the definition of the risk premium, we obtain the pricing condition

$$\pi = \mu - p - r. \tag{1}$$

Thus, given p and r, we can solve for the risk premium.

The demand for the risky asset for investor j is given by $Q_j = \alpha_j[1 - c_j(r, \pi)]\frac{W_j}{P}$. Using Equation (1), we can express the demand for the risky asset as a function of p, r, and $\overline{\alpha}_p$: $Q_j = F_j(p, r, \overline{\alpha}_p)$. We can write the market clearing condition for the risky asset as follows:

$$\underbrace{\omega_a F_a(p, r, \overline{\alpha}_p)}_{\text{active demand}} = \underbrace{1 - \omega_p F_p(p, r, \overline{\alpha}_p)}_{\text{net supply}}. \tag{2}$$

The market for the risky asset reaches equilibrium when demand from active investors equals the net supply available to them—that is, the total supply minus the amount held by passive investors. Notably, active demand is independent of $\overline{\alpha}_p$, meaning that changes in $\overline{\alpha}_p$ lead to shifts of the net supply curve.

Market for goods. The sensitivity of the consumption-wealth ratio to changes in the interest rate and the risk premium plays an important role in the analysis. In our dynamic model in Section 3, the average consumption-wealth ratio depends on $r + \pi$, meaning that sensitivity to interest rates is equal to sensitivity to the risk premium. Assumption 1 below ensures that the consumption-wealth ratio in our simple model remains consistent with standard micro-founded models.

Assumption 1. *The average consumption-wealth ratio is a function of the expected return in the risky asset $r+\pi$, that is, $\sum_{j\in\{p,a\}} x_j c_j(r,\pi) = c(r+\pi)$, where $x_j \equiv \frac{\omega_j W_j}{\omega_a W_a + \omega_p W_p}$ is the wealth share.*

Intuitively, since bonds are in zero net supply, the average return on investors' portfolios corresponds to the return on the risky asset. The market-clearing condition for goods is given by:

$$c(\mu - p) = \frac{1}{1+e^{p}}. \tag{3}$$

Assumption 1 implies that the system determining p and r satisfies an important *recursivity property*: condition (3) depends only on p, while condition (2) depends on both p and r.

2.2 The general equilibrium implications of portfolio flows

We are interested in examining how the price of the risky asset p and the return on the riskless asset r respond to a portfolio flow toward the risky asset and away from the riskless bond. Let $\overline{\alpha}_p^*$ denote the portfolio share of passive investors in an initial equilibrium, with corresponding prices (p^*, r^*). For a small deviation of the passive portfolio, $\overline{\alpha}_p = \overline{\alpha}_p^* e^{\hat{\alpha}_p}$, we linearized the demand around the initial equilibrium to obtain $q_j \equiv \log Q_j/Q_j^*$:

$$q_j = -\zeta_{j,p}^q (p - p^*) - \zeta_{j,r}^q (r - r^*) + f_j, \tag{4}$$

where $\zeta_{j,p}^q \equiv -\frac{\partial \log F_j}{\partial p}$ represents investor j's price elasticity, and $\zeta_{j,r}^q \equiv -\frac{\partial \log F_j}{\partial r}$ denotes investor j's *cross-elasticity*, which measures the sensitivity of demand for the risky asset to changes in the price of the riskless asset. The term $f_j \equiv \frac{\partial \log F_j}{\partial \log \overline{\alpha}_p} \hat{\alpha}_p$ represents a *flow shock*, capturing an exogenous flow from bonds to stocks. Since active demand is independent of $\overline{\alpha}_p$, we have $f_a = 0$, meaning only passive investors' demand is affected by the flow shock. Moreover, because α_p is independent of the risk premium, the cross-elasticity for the passive investor is zero.

The right panel in Figure 2 illustrates the equilibrium in the market for the risky asset. The solid downward-sloping curve represents active demand. An increase in p, given r, reduces asset demand

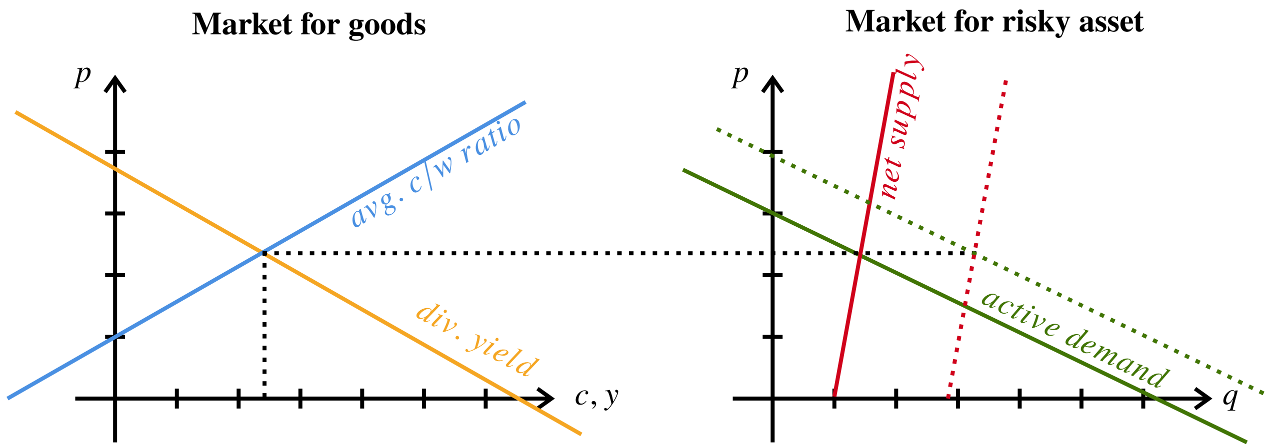

Figure 2. Equilibrium in the goods and risky asset markets

by active investors, as it corresponds to a decline in the risk premium. The net supply of risky assets is depicted as the solid upward-sloping curve. The net supply curve will be upward-sloping provided that $c_p'(r^* + p^*)$ is negative or sufficiently small.[9]

The market-clearing condition for the goods market can be written as

$$\zeta_p^c(p - p^*) = -\frac{e^{p^*}}{(1 + e^{p^*})^2}(p - p^*),$$

where $\zeta_p^c \equiv -c'(\mu - p^*)$. The left-panel of Figure 2 shows the equilibrium in the goods market.

The inelastic markets hypothesis. First, consider a *partial equilibrium* version of this model, where we set $r = r^*$ and drop the market clearing condition for goods. Alternatively, we could take the limit as $x_p \to 1$, in which case the cross-elasticity equals zero, $\zeta_r^q = 0$.[10] Under this setup, the price of the risky asset satisfies:

$$p - p^* = \underbrace{\frac{1}{\zeta_p^q}}_{\substack{\text{inverse market} \\ \text{elasticity}}} \times \underbrace{f}_{\text{flow shock}}, \tag{5}$$

where $\zeta_p^q \equiv \sum_{j \in \{a,p\}} x_j \zeta_{j,p}^q$ and $f \equiv x_p f_p$.

[9]Notice that the net supply would be vertical if we assumed that Q_p was exogenously given instead of α_p.

[10]For this limit to be well-defined, we require $\zeta_p^c + \frac{e^{p^*}}{(1+e^{p^*})^2} = 0$, making the interest indeterminate.

Equation (5) shows that the price-dividend ratio responds to portfolio flows, with the price impact given by the inverse of market elasticity. This elasticity depends on the average price elasticity across all investors. This result is consistent with the *inelastic market hypothesis* of Gabaix and Koijen (2020), who estimate the price impact of flows into the stock market and find a large price response, suggesting that investors are relatively inelastic.

Figure 2 illustrates the intuition behind this result. When passive investors reallocate funds from the risky asset to the riskless asset, the net supply curve effectively shifts to the right, as shown by the upward-sloping dashed line in the right panel. To incentivize active investors to hold more of the risky asset, its expected return must increase—meaning the price-dividend ratio must decline. If both passive and active investors are relatively inelastic, then even a small flow shock can lead to a substantial price drop.

The crucial role of the cross-elasticity. The partial equilibrium analysis abstracts from the role of cross-elasticity. The proposition below shows that the aggregate market elasticity is fundamentally different in the general equilibrium model where r and p are allowed to adjust.

Proposition 1 (Infinite market elasticity)**.** *Suppose* $\zeta_r^q \equiv \sum_{j \in \{a,p\}} x_j \zeta_{j,r}^q \neq 0$*. Then, a portfolio flow shock has no price impact,* $p - p^* = 0$*. Moreover, the portfolio flow shock has opposite effects on the risk premium and the interest rate:* $r - r^* = \frac{1}{\zeta_r^q} f$ *and* $\pi - \pi^* = -\frac{1}{\zeta_r^q} f$*.*

Give a non-zero cross-elasticity, $\zeta_r^q \neq 0$, the market is infinitely elastic, regardless of how price-inelastic individual investors are. Even if investors are extremely price-inelastic ($\zeta_p^q \approx 0$), the market elasticity remains infinite. Moreover, the result holds for any non-zero value of the cross-elasticity. Markets would be infinitely elastic even with a small cross-elasticity ($\zeta_r^q \approx 0$).

The disconnect between individual price elasticity and market elasticity arises from a general equilibrium effect. Figure 2 illustrates this mechanism.

In the goods market, the price-dividend ratio is determined by consumption behavior. This is clearly the case in models with unit elasticity of intertemporal substitution (EIS), where the price-dividend ratio is pinned down by the investor's discount rate. Our formulation generalizes this approach. To restore equilibrium in the market for the risky asset, the interest rate must adjust until the demand from active investors meets the net supply at the initial price. This adjustment is represented in the right panel by the dashed downward-sloping curve, which shifts upward until it intersects the new net supply curve at the original price. Since the interest rate fully offsets movements in the risk premium, the price-dividend ratio remains unaffected by the flow shock.

Intuitively, as funds shift towards riskless bonds, the increased supply of bonds pushes interest rates down. Similarly, the reduction in the demand for risky assets (or the increase in net supply) raises the risk premium. In this economy, these two effects exactly offset each other, regardless of how price-inelastic individual investors are. If the decline in r were smaller than the increase in risk premium, it would create excess demand for goods or, equivalently, an excess supply of bonds. Thus, in this economy, achieving simultaneous equilibrium in the markets for both risky and riskless assets requires an infinite market elasticity.

The need for a micro-founded demand system. The analysis above highlights the importance of the cross-price elasticity. In particular, it provides an example where the market can be infinitely elastic, regardless of how price-inelastic investors are. It is essential for this result that the average consumption-wealth ratio is independent of the portfolio flow shock f. If a portfolio flow shock were to simultaneously shift the net supply curve and the average consumption-wealth ratio, then the price impact would not be zero, and the market elasticity would be finite. [11]

Therefore, to understand the determinants of the aggregate market elasticity, we need a theory of how portfolio reallocation shocks affect not only the demand for risky assets, but also the demand for safe bonds, which is ultimately driven by investors' savings behavior. To obtain a micro-founded demand system, we next consider a dynamic asset pricing model with heterogeneous agents and introduce a new methodology to derive investor demand in this setting.

[11] Similar logic applies to the effects of an increase in uncertainty. Higher uncertainty has no impact on the price when the EIS is equal to one, given a constant consumption-wealth ratio. A high EIS is necessary for higher uncertainty to reduce the price, as it reduces the consumption-wealth ratio in this case (see, e.g., Bansal and Yaron 2004).

3 A dynamic asset pricing model with passive investors

In this section, we consider a general equilibrium continuous-time economy with rich investor heterogeneity. Investors differ in their risk aversion and the extent to which they participate in financial markets, where they can be either active or passive. Active investors are also subject to a state-dependent leverage constraint. This occasionally binding constraint limits the maximum risk exposure of active investors. The presence of passive investors allows us to study the price impact of (passive) portfolio flows from one asset class into another. Investor heterogeneity and financial frictions are essential in assessing (active) market participants' responses to portfolio flow shocks.

3.1 Environment

3.1.1 Endowment and financial markets

The aggregate endowment, Y_t, follows a geometric Brownian motion:

$$\frac{dY_t}{Y_t} = \mu dt + \sigma dZ_t, \tag{6}$$

where $\mu \in \mathbb{R}$ and the $1 \times d$ vector $\sigma \in \mathbb{R}^d$ are constants, and the $d \times 1$ vector $Z = \left\{Z_t \in \mathbb{R}^d; \mathcal{F}_t, t \geq 0\right\}$ is a standard Brownian motion defined in a probability space $(\Omega, P, \mathcal{F})$ equipped with a filtration $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ with the usual conditions. Z is multi-dimensional, enabling us to capture an arbitrary correlation between endowment shocks and portfolio-flow shocks, which we describe below.

Investors can trade a risky asset in unit-supply which is a claim to the aggregate endowment Y_t. The (cumulative) return on the risky asset R_t satisfies:

$$dR_t = \frac{dP_t + Y_t dt}{P_t} \equiv \mu_{R,t} dt + \sigma_{R,t}\, dZ_t, \tag{7}$$

where P_t is the price of the risky claim, $\mu_{R,t}$ is the expected return on the risky asset, and the $1 \times d$ vector $\sigma_{R,t}$ represents its exposure to the different shocks, all to be determined in equilibrium. The instantaneous return volatility is given by $||\sigma_{R,t}||$.

Investors have also access to a instantaneous risk-free asset paying interest rate r_t. In equilibrium, asset prices are a function of the aggregate state variable $X_t \in \mathbb{R}^N$, which allow us to write e.g., $r_t = r(X_t)$, $\mu_{P,t} = \mu_P(X_t)$, and $\sigma_{R,t} = \sigma_R(X_t)$, with some abuse of notation. The aggregate state variable evolves according to

$$dX_t = \mu_{X,t} dt + \sigma_{X,t}\, dZ_t, \tag{8}$$

where $\mu_{X,t} \in \mathbb{R}^N$ is the vector of drifts, and $\sigma_{X,t} \in \mathbb{R}^{N\times d}$ is the matrix of exposures, both determined in equilibrium.

3.1.2 Investors and financial constraints

The economy is populated by $(J+1)$ types of investors, indexed by $j = 0, 1, \ldots, J$. The mass of type j investors is denoted by ω_j. Investors die with Poisson intensity κ and a mass $\kappa\omega_j$ of type j agents are born every period, such that total population is constant and normalized to one. Newborn agents inherit the wealth from their parents. The purpose of this overlapping generation structure is to guarantee that a non-degenerate stationary distribution of wealth exists in our economy. Investors have Duffie and Epstein (1992) recursive preferences, where investor j has risk aversion coefficient γ_j and all investors have a common EIS equal to ψ. Investors choose a process for consumption, $C_{j,t}$, and the share of wealth invested in the risky asset, $\alpha_{j,t}$, subject to portfolio constraints. Given the wealth of investor j, $W_{j,t}$, the demand for the risky asset is given by $Q_{j,t} = \alpha_{j,t} W_{j,t}/P_t$ and the demand for the riskless asset is $B_{j,t} = (1 - \alpha_{j,t})W_{j,t}$.

Passive investors. Investors face different financial constraints. Agents of type $j = 0$ are *passive investors*, that is their portfolio share is given by $\alpha_{0,t} = \overline{\alpha}_{p,t}$, for a given process $\overline{\alpha}_{p,t}$. Outside a boundary region, $\overline{\alpha}_{p,t}$ follows an exogenous square-root process (Cox, Ingersoll, and Ross, 1985):

$$d\overline{\alpha}_{p,t} = \theta_p(\overline{\alpha} - \overline{\alpha}_{p,t})dt + \sigma_p\sqrt{\overline{\alpha}_{p,t}}\, dZ_t, \tag{9}$$

where $\overline{\alpha}$ is the long-run mean and θ_p is the mean-reversion parameter that controls the speed which passive investors rebalance their portfolios. We refer to the innovations to $\overline{\alpha}_{p,t}$ as *portfolio flow shocks*, capturing fluctuations in the amount that passive investors allocate to the risky asset.[12] These movements in portfolio positions may happen in response to aggregate shocks, potentially amplifying the effect of shocks to endowments. Alternatively, they could represent an independent source of fluctuation. Our specification in Equation (9) accommodates both cases. For instance, suppose $d = 2$, $\sigma = (\sigma_1, 0)$, and $\sigma_p = \left(\sigma_{p1}, \sigma_{p2}\right)$. In this case, the first element of the Brownian motion dZ captures aggregate endowment shocks, while the second element of the Brownian motion captures pure portfolio flow shocks.

The formulation above is meant to capture the various forms of passive investor behavior documented in the literature. For instance, Ameriks and Zeldes (2004) and Brunnermeier and Nagel (2008) document substantial inertia in households' portfolios, with very limited or slow rebalancing. This implies that portfolio shares would move with shocks to returns, consistent with the exposure of $\overline{\alpha}_p$ to aggregate shocks in Equation (9). Additionally, Parker, Schoar, and Sun (2023) show that the introduction of target-date funds (TDF) has led to a relatively stable portfolio share. Gabaix and Koijen (2020) document similar behavior for the equity share of different institutional investors, consistent with the idea that these investors do not actively adjust their portfolio share as market conditions vary. Moreover, Parker, Schoar, Cole, and Simester (2022) show that changes in regulation led to an sharp increase in the share of stocks held by households, as TDFs became widely adopted, consistent with the idea that portfolio flows may be driven by factors that are orthogonal to changes in fundamentals.

Behavior at the boundary region. For technical regions, we specify a different behavior for the portfolio share as the wealth share of passive investors approaches one. If passive investors hold nearly all the wealth, but a disproportionately smaller share of the risky assets, the average portfolio share of active investors would increase without bound. This behavior can lead to equilibrium

[12] Alternatively, one could allow for shocks in the mass of passive investors ω_0. Ultimately, it is the total amount of risky assets held by passive investors that will be relevant in equilibrium.

multiplicity, and even failures of no-arbitrage conditions (see, e.g., Hugonnier 2012). To avoid these pathologies, we assume that passive investors increase their portfolio share when their wealth share exceeds a threshold, so the active investors' leverage remains bounded.[13] This ensures that the portfolio share of passive investors converges to one as their wealth also converges to one. In our calibration, the economy spends nearly all its time away from this boundary region.

Active investors. We refer to investors of type $j = 1, \ldots, J$ as *active investors*. Active investors continuously rebalance their portfolios subject to state-dependent leverage constraints. Following the literature on leverage or margin constraints, we assume that the maximum portfolio share for an active investor is decreasing in aggregate volatility:[14]

$$\alpha_{j,t} \leq \frac{\overline{\sigma}}{\|\sigma_{R,t}\|}, \tag{10}$$

where we assume $\overline{\sigma} \geq \|\sigma\|$.

Constraint (10) resembles a Value-at-Risk (VaR) constraint, which is common for banks and other leveraged financial institutions. It captures the fact that either margin requirements or financial intermediaries' regulatory or risk-management constraints become tighter in periods of high volatility. Allowing for leverage constraints plays a potential important role in determining the aggregate market elasticity, as they may limit the ability of some active investors to provide an elastic response to shocks during periods of high volatility.

Investors' problem. The problem of an investor of type $j = 0, \ldots, J$ is given by

$$V_{j,t} = \max_{[C_j, \alpha_j]} \mathbb{E}_t \left[\int_t^\infty f_j(C_{j,s}, V_{j,s}) ds \right], \tag{11}$$

[13]This is analogous to the free-entry mechanism in Khorrami (2022), who shows that these pathologies are avoided in models with entry.

[14]See e.g., Brunnermeier and Pedersen (2009), Garleanu and Pedersen (2011), and Adrian and Shin (2014).

subject to the flow budget constraint,

$$dW_{j,t} = \left[(r_t + \pi_t \alpha_{j,t})W_{j,t} - C_{j,t}\right] dt + \alpha_{j,t} W_{j,t} \sigma_{R,t} dZ_t, \tag{12}$$

a non-negativity condition on wealth, $W_{j,t} \geq 0$, and the portfolio constraint $\alpha_{j,t} \in \Omega_{j,t}$, where $\Omega_{0,t} = \left\{\alpha_0 : \alpha_0 = \overline{\alpha}_{p,t}\right\}$ and $\Omega_{j,t} = \{\alpha_j : \alpha_j \leq \frac{\overline{\sigma}}{\|\sigma_{R,t}\|}\}$ for $j = 1, \ldots, J$, where $\overline{\alpha}_{p,t}$ follows the process (9), and $\pi_t = \mu_{R,t} - r_t$ denotes the risk premium. The aggregator for investor j, $f_j(C, V)$, is given by

$$f_j(C, V) = \rho \frac{(1-\gamma_j)V}{1-\psi^{-1}} \left\{ \left(\frac{C}{\left((1-\gamma_j)V\right)^{\frac{1}{1-\gamma_j}}} \right)^{1-\psi^{-1}} - 1 \right\}, \tag{13}$$

where the discount factor $\rho \equiv \hat{\rho} + \kappa$ incorporates not only the investors' impatience $\hat{\rho}$, but also the death probability κ.

3.1.3 Market clearing and equilibrium

We provide the definition of equilibrium below.

Definition 1. *A competitive equilibrium is a set of stochastic processes adapted to the filtration created by Z_t: the aggregate endowment Y, the price of the claim on the aggregate endowment P, and the risk-free rate r; and a set of stochastic processes for each investor $j \in \{0, \ldots, J\}$: wealth W_j, consumption C_j, and stock holdings α_j, such that*

(i) Aggregate endowment evolves according to (6), given $Y_0 > 0$.

(ii) Given the stochastic processes (P_t, r_t), choices $\left(C_j, \alpha_j\right)$ solve agent j's problem in (11).

(iii) Markets for consumption goods, risky asset, and risk-free bonds clear

$$\sum_{j=0}^{J} \omega_j C_{j,t} = Y_t, \qquad \sum_{j=0}^{J} \omega_j Q_{j,t} = 1, \qquad \sum_{j=0}^{J} \omega_j B_{j,t} = 0, \tag{14}$$

where $Q_{j,t} = \alpha_{j,t} W_{j,t} / P_t$ and $B_{j,t} = (1 - \alpha_{j,t}) W_{j,t}$.

3.2 Equilibrium characterization

In this section, we provide a characterization of the equilibrium conditions and define a Markov equilibrium in terms of the wealth distribution and the portfolio share of passive investors.

Investors' problem. Given the homotheticity of preferences, the value function for investor j can be written as

$$V_{j,t} = \left(\frac{c_{j,t}}{\rho^{\psi}}\right)^{\frac{1-\gamma_j}{1-\psi}} \frac{W_{j,t}^{1-\gamma_j}}{1-\gamma_j},$$

where $c_{j,t}$ is a function of the aggregate state variable X_t, that is, $c_{j,t} = c_j(X_t)$. The function $c_{j,t}$ evolves according to

$$\frac{dc_{j,t}}{c_{j,t}} = \mu_{c_{j,t}} dt + \sigma_{c_{j,t}} dZ_t,$$

where $\mu_{c_{j,t}}$ and $\sigma_{c_{j,t}}$ are given by Ito's lemma. Given the process for $c_{j,t}$, one can solve for investors' policy functions. In particular, the function $c_{j,t}$ corresponds to agent j's consumption-wealth ratio:

$$\frac{C_{j,t}}{W_{j,t}} = c_{j,t}.$$

We can express the portfolio weight $\alpha_{j,t}$ in terms of $c_{j,t}$ and asset prices:

$$\alpha_{j,t} = \min\left\{\frac{\pi_t}{\gamma_j \|\sigma_{R,t}\|^2} + \varsigma_{j,t}, \frac{\overline{\sigma}}{\|\sigma_{R,t}\|}\right\}, \tag{15}$$

where $\varsigma_{j,t} \equiv \frac{1-\gamma_j^{-1}}{\psi-1} \frac{\sigma_{c_{j,t}}\sigma_{R,t}'}{\|\sigma_{R,t}\|^2}$ is the hedging demand component. The risk exposure of unconstrained investors is given by the usual myopic and hedging components. The myopic component depends on the investor's risk tolerance $1/\gamma_j$, the risk premium π_t, and return variance $\|\sigma_{R,t}\|^2$. The hedging demand depends on the correlation of $c_{j,t}$, which affects the investor's marginal utility of wealth, and the risky asset. The leverage constraint limits the maximum risk exposure an investor can achieve at any point in time.

From the Hamilton-Jacobi-Bellman (HJB) equation, we obtain an expression for $c_{j,t}$:

$$c_{j,t} = \psi\rho + (1-\psi)\left[r_t + \pi_t\alpha_{j,t} - \frac{\gamma_j}{2}\|\sigma_{R,t}\|^2\alpha_{j,t}^2\right] + \xi_{j,t}, \tag{16}$$

where $\xi_{j,t} \equiv \mu_{c_j,t} + (1-\gamma_j)\sigma_{c_j,t}\sigma'_{R,t}\alpha_{j,t} + \frac{\psi-\gamma_j}{1-\psi}\frac{\|\sigma_{c_j,t}\|^2}{2}$ corresponds to a forward-looking shifter of the consumption-wealth ratio, which depends on the drift and diffusion of c_j.

The consumption-wealth ratio depends on current investment opportunities, captured by the risk-adjusted return $r_t + \pi_t\alpha_{j,t} - \frac{\gamma_j}{2}\alpha_{j,t}^2\|\sigma_{R,t}\|^2$, as well as future investment opportunities, captured by $\xi_{j,t}$. Note that, as usual, movements in returns have income and substitution effects and the net response of the consumption-wealth ratio depends on the EIS.

Pricing condition. Let $p_t \equiv P_t/y_t$ denote the price-dividend ratio for the risky asset. Since the expected return is given by $r_t + \pi_t = \frac{1}{p_t} + \mu_{P,t}$, the price-dividend ratio satisfies the condition

$$\frac{1}{p_t} = r_t + \pi_t - (\mu + \mu_{p,t} + \sigma\sigma'_{p,t}), \tag{17}$$

where $\sigma_{R,t} = \sigma + \sigma_{p,t}$, $(\mu_{p,t}, \sigma_{p,t})$ is given by Ito's lemma, and we used $\mu_{P,t} = \mu + \mu_{p,t} + \sigma\sigma'_{p,t}$.

Aggregate state variable. Let x_j be the wealth share of type-j investors:

$$x_{j,t} \equiv \frac{\omega_j W_{j,t}}{P_t}.$$

We define the aggregate state variable as $X_t = (x_t, \overline{\alpha}_{p,t})$, where $x_t \equiv \left(x_{1,t}, x_{2,t}, \ldots, x_{J,t}\right)$. The law of motion of $\overline{\alpha}_{p,t}$ is given by (9). From Ito's lemma, the law of motion of $x_{j,t}$ is given by

$$\frac{dx_{j,t}}{x_{j,t}} = \left(r_t + \pi_t\alpha_{j,t} - c_{j,t} - \mu - \mu_{p,t} - \sigma\sigma'_{p,t} + (1-\alpha_{j,t})\|\sigma_{R,t}\|^2 + \kappa\frac{\omega_j - x_{j,t}}{x_{j,t}}\right)dt + \left(\alpha_{j,t} - 1\right)\sigma_{R,t}dZ_t.$$

Risk premium and interest rate. Let $\mathcal{J}_t^u \subseteq \{1, 2, \ldots, J\}$ and $\mathcal{J}_t^c \subseteq \{1, 2, \ldots, J\}$ denote the set of unconstrained and constrained active investors at time t, respectively. Define the wealth share of unconstrained investors as $x_{u,t} \equiv \sum_{j \in \mathcal{J}_t^u} x_{j,t}$, and the wealth share of constrained investors as $x_{c,t} \equiv \sum_{j \in \mathcal{J}_t^c} x_{j,t}$ From the market clearing for the risky asset, we obtain

$$\underbrace{\sum_{j \in \mathcal{J}_t^u} x_{j,t} \alpha_{j,t}}_{\substack{\textbf{active unconstrained} \\ \textbf{demand}}} = \underbrace{1 - x_{0,t} \overline{\alpha}_{p,t} - x_{c,t} \overline{\alpha}_{c,t}}_{\textbf{net supply}}, \tag{18}$$

where $\overline{\alpha}_{c,t} \equiv \frac{\overline{\sigma}}{\|\sigma_{R,t}\|}$ is the portfolio share of constrained agents. The expression above is the analogous in our dynamic setting of Equation (2), as $\omega_j Q_{j,t} = x_{j,t} \alpha_{j,t}$. The net supply of risky assets to unconstrained (or marginal) investors is given by $1 - x_{0,t} \overline{\alpha}_{p,t} - x_{c,t} \overline{\alpha}_{c,t}$, that is, the total supply minus the demand from constrained (or infra-marginal) investors. When all active investors are unconstrained, we recover the condition from Section 2, Equation (2), where the demand from active investors must equal the total supply minus the amount held by passive investors.

Combining the expression above with the optimal portfolio of active investors, we obtain an expression for the risk premium:

$$\pi_t = \frac{\gamma_{u,t} \|\sigma_{R,t}\|^2}{x_{u,t}} \left[1 - x_{0,t} \overline{\alpha}_{p,t} - x_{c,t} \overline{\alpha}_{c,t} - x_{u,t} \varsigma_t\right], \tag{19}$$

where $\gamma_{u,t} \equiv \left[\sum_{j \in \mathcal{J}_t^u} \frac{x_{j,t}}{x_{u,t}} \frac{1}{\gamma_j}\right]^{-1}$ denotes the average risk aversion and $\varsigma_t \equiv \sum_{j \in \mathcal{J}_t^u} \frac{x_{j,t}}{x_{u,t}} \varsigma_{j,t}$ the average hedging demand of unconstrained investors.

Equation (19) captures the effect of heterogeneity, passive investment, and leverage constraints on the risk premium. Keeping everything else constant, an increase in the average risk aversion of unconstrained agents (higher $\gamma_{u,t}$), a reduction in passive investors' demand (lower $\overline{\alpha}_{p,t}$), a reduction in the risk bearing capacity of constrained investor (lower $\overline{\alpha}_{c,t}$), or a reduction in hedging demands (lower ς_t) all tend to increase the risk premium.

We can write the market-clearing condition for goods as follows:

$$\sum_{j=0}^{J} x_j c_{j,t} = \frac{1}{p_t}, \tag{20}$$

From the equation above and Equation (16), we obtain

$$r_t = \rho + \psi^{-1}\mu_{P,t} + \left(1 - \psi^{-1}\right)\sum_{j=0}^{J} x_{j,t}\frac{\gamma_j \alpha_{j,t}^2}{2}||\sigma_{R,t}||^2 - \pi_t + \psi^{-1}\xi_t, \tag{21}$$

where $\xi_t \equiv \sum_{j=0}^{J} x_{j,t}\xi_{j,t}$.

The first two terms of the right-hand side capture the effect of impatience and intertemporal substitution, the next term capture the effect of uncertainty, while the last term captures the effect of time-varying investment opportunities. Note that the risk-free interest rate depends on the risk premium π_t and the distribution of risk across investors $\{\alpha_j\}$, which will be important when considering the equilibrium impact of portfolio flows in Section 4.

Endogenous volatility. The exposure of returns to shocks has both an exogenous component and an endogenous component, $\sigma_{R,t} = \sigma + \sigma_{p,t}$. The exogenous component corresponds to the volatility of cash flows. The endogenous component corresponds to changes in the valuation ratio p_t. As dividend growth is iid, movements in the price-dividend ratio are entirely driven by movements in expected returns. The term $\sigma_{p,t}$ is given by Ito's lemma:

$$\sigma_{p,t} = \frac{p_x(X_t)}{p(X_t)}\sigma_{x,t} + \frac{p_{\overline{\alpha}_p}(X_t)}{p(X_t)}\sigma_p\sqrt{\overline{\alpha}_{p,t}}. \tag{22}$$

The endogenous volatility then depends on two terms. First, the product of the sensitivity of the price-dividend ratio to changes in the wealth distribution (given by $x_{j,t}$ for $j = 1, \ldots, J$) and the response of the wealth distribution to shocks, $\sigma_{x,t}$. Second, the product of the sensitivity of the price-dividend ratio to changes in the passive portfolio and the response of the passive portfolio to shocks. The first term is standard in heterogeneous-agent models (see e.g., Panageas 2020). The second term is only present in economies with exogenous portfolio-flow shocks, and it is only quantitatively relevant in economies where the market is sufficiently inelastic, or $\frac{p_{\overline{\alpha}_p}(X_t)}{p(X_t)}$ is sufficiently large. Therefore, Equation (22) provides the link between the aggregate market elasticity and return volatility $\|\sigma + \sigma_{p,t}\|$.

Markov equilibrium. Equations (19) and (21) allow us to express π_t and r_t as functions of $c_{j,t}$ and p_t and their derivatives, after expressing $(\mu_{c_j,t}, \mu_{p,t})$ and $(\sigma_{c_j,t}, \sigma_{p,t})$ as a function of $(\mu_{X,t}, \sigma_{X,t})$ and the derivatives of $c_{j,t}$ and p_t using Ito's lemma. Plugging the values of π_t and r_t into (16) and (17), and using the expression for the law of motion of the state variables, we obtain a system of $J+2$ PDEs involving $c_j(X_t)$, for $j = 0, 1, \ldots, J$, and $p(X_t)$. These functions depend on $J+1$ state variables, corresponding to J wealth shares, $x_{j,t}$ for $j = 1, 2, \ldots, J$, and the portfolio share of passive investors $\overline{\alpha}_{p,t}$. We define the Markov equilibrium in state variable X_t below.

Definition 2. *A Markov equilibrium in state variable $X_t = \left(x_t, \overline{\alpha}_{p,t}\right)$, where $x_t \equiv \left(x_{1,t}, x_{2,t}, \ldots, x_{J,t}\right)$ and law of motion for $\overline{\alpha}_p$ is given in (9), is the set of functions: price-dividend ratio $p(X)$, interest rate $r(X)$, consumption-wealth ratio $c_j(X)$, policy functions $\{C_j(X), \alpha_j(X)\}$, for $j \in \{0, \ldots, J\}$, and laws of motion for the endogenous state variable $\mu_X(X)$ and $\sigma_X(X)$, such that:*

(i) The consumption-wealth ratio c_j solves agent j's HJB equation (16), and C_j and α_j are the corresponding policy functions, taking p, r and laws of motion for X as given.

(ii) Markets for the consumption good, the risky asset, and the risk-free bond clear:

$$\sum_{j=0}^{J} x_{j,t} c_{j,t} = \frac{1}{p_t}, \qquad \sum_{j=0}^{J} x_{j,t} \alpha_{j,t} = 1, \qquad \sum_{j=0}^{J} x_{j,t}(1 - \alpha_{j,t}) = 0. \tag{23}$$

4 The Determinants of the Aggregate Market Elasticity

In this section, we consider the effects of portfolio flows on asset prices. The response of the price of the risky asset to passive portfolio flows is determined by the (inverse) aggregate market elasticity. Computing this elasticity requires solving the system of PDEs given by (16) and (17), which is not available in closed-form. To isolate the economic mechanisms by which different frictions affect the aggregate market elasticity, we extend the perturbation method used in Silva (2020) to obtain asymptotic closed-form expression for this elasticity.

4.1 Perturbation method

We consider a family of economies indexed by the parameter $\epsilon > 0$. This parameter simultaneously controls the degree of preference heterogeneity, risky asset demand by passive investors, and the tightness of the leverage constraint that active investors face. The coefficient of relative risk aversion of investor $j = 0, \ldots, J$ is given by

$$\gamma_j = \gamma(1 + \hat{\gamma}_j \epsilon), \tag{24}$$

where

$$\sum_{j=0}^{J} \omega_j \hat{\gamma}_j = 0.$$

Parameter γ captures the average risk aversion in the economy weighted by population shares, and $\hat{\gamma}_j$ controls the proportional deviation of investor j's risk aversion from this weighted average. When $\epsilon = 0$, we have an economy with homogeneous preferences, and when $\epsilon = 1$, we have our economy of interest with heterogeneous investors.

Parameter ϵ also affects the portfolio share of passive investors. For simplicity, we abstract from time-variation in the portfolio of passive investors, that is, we assume throughout this section that $\theta_p = 0$ and $\sigma_p = 0$ in Equation (9). We then assume that Z_t is uni-dimensional and write σ instead of $||\sigma||$. The portfolio share of passive investors is given by

$$\overline{\alpha}_p = 1 + \hat{\alpha}_p \epsilon. \tag{25}$$

Note that when $\epsilon = 0$, passive investors are *fully invested* in the risky asset. Parameter $\hat{\alpha}_p$ controls the deviations from this benchmark.

To allow for a first-order role of the leverage constraints, we also assume that $\overline{\sigma}$ in Equation (10) is given by

$$\overline{\sigma} = \sigma + \hat{\sigma} \epsilon. \tag{26}$$

This assumption guarantees that the tightness of the leverage constraint will be of the order $O(\epsilon)$, the same order as the demand for leverage in this economy. Finally, to focus on the role of heterogeneity, passive demand, and leverage constraints, we abstract from the overlapping generations feature, that is, we set $\kappa = 0$.

By considering a family of economies, the values of endogenous variables depend not only on the aggregate state variable X_t but also on the parameter indexing the specific economy, i.e., ϵ. For instance, the price-dividend and the consumption-wealth ratios for investor j are now given by $p(X, \epsilon)$ and $c_j(X, \epsilon)$, respectively.

We are interested in a second-order expansion of the equilibrium objects on the parameter ϵ:

$$p(X, \epsilon) = p_0(X) + p_1(X)\epsilon + p_2(X)\epsilon^2 + O(\epsilon^3), \tag{27}$$

$$c_j(X, \epsilon) = c_{j,0}(X) + c_{j,1}(X)\epsilon + c_{j,2}(X)\epsilon^2 + O(\epsilon^3), \tag{28}$$

where the k-th order corrections $p_k(X)$ and $c_{j,k}(X)$, for $k \in \{0, 1, 2\}$, are functions of the state variable X that we need to determine.

Notice that our method is different from the standard perturbation of dynamic stochastic general equilibrium (DSGE) models in which the function $p(X, \epsilon)$ is typically linearized in both X *and* ϵ. This standard approach makes the analysis local in both ϵ and the distance of X to the non-stochastic steady state. In contrast, we do not assume that the aggregate state is close to the steady state, which requires us to solve for arbitrary functions of X instead of coefficients on a linear or quadratic approximations in DSGE models.[15] Given that our approach provides a global method with respect to the states, we refer to this procedure as *state-global perturbations*.[16]

We proceed by computing the functions $\left(p_k(X), c_{j,k}(X)\right)$, $k = 0, 1, 2$, in three steps. First, we consider the behavior in the benchmark economy, that is, $\epsilon = 0$. We then solve for the first-order

[15]The linearization of p gives $p(X, \epsilon) = \overline{p}_0 + \overline{p}_{1,X}\left(X - \overline{X}\right) + \overline{p}_{1,\epsilon}\epsilon + O\left(\|X - \overline{X}\|^2, \epsilon^2\right)$, where $\overline{X}$ is the non-stochastic steady state, and the coefficients are independent of X. Parameter ϵ typically controls the variance of aggregate shocks in these applications. See Schmitt-Grohé and Uribe (2004) for a discussion of these methods.

[16]See Kargar, Passadore, and Silva (2020) for an application of the state-global perturbation method to an environment with endogenous transaction costs.

and second-order corrections, $p_1(X)$ and $p_2(X)$, respectively.

4.1.1 The benchmark economy

In the absence of preference heterogeneity, and with passive investors fully invested in the risky asset, the economy effectively behaves as the frictionless Lucas economy with a representative agent.

Lemma 1 (Benchmark economy)**.** *Suppose* $\rho > \left(1-\psi^{-1}\right)\left(\mu - \frac{\gamma\sigma^2}{2}\right)$.[17] *Then, for the* $\epsilon = 0$ *economy,*

(i) Investors' consumption-wealth ratio and risk exposure are given by:

$$c_{j,0}(X) = \rho - \left(1-\psi^{-1}\right)\left(\mu - \frac{\gamma\sigma^2}{2}\right), \qquad \alpha_{j,0}(X) = 1, \qquad \textit{for } j = 0, 1, \ldots, J.$$

(ii) Risk premium, risk-free rate, and price-dividend ratio are given by:

$$\pi_0(X) = \gamma\sigma^2, \qquad r_0(X) = \rho + \psi^{-1}\mu - \left(1+\psi^{-1}\right)\frac{\gamma\sigma^2}{2}, \qquad p_0(X) = \frac{1}{c_{j,0}(X)}.$$

Proof. See Appendix C.1. □

Lemma 1 shows that there is no time variation in the expected returns in the benchmark economy in the absence of frictions. In particular, the price-dividend ratio $p_0(X)$ is constant in the benchmark economy, so the risk premium and interest rate are given by the standard Lucas economy formulae.

4.2 The first-order demand system

We consider next the first-order approximation of the demand system. Investor j's demand for the risky asset is given by $Q_j(X;\epsilon) = \alpha_j(X;\epsilon)\frac{x_j}{\omega_j}$. Let $q_{j,t} = \log Q_{j,t}/Q_{j,0}(X_t)$ denote the log

[17] The condition $\rho > \left(1-\psi^{-1}\right)\left(\mu - \frac{\gamma\sigma^2}{2}\right)$ is standard in economies with growth and risk, guaranteeing that investors achieve finite utility.

deviation of investor j's demand from her demand in the benchmark economy. Up to first order, we can write $q_{j,t}$ as follows:

$$q_{j,t} = \alpha_{j,1}(X_t)\epsilon + O(\epsilon^2), \tag{29}$$

where $\alpha_{j,1}(X_t)$ is the first-order correction for portfolio weight $\alpha_j(X;\epsilon)$. Three observations are important in computing $\alpha_{j,1}$. First, the endogenous volatility is equal to zero up to first order, as $\sigma_{p,t}$ is given by

$$\sigma_{p,t} = \frac{p_x(X_t)}{p(X_t)}\sigma_x(X_t) = \left[\frac{p_{x,1}(X_t)}{p_0(X_t)}\sigma_{x,0}(X_t) + \frac{p_{x,0}(X_t)}{p_0(X_t)}\sigma_{x,1}(X_t)\right]\epsilon + O(\epsilon^2) = O(\epsilon^2), \tag{30}$$

where $p_{x,0}(X_t) = \sigma_{x,0}(X_t) = 0$, as shown in Lemma 1. Second, the hedging demand is also second-order in ϵ, $\varsigma_{j,t} = O(\epsilon^2)$, as a similar argument shows that $\sigma_{c_j}(X_t) = O(\epsilon^2)$. Third, expanding the pricing condition (17), we obtain the analogous of Equation (1):

$$\hat{\pi}_t = -\hat{r}_t - \frac{1}{p^*}\hat{p}_t + O(\epsilon^2), \tag{31}$$

where $\hat{\pi}_t \equiv \pi_t - \pi_0(X_t)$, $\hat{r}_t \equiv r_t - r_0(X_t)$, and $\hat{p}_t = \frac{p_t - p^*}{p^*}$, using the fact that $\mu_{p,t} = O(\epsilon^2)$ and $p^* \equiv p_0(X)$. As in Section 2, we can use the pricing condition to write the risk premium in terms of the price-dividend ratio and the interest rate.

Using the three observations above, we can compute the first-order expansion of the portfolio share in Equation (15) to obtain investor j's demand for an unconstrained investor :

$$q_{j,t} = -\zeta^q_{j,p}\hat{p}_t - \zeta^q_{j,r}\hat{r}_t + f_{j,t} + O(\epsilon^2), \tag{32}$$

where

$$\zeta^q_{j,p} \equiv \frac{1}{p^*}\frac{1}{\gamma\sigma^2}, \qquad \zeta^q_{j,r} \equiv \frac{1}{\gamma\sigma^2}, \qquad f_{j,t} \equiv -\hat{\gamma}_j\epsilon.$$

A similar equation holds for a constrained investor, but in this case $\zeta^q_{j,p} = \zeta^q_{j,r} = 0$ and $f_{j,t} = \frac{\hat{\sigma}}{\|\sigma\|}\epsilon$. The demand from passive investors also takes the form above, with $\zeta^q_{0,p} = \zeta^q_{0,r} = 0$, and $f_{0,t} = \hat{\alpha}_p\epsilon$.

The demand for the risky asset for unconstrained agents is analogous to the active demand in Section 2. However, we can now relate the elasticities to the underlying parameters. The price elasticity $\zeta^q_{j,p}$ is given by $\frac{1/p^*}{\gamma\sigma^2}$, the ratio of the dividend yield and the risk premium in the benchmark economy, while the cross price-elasticity $\zeta^q_{j,r}$ is given by the inverse of the risk premium. Importantly, the cross price-elasticity is non-zero for unconstrained investors. As shown above, the direct and cross price elasticities are equal to zero for a constrained investor. Finally, even if $\hat{r}_t = \hat{\pi}_t = 0$, investors' demand would deviate from the one in the benchmark economy due to differences in risk aversion, the leverage limit, or the passive portfolio. These effects are captured by the demand shifters $f_{j,t}$.

We can proceed in an analogous way to compute the first-order expansion of the consumption-wealth ratio in Equation (16):

$$c_{j,t} = \psi\rho + (1-\psi)\left[r_t + \pi_t - \frac{\gamma_j\sigma^2}{2}\right] + O(\epsilon^2), \tag{33}$$

where we used the fact that $\xi_t = O(\epsilon^2)$. In line with Assumption 1, both the risk-free rate and risk premium have a similar first-order impact on the consumption-wealth ratio. Using the pricing condition, we can write $\hat{c}_{j,t} \equiv c_{j,t} - c_{j,0}(X_t)$ in terms of $\hat{p}_t$ and a shifter:

$$\hat{c}_{j,t} = \zeta^c_{j,p}\hat{p}_t + \zeta^c_{j,0}, \tag{34}$$

where $\zeta^c_{j,p} \equiv (\psi-1)\frac{1}{p^*}$ and $\zeta^c_{j,0} \equiv (\psi-1)\frac{\gamma\sigma^2}{2}\hat{\gamma}_j\epsilon$. Notice that the consumption-wealth ratio does not depend on r_t given p_t, as in Section 2. The consumption-wealth ratio can be either increasing or decreasing in p_t. The case depicted in Figure 2, where the consumption-wealth ratio is increasing in p_t, corresponds to $\psi > 1$, a common assumption in macro-finance models (see e.g. Bansal and Yaron 2004). Another important property is that the consumption-wealth ratio is independent of $\overline{\alpha}_{p,t}$, which has implications for the determination of the market elasticity.

4.2.1 Aggregate market elasticity in frictionless economies

Given the demand system, we can solve for the equilibrium prices. The market clearing conditions for the risky asset and for goods can be written as

$$\sum_{j=0}^{J} x_j q_{j,t} = 0, \qquad \sum_{j=0}^{J} x_j \hat{c}_{j,t} = -\frac{1}{p^*}\hat{p}_t,$$

Aggregating the demand for the risky asset and the consumption-wealth ratio, we obtain

$$\begin{bmatrix} \zeta_p^q & \zeta_r^q \\ \zeta_p^c + \frac{1}{p^*} & 0 \end{bmatrix} \begin{bmatrix} \hat{p}_t \\ \hat{r}_t \end{bmatrix} = \begin{bmatrix} f_t \\ -\zeta_0^c \end{bmatrix},$$

where $\zeta_k^q \equiv \sum_{j=0}^{J} x_j \zeta_{j,k}^q$, for $k \in \{p, r\}$, $\zeta_k^c \equiv \sum_{j=0}^{J} x_j \zeta_{j,k}^c$, for $k \in \{p, 0\}$, and $f_t \equiv \sum_{j=0}^{J} x_j f_{j,t}$.

Solving the system above, we obtain asset prices in this economy:

$$\hat{p}_t = -\frac{\zeta_0^c}{\zeta_p^c + \frac{1}{p^*}}, \qquad \hat{r}_t = \frac{f_t}{\zeta_r^q} + \frac{\zeta_p^q}{\zeta_r^q\left(\zeta_p^c + \frac{1}{p^*}\right)}\zeta_0^c.$$

The aggregate market elasticity, $\epsilon_{M,t}$, is defined as the inverse of the proportional change in the price of the risky asset in response to a portfolio flow shock f_t:

$$\epsilon_{M,t}^{-1} = \left(\frac{\partial \hat{p}_t}{\partial f_t}\right)^{-1}.$$

As $\hat{p}_t$ is independent of the portfolio flow shock f_t, the inverse elasticity is equal to zero up to first order:

$$\epsilon_{M,t}^{-1} = 0. \tag{35}$$

This result echoes the findings in Section 2 which showed how the aggregate elasticity could be infinite in the context of a two-period economy. We obtain the same result now in the context of our dynamic economy using a first-order approximation.

Interpretation: market elasticity in a frictionless economy. Why is the market infinitely elastic? The fact that we are considering a first-order approximation is relevant. The elasticity we obtain coincides with the one in an economy without preference heterogeneity or leverage constraints, and passive investors are initially fully invested in the risky asset. Therefore, we can interpret $\epsilon_{M,t}$ as capturing the effect of a small deviation from a Lucas economy. In this effectively frictionless environment, portfolio flows have opposing effects on interest rates and risk premium, leaving the price of the risky asset unchanged, as shown in the next proposition.

Proposition 2 (First-order impact of portfolio flows)**.** *Suppose* $\rho > \left(1-\psi^{-1}\right)\left(\mu - \frac{\gamma\sigma^2}{2}\right)$. *Then,*

(i) The price-dividend ratio is given by

$$\hat{p}(X) = -(1-\psi^{-1})p^*\frac{\gamma\sigma^2}{2}\sum_{j=0}^{J} x_j\hat{\gamma}_j\epsilon + O(\epsilon^2).$$

(ii) The interest rate and the risk premium are given by

$$\hat{\pi}(X) = \gamma\sigma^2\left[\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j - \frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c}\right]\epsilon + O(\epsilon^2),$$
$$\hat{r}(X) = \gamma\sigma^2\left[-\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j + \frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c} + \left(1-\psi^{-1}\right)\sum_{j=0}^{J}x_{j,t}\frac{\hat{\gamma}_j}{2}\right]\epsilon + O(\epsilon^2),$$

where $x_{u,t} \equiv \sum_{j\in\mathcal{J}_t^u} x_{j,t}$ *and* $x_{c,t} \equiv \sum_{j\in\mathcal{J}_t^c} x_{j,t}$ *are the wealth shares of unconstrained and constrained active investors, respectively, and* $\mathcal{J}^u$ *and* $\mathcal{J}^c$ *denote, respectively, the set of unconstrained and constrained active investors, defined in Section 3.2.*

Proposition 2 shows that if passive investors reduce their position in the risky asset, this raises the risk premium and reduces the interest rate by the same amount. The intuition is analogous to the case shown in Figure 2. The risk premium increases to induce active investors to hold more of the risky asset, while the interest rate adjusts to ensure the price-dividend ratio is consistent with equilibrium in the goods market.

4.3 The second-order demand system

Proposition 2 shows that the market is infinitely elastic up to first order. However, a first-order approximation is unable to capture how different frictions interact. For this reason, we consider next a second-order approximation of the demand system. To isolate the role of each friction, we consider first the case without preference heterogeneity or leverage constraints, we then add preference heterogeneity, and finally consider the case with heterogeneity and leverage constraints.

4.3.1 Inefficient passive demand

Suppose the risk aversion is the same for all investors and they are not subject to a leverage constraint. We focus on how the initial portfolio of passive investors, $\overline{\alpha}_p$, affects the aggregate market elasticity. In particular, whether the portfolio of passive investors deviates from its optimal level plays an important role.

The role of risk misallocation. In the absence of heterogeneity and leverage constraints, the demand for the risky asset under a second-order approximation takes the form in Equation (32), as shown in the appendix. In contrast, the consumption-wealth ratio is now given by

$$c_{j,t} = \psi\rho + (1-\psi)\left[r_t + \pi_t\alpha_{j,t} - \frac{\gamma\sigma^2}{2}\alpha_{j,t}^2\right] + O(\epsilon^3), \tag{36}$$

where $\alpha_{0,t} = \overline{\alpha}_p$ and $\alpha_{j,t} = \frac{1-x_{0,t}\overline{\alpha}_p}{1-x_{0,t}}$ for $j = 1, \ldots, J$. Aggregating across investors, and using the market clearing condition $\sum_{j=0}^{J} x_j\alpha_{j,t} = 1$, we obtain

$$c_t = \psi\rho + (1-\psi)\left[r_t + \pi_t - \frac{\gamma\sigma^2}{2}\sum_{j=0}^{J} x_j\alpha_{j,t}^2\right] + O(\epsilon^3). \tag{37}$$

The average consumption-wealth ratio now depends on the distribution of the risky asset in the economy. Under the optimal allocation, the portfolio share is equal to one, $\alpha_{j,t} = 1$, for all investors, so $\sum_{j=0}^{J} x_j\alpha_{j,t}^2 = 1$. Any deviation of the optimal allocation creates dispersion in portfolios, so

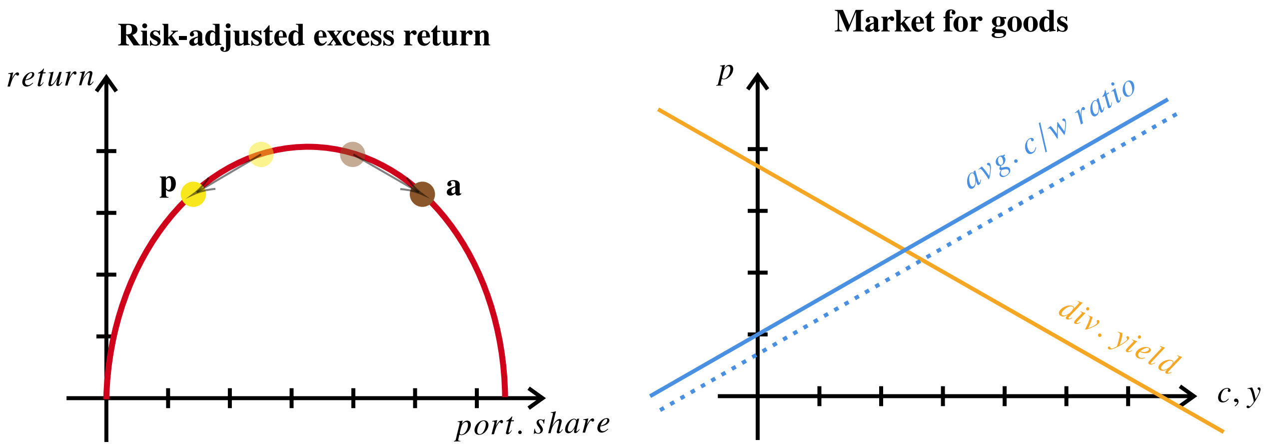

Figure 3. Consumption-wealth ratio when risk is nearly perfectly allocated

$\sum_{j=0}^{J} x_j \alpha_{j,t}^2 > 1$, which reduces the risk-adjusted expected return $r_t + \pi_t - \frac{\gamma\sigma^2}{2} \sum_{j=0}^{J} x_j \alpha_{j,t}^2$. If $\psi > 1$, the reduction in risk-adjusted returns weakens investors incentive to save, for any given level of $r_t + \pi_t$. Therefore, the consumption-wealth ratio depends on how risk is allocated in the economy.

Figure 3 illustrates how an increase in risk misallocation affects the equilibrium in the goods market. The left panel represents the risk-adjusted excess return, $\pi_t \alpha_{j,t} - \frac{\gamma\sigma^2}{2} \alpha_{j,t}^2$, as a function of the portfolio share $\alpha_{j,t}$. Suppose that the passive investor has initially a portfolio share that is less than one and then decides to further reduce its holding of the risky asset, as represented by the yellow dot in the figure. Even if the risk premium is constant, the risk-adjusted return is reduced, as the investor moves away from the optimal holding of the risky asset. The portfolio share of the active investors also moves away from the optimal, as represented by the brown dot in the figure, so the risk-adjusted return is reduced for them as well. Therefore, both agents have a weaker incentive to save, which leads to a shift in the average consumption-wealth ratio, as shown in the right panel.

The magnitude of the shift in the average consumption-wealth ratio depends on the initial distribution of risk. This point can be seen more clearly by writing the average consumption-wealth in terms of deviations from its value in the benchmark economy:

$$\hat{c}_t = \zeta_p^c \hat{p}_t + \zeta_0^c, \tag{38}$$

where $\zeta_0^c = (\psi - 1) \frac{\gamma\sigma^2}{2} \sum_{j=0}^{J} x_j (\alpha_{j,t} - 1)^2$. Notice that ζ_0^c is a function of $\overline{\alpha}_p$ and the derivative of

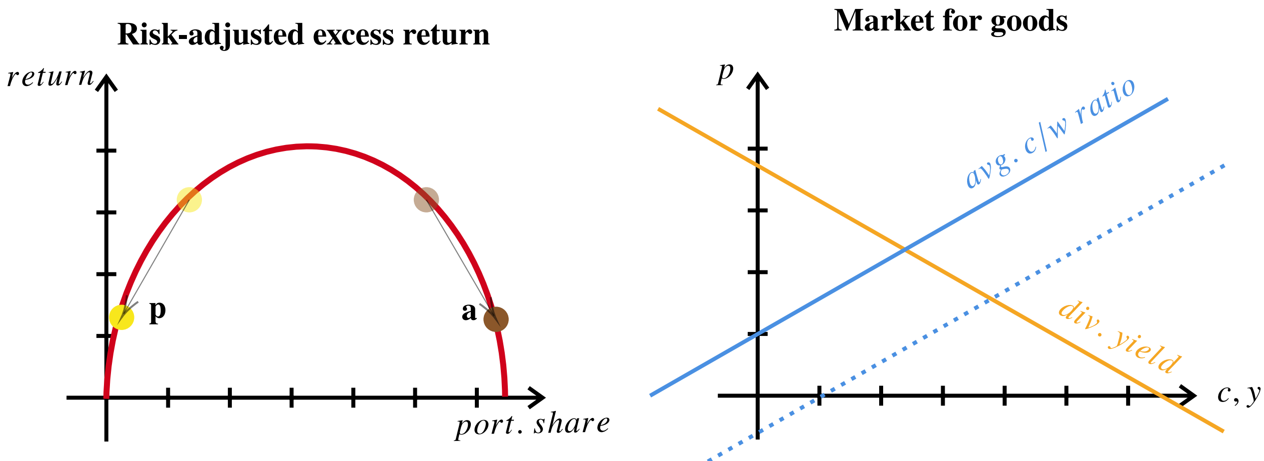

Figure 4. Consumption-wealth ratio when risk is highly misallocated

ζ_0^c with respect to $\overline{\alpha}_p$ is given by

$$\left.\frac{\partial \zeta_0^c}{\partial \overline{\alpha}_p}\right|_{\overline{\alpha}_p=1} = (\psi - 1)\gamma\sigma^2 \sum_{j=0}^{J} x_j(\alpha_{j,t} - 1) \left.\frac{\partial \alpha_{j,t}}{\partial \overline{\alpha}_p}\right|_{\overline{\alpha}_p=1} = 0. \tag{39}$$

If the passive investor initially holds the optimal portfolio, then a small change in $\overline{\alpha}_p$ has no impact in the consumption-wealth ratio. This corresponds to the frictionless benchmark, which is analogous to the result under the first-order approximation. However, if the initial allocation of risk is not optimal, changes in $\overline{\alpha}_p$ affect the consumption-wealth ratio. Moreover, the effect is stronger the further the initial allocation is from the optimal. This point is illustrated in Figure 4. The initial allocation is now further away from the optimal relative to the case in Figure 3, which leads to a larger decline in the risk-adjusted returns, as shown in the left panel. In this case, we see a larger shift in the average consumption-wealth curve in the right panel.

The market elasticity with inefficient passive demand. The equilibrium value for asset prices takes the same shape as in the case of the first-order approximation:

$$\hat{p}_t = -\frac{\zeta_0^c}{\zeta_p^c + \frac{1}{p^*}}, \qquad \hat{r}_t = \frac{f_t}{\zeta_r^q} + \frac{\zeta_p^q}{\zeta_r^q\left(\zeta_p^c + \frac{1}{p^*}\right)}\zeta_0^c.$$

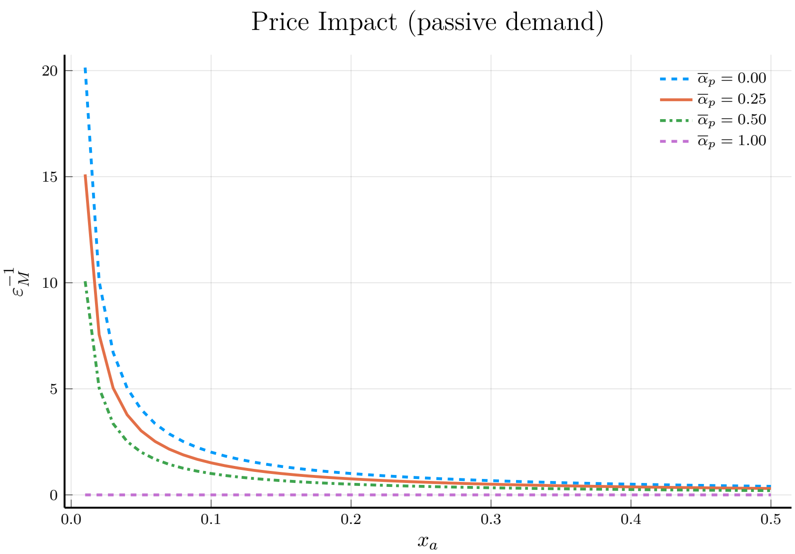

Figure 5. Price impact: inefficient passive demand.

As before, the price of the risky asset is independent of the (micro) price elasticity ζ_p^q. However, while the shifter ζ_0^c was independent of $\overline{\alpha}_p$ under a first-order approximation, this is not the case now, as ζ_0^c is a function of $\overline{\alpha}_p$. The next proposition derives the aggregate market elasticity in this case.

Proposition 3 (Aggregate elasticity: inefficient passive demand)**.** *Suppose* $\rho > \left(1-\psi^{-1}\right)\left(\mu - \frac{\gamma\sigma^2}{2}\right)$. *If there is no preference heterogeneity and active investors do not face leverage constraints, the inverse aggregate market elasticity,* $1/\varepsilon_M$, *is given by:*

$$\varepsilon_M^{-1} = \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{y_0(X)}\frac{1-\overline{\alpha}_p}{x_a} + O\left(\epsilon^2\right), \tag{40}$$

where $x_a \equiv 1 - x_0$ *denotes the wealth share of active investors and* $y_0(x) = 1/p_0(X)$.

Proof. See Appendix D.1. □

We highlight several points from Proposition 3. First, the price impact is equal to zero if $\overline{\alpha}_p$, that is, the market is infinitely elastic when risk is initially optimally allocated. Second, the price impact is also zero if the EIS is equal to one. In this case, the consumption-wealth ratio is constant and $\zeta_0^c = 0$, so $\overline{\alpha}_p$ does not affect investors' incentive to save. Third, the market elasticity is positive if $\psi > 1$ and $\overline{\alpha}_p < 1$, so passive investors hold an inefficiently low share of the risky asset. This result is consistent with the discussion in Section 2, which shows that we obtain a finite elasticity when movements in the passive portfolio share shifts not only the net supply of risk to active investors, but also the consumption-wealth ratio. As discussed above, risk misallocation provides a mechanism linking changes in $\overline{\alpha}_p$ to shifts in the consumption-wealth ratio. Moreover, everything else constant, the price impact is largest in the case of no market participation by passive investors ($\overline{\alpha}_{p,t} = 0$), as in e.g. Basak and Cuoco (1998), given that this corresponds to the maximum level of misallocation (given $\overline{\alpha}_p \geq 0$). Fourth, the elasticity is state-dependent and the price impact is larger when active investors are under-capitalized, that is, they hold a relatively small share of wealth. Figure 5 illustrate these points. We can see that the price impact is equal to zero when $\overline{\alpha}_p = 1$, it is positive when $\overline{\alpha}_p < 1$, it is decreasing in wealth share of active investors x_a, and it is largest when $\overline{\alpha}_p = 0$ for any given level of x_a.

4.4 Preference heterogeneity

We consider next the case where investors have heterogeneous preferences, but they are no subject to leverage constraints. Proposition 4 derives the aggregate market elasticity for this case.

Proposition 4 (Aggregate elasticity: preference heterogeneity)**.** *Suppose $\rho > \left(1 - \psi^{-1}\right)\left(\mu - \frac{\gamma\sigma^2}{2}\right)$. If active investors have heterogeneous risk aversions but face no leverage constraints, the inverse aggregate market elasticity, $1/\varepsilon_M$, is given by:*

$$\varepsilon_M^{-1} = \left(1 - \psi^{-1}\right) \frac{\gamma\sigma^2}{y_0(X)} \left[\frac{1 - \overline{\alpha}_p}{x_a} - \frac{\gamma_0 - \mathbb{E}^u[\gamma_j]}{\gamma}\right] + O\left(\epsilon^2\right), \tag{41}$$

where $\mathbb{E}^u[\gamma_j] \equiv \sum_{j \in \mathcal{J}^u} \frac{x_j}{x_u}\gamma_j$.

Proof. See Appendix D.2. □

Several points are worth emphasizing. First, note that with heterogeneous active investors even in the case in which $\overline{\alpha}_p = 1$, the aggregate elasticity is finite. As we show later in the next section, this effect is due to impact of flows on the risk-free rate.

Second, assuming that the passive agents are more risk-averse than active investors, that is $\gamma_0 \geq \mathbb{E}^u[\gamma_j]$, with $\psi > 1$, introducing preference heterogeneity attenuates the price response to flows, leading to more elastic markets relative to the economy without heterogeneity in Proposition 3. This result is due to risk misallocation, as passive investors are effectively closer to their optimal portfolio when $\gamma_0 > \mathbb{E}^u[\gamma_j]$, for a given initial value $\overline{\alpha}_p < 1$. To see this point, let's compute the optimal portfolio share that investor $j = 0$ would choose if she was an active investor:

$$\overline{\alpha}_p^{optimal} = 1 - x_a \left[\frac{\gamma_0 - \mathbb{E}^a[\gamma_j]}{\gamma} \right], \tag{42}$$

If type $j = 0$ investors have high risk aversion, then it is optimal for them to invest less than 100% in the risky asset. In this case, $\overline{\alpha}_p < 1$ is not an indication of risk misallocation, but it reflects the optimal risk sharing among investors. It turns out that the elasticity will be positive if $\alpha_p < \overline{\alpha}_p^{optimal}$:

$$\varepsilon_M^{-1} = \left(1 - \psi^{-1}\right) \frac{\gamma\sigma^2}{y_0(X)} \frac{\overline{\alpha}_{p,t}^{optimal} - \overline{\alpha}_p}{x_a} + O\left(\epsilon^2\right). \tag{43}$$

This shows that risk misallocation is again the key ingredient necessary to obtain a finite elasticity. Notice that a reduction in the wealth share of active investors raises the price impact by more in the presence of heterogeneous preferences, everything else constant. The reason is that a reduction in x_a now raises the optimal portfolio share of passive investors. Optimal risk sharing implies they should hold more of the risky asset when x_a is low, which is typically after a negative aggregate shock. For a given level of $\overline{\alpha}_p$, this implies that there is more risk misallocation when x_a is low, that is, $\overline{\alpha}_p$ is further away from the optimal.

Finally, from Equation (41), we see that the price impact of flows depends not only on the wealth share of active investors x_a as before, but also on the wealth distribution among active investors through $\mathbb{E}^u[\gamma_j]$. This results reiterates the importance of heterogeneity among financial intermediaries that active investors in our model represent (e.g., Veronesi, 2019; Kargar, 2021).

In Figure 6, we plot inverse aggregate elasticity for the case with one passive and two active investors ($J = 3$) with $\gamma_0 \geq \gamma_1 > \gamma_2$. In this economy, the two key state variables are the wealth

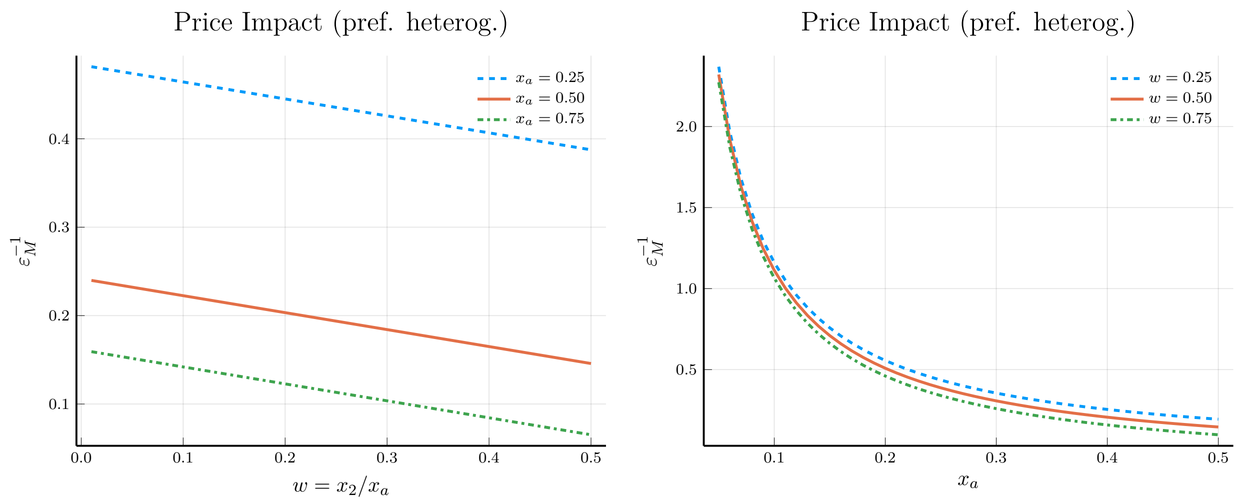

Figure 6. Price impact: passive demand and preference heterogeneity.

share of active investors, x_a, and the wealth share the most risk-tolerant active investor (agent 2) among active investors, w:

$$x_a = 1 - x_0, \qquad w = \frac{x_2}{x_a}. \tag{44}$$

From Figure 6, we see that as the active investors become less capitalized and x_a declines, market become more inelastic. Moreover, as less risk averse active investors become less capitalized and w goes down, we see a larger price impact from passive flows.

4.5 Leverage constraints

We consider next the case where investors have heterogeneous preferences and they are subject to leverage constraints. In Proposition 5, we derive the aggregate market elasticity when heterogeneous active investors also face leverage constraints.

Proposition 5 (Aggregate elasticity: leverage constraints)**.** *Suppose* $\rho > \left(1 - \psi^{-1}\right)\left(\mu - \frac{\gamma\sigma^2}{2}\right)$. *When active investors have heterogeneous preferences and also face leverage constraints in* (10), *the inverse aggregate market elasticity,* $1/\varepsilon_M$, *is given by:*

$$\varepsilon_M^{-1} = \left(1 - \psi^{-1}\right) \frac{\gamma\sigma^2}{y_0(X)} \left[\frac{(1 - x_c)(1 - \overline{\alpha}_p)}{x_u} - \frac{\gamma_0 - \mathbb{E}^u[\gamma_j]}{\gamma} - \frac{x_c(\frac{\overline{\sigma}}{\sigma} - 1)}{x_u} \right] + O\left(\epsilon^2\right). \tag{45}$$

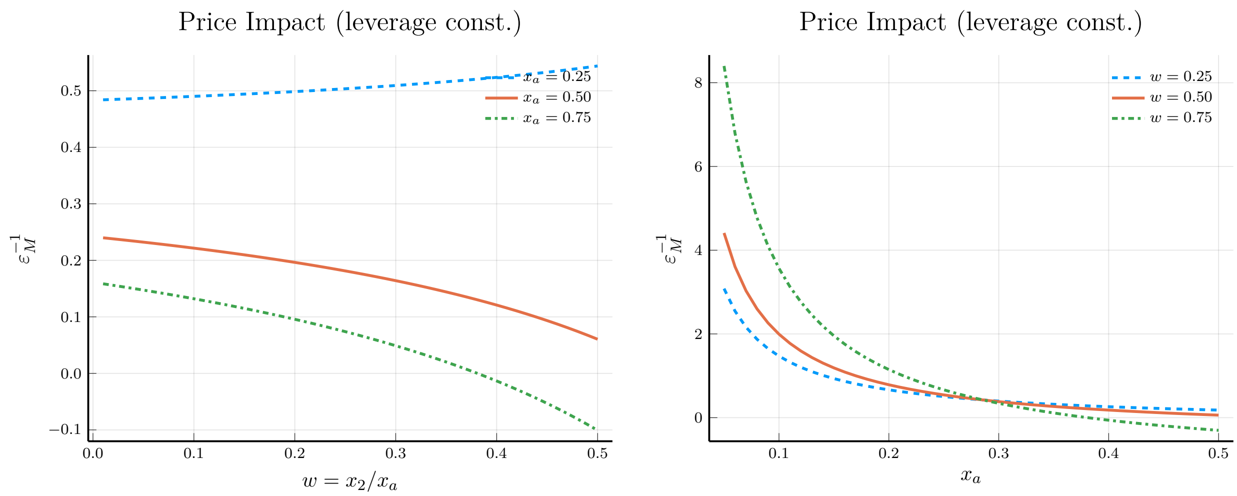

Figure 7. Price impact: leverage constraints.

Proof. See Appendix D.3. □

We highlight several points from Proposition 5. First, note than when all active agents are unconstrained, i.e., $x_c = 0$, we get the price impact from Proposition 4. Similar to the case with heterogeneous investors, the aggregate elasticity is finite when $\overline{\alpha}_p = 1$.

As in the previous two cases, risk misallocation plays an important role. The optimal portfolio share of type $j = 0$ investors, i.e. the portfolio they would choose if they were active, is given by

$$\overline{\alpha}_{p,t}^{optimal} = 1 - \frac{x_u}{1 - x_c} \frac{\gamma_0 - \mathbb{E}^u[\gamma_j]}{\gamma} - \frac{x_c\left(\frac{\overline{\sigma}}{\|\sigma\|} - 1\right)}{1 - x_c}. \tag{46}$$

Plugging the formula for the optimal portfolio share in the expression for the aggregate market elasticity, we obtain

$$\varepsilon_M^{-1} = \left(1 - \psi^{-1}\right) \frac{\gamma\sigma^2}{y_0(X)} \frac{\overline{\alpha}_{p,t}^{optimal} - \overline{\alpha}_p}{x_a}(1 - x_c) + O\left(\epsilon^2\right). \tag{47}$$

As before, the market elasticity is finite and positive when $\overline{\alpha}_p$ is below its optimal level. The optimal portfolio now depends on the average risk aversion of unconstrained agents, the leverage constraint, and the entire distribution among active investors. Figure 8 illustrates the behavior of the market elasticity in the presence of leverage constraints. The left panel shows the price impact

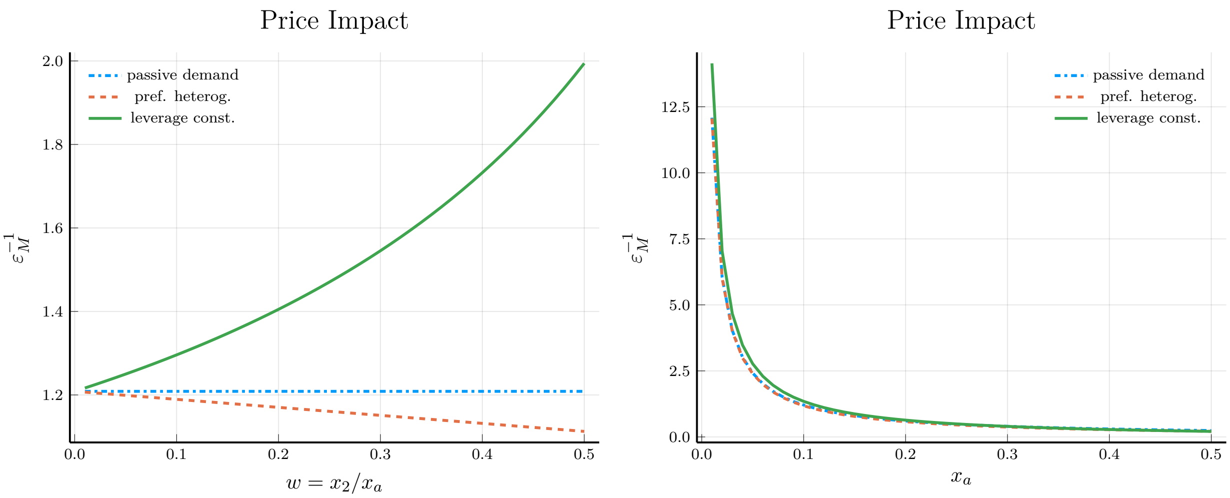

Figure 8. Decomposition of the inverse aggregate market elasticity.

for different values of $w = x_2/x_a$, that is, the share of wealth of the low risk aversion investor among active investors. The right panel shows the inverse elasticity as a function of the wealth share of active investors.

5 Quantitative Implications

In this section, we consider the quantitative implications of the model described in Section 3.

5.1 Calibration strategy

Sectors. We process raw flow of funds (FOF) data to arrive at $J = 6$ major sectors: household mutual fund holdings, households non-mutual fund holdings, hedge funds, broker-dealers (L130), rest of the world (L133), and hedge funds.

To visualize the size of these ultimate sectors, Figure 9 plots the fraction of the U.S. stock market held by each sector from 2012 when hedge fund holdings became available. Figure 9's legend reports the average fraction held by each sector. The largest holder is the household sector which holds 17.3% via mutual funds and 43.3% through other means. The net largest sectors are foreign investors ("rest of the world") and pension/insurance sectors which hold 17.5% and 17.3%, respectively. Hedge funds and broker-dealers are relatively small and hold 5.8% and 0.5%,

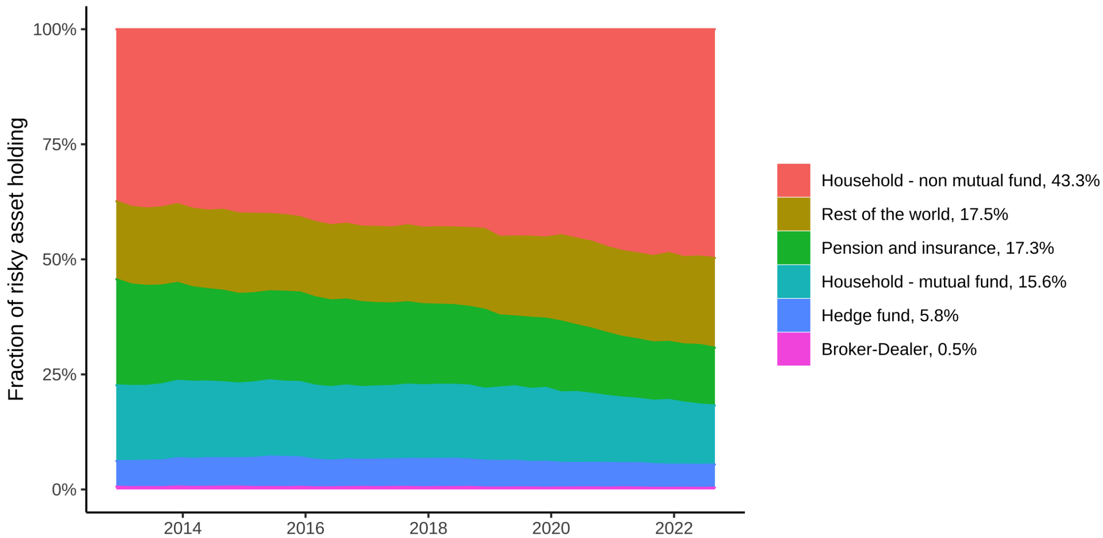

Figure 9. Fraction of stock market holding by sectors from 2012 to 2022. The average fraction held is reported in the legend. Source: Flow of Funds.

respectively, consistent with Figure 1 in Koijen, Richmond, and Yogo (2024).

We first drop sectors that have no or minimal investments in the stock market. This includes nonfinancial business (L102), general government (L105), monetary authority (L109), Private depository institutions (L110), money market funds (L121), Government-sponsored enterprises (L125), Agency-and GSE-backed mortgage pools (L126), and Issuers of asset-backed securities (L127), Finance companies (L128), Real estate investment trusts (L129), Holding companies (L131), and Other financial business (L132). We then aggregate sectors that invest in the stock market but are pass-throughs onto the end users' balance sheets.

Mutual funds (L122), Closed-end funds (L123), and Exchange-traded funds (L124): we aggregate L123 and L124 onto the household (L101) balance sheet. For mutual fund holdings (L122), FOF provides a detailed breakdown of end investors. We aggregate the non-household holdings onto the respective end investor balance sheets but separately consider the household holdings as a separate sector. We aggregate all defined contribution pensions (Tables L118c, L119c, and L120c) onto household balance sheet.

We then separate out the hedge fund sector. Both the household (L101) and foreign (L133) sectors contain hedge funds. Because hedge funds likely behave differently from other sectors, we separate them out. We use Table B.101.f to capture domestic hedge funds and use the hedge fund table in enhanced financial accounts to also capture foreign hedge funds. We consider all hedge funds a single sector, and we subtract their holdings from the household and foreign sectors. Finally, we aggregate together all remaining insurance and pension sectors. This includes property-casualty insurance (L115), life insurance (L116), and defined benefit pension and retirement plans (L118b, L119b, L120b).

Mass of investors and risk aversion coefficients. We choose ω_j and γ_j for each sector j to match the following moments:

- $\frac{1}{T}\sum_t \frac{Assets_{j,t}}{\sum_i Assets_{i,t}}$, which corresponds to $\frac{1}{T}\frac{\sum_t \omega_j W_{j,t}}{\sum_t \sum_i \omega_i W_{i,t}}$ in the model
- $\{b_j\}_j$ from time-series regressions of

  $$\text{Risky Asset}_{j,t} = a_j + b_j \left(\sum_j \text{Risky Asset}_{i,t} \right),$$

  which corresponds in the model to the regression coefficient in the following regression in the model:

  $$\omega_j W_{j,t} \alpha_{j,t} = a_j + b_j \sum_i \omega_i W_{i,t} \alpha_{i,t}$$

- Only for active sectors $(j = 1, 2, \ldots, J)$

Average passive Share $\overline{\alpha}$. There seems to be a big range in the literature from a third to over two thirds:

- Chinco and Sammon (2024) find passive share of 33%: Each time that a stock gets added to or dropped from an index they ask: "How much money would have to be tracking that index to explain the enormous burst in closing volume on reconstitution day that we observe in the data?"

- Koijen et al. (2024) find a much larger passive share of 67.2% in 2016Q4. They use the active share, modified for their application, as a measure of active investment management (Cremers and Petajisto, 2009).

We need to find the total passive wealth and divide to get portfolio share.

Calibrating the Passive Demand Process. Type $j = 0$ investors are passive, and their portfolio weight follows and exogenous CIR process in Equation (9), reproduced below:

$$d\overline{\alpha}_{p,t} = \theta_p(\overline{\alpha} - \overline{\alpha}_{p,t})dt + \sigma_p\sqrt{\overline{\alpha}_{p,t}}\, dZ_t,$$

1. Average passive demand ($\overline{\alpha}$):
   - Calibrating the average risky asset share of the passive sector is discussed above.
2. Volatility of the passive demand (σ_p)
   - Following Appendix D3 of Gabaix and Koijen (2020), to measure equity flows, we scale the dollar equity flows for each sector j, $\Delta F^{\varepsilon}_{j,t}$, by the size of the aggregate market in the previous quarter, ε_{t-1} i.e., $\Delta F^{\varepsilon}_{j,t}/\varepsilon_{t-1}$.
   - We use indirect inference by matching total flows (the blue line in the left panel of Figure 1). This is our attempt to replicate the red dashed line in Figure D.7 in Gabaix and Koijen (2020).
3. Persistence of passive demand (θ_p)
   - We follow the literature on household portfolio inertia. Brunnermeier and Nagel (2008), using PSID data, document substantial inertia in households' portfolios, with very limited or slow rebalancing. This implies that portfolio shares would move with shocks to returns, consistent with the exposure of $\overline{\alpha}_p$ to aggregate shocks.

- Brunnermeier and Nagel (2008) "use the information on net purchases or sales of risky assets to construct a variable $\Delta_k \text{Inert}_t$, representing the (counterfactual) change in the liquid risky asset share that the household would have experienced between $t-k$ and t under perfect inertia—that is, if it had not undertaken any purchases or sales of risky assets between $t-k$ and t." They regress household portfolio weight on inertia ($\Delta_k \text{Inert}_t$), and k-period difference in post consumption wealth and major changes in family composition or asset ownership. They find the coefficient on the inertia variable is large, around 0.75, with small standard errors. Taken at face value, it suggests that there is huge inertia. Households' asset allocations seem to fluctuate strongly as a function of in- and outflows, and capital gains and losses, without much rebalancing taking place.
- Similarly, Parker et al. (2023) show that the introduction of target-date funds has led to a relatively stable portfolio share, consistent with the evidence in Brunnermeier and Nagel (2008).

4. Correlation between flow shocks, i.e., innovations in $\overline{\alpha}_p$, and the aggregate shock
    - We regress quarterly flows, the blue line in the left panel of Figure 1, on the market. The beta coefficient can be used to indirectly get the correlation between flow shocks and aggregate socks.

Table 1 lists the parameter values used in calibrating the model.

5.2 Quantitative results

Numerical solution. To assess the model's quantitative implications, we rely on a global solution method instead of the perturbation approach described in Section 4. A method able to handle dimensional state spaces is necessary, as we have a total of $J+1$ state variables, and a total of $J=6$ active sectors. We adopt the neural-networks based method of Duarte, Duarte, and Silva (2024). We discuss the solution method in more detail in the appendix.

Table 1. Parameter values

This table reports the parameter values used in calibrating the model.

	Parameter	Choice
	Preferences & distribution	
ψ	EIS	1.5
γ_0	Risk aversion of passive investors	10
γ_j	Risk aversion of active investors	(9.368, 4.925, 3.212, 1.552)
ρ	Rate of time preference	0.01
	Technology	
μ	Endowment growth rate	0.022
σ	Endowment volatility	0.035
	Passive demand	
$\overline{\alpha}$	Mean	0.25
θ_p	Mean reversion parameter	0.9
	Leverage constraints	
$\overline{\sigma}$	Tightness of the leverage constraint	0.05

Figure 10 plots the consumption-wealth ratio, dividend yield, interest rate, and price of risk as a function of the wealth share of active agents.

Figure 11 displays the price impact in the left panel and the aggregate elasticity (i.e., its inverse) in the right panel. We see that the price impact is comparable with the solution obtained using the global perturbation method in Section 4.

6 Conclusion

This paper provides an analysis of the determinants of aggregate market elasticity in a general equilibrium framework with rich investor heterogeneity, passive investment, and financial constraints. Our analysis yields several important insights about market inelasticity and its implications for asset prices and market volatility.

A central finding of our work is the crucial role of cross-price elasticity–the sensitivity of demand for risky assets to changes in interest rates. When cross-elasticity is zero, aggregate market elasticity is simply an average of individual investors' price elasticities, consistent with partial equilibrium intuition. However, with non-zero cross-elasticity, the market can be infinitely elastic

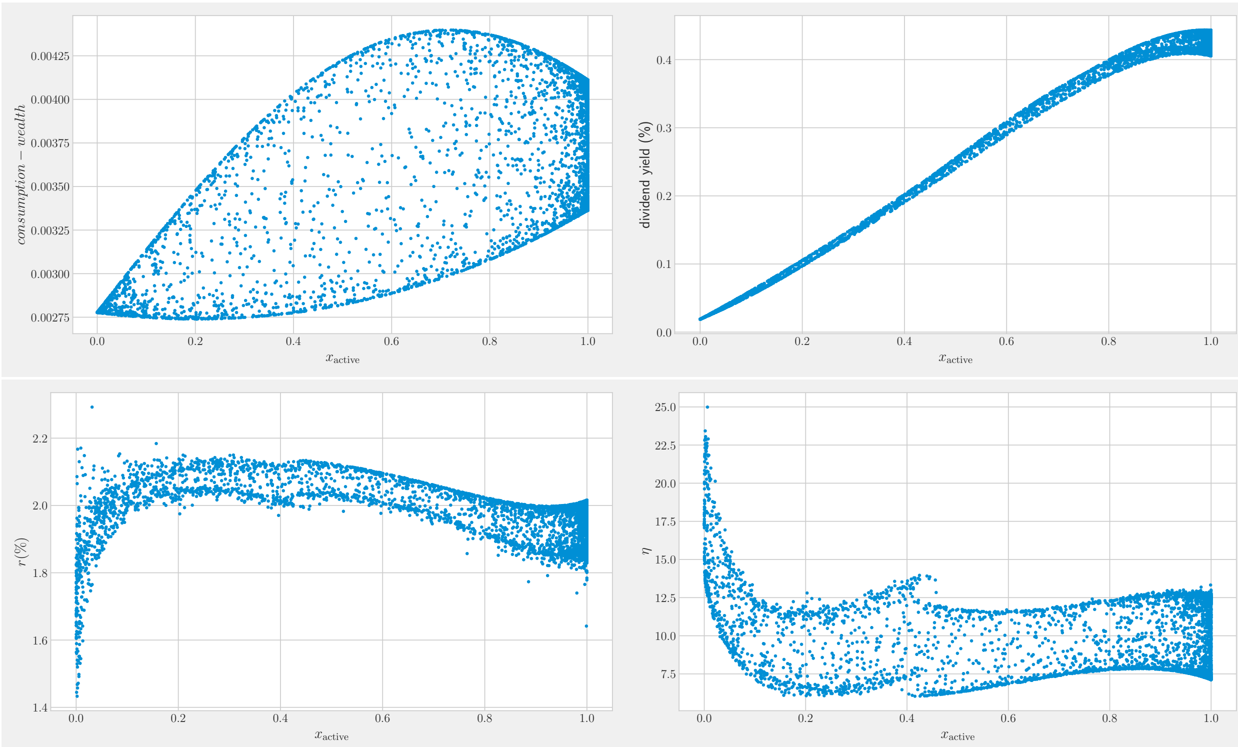

Figure 10. Model results. This figure plots the consumption-wealth ratio, dividend yield, interest rate, and price of risk as a function of the wealth share of active agents.

even when individual investors are highly inelastic, as changes in interest rates offset movements in risk premia. This highlights how general equilibrium effects fundamentally alter the relationship between individual and aggregate elasticities.

We show that beyond cross-elasticity effects, the key determinant of market elasticity is how risk is allocated in the economy. When passive investors hold an efficient share of risky assets, the market remains infinitely elastic regardless of individual investor preferences. However, when risk is misallocated, portfolio flows have meaningful price impacts.

Our model demonstrates that market elasticity is both state-dependent and time-varying, influenced by the distribution of wealth between active and passive investors as well as the allocation of wealth among active investors themselves. While passive investment and leverage constraints amplify the price impact of flows, preference heterogeneity can make markets more elastic by improving risk allocation. These findings highlight that market inelasticity serves as an indicator of underlying inefficiencies in risk allocation rather than a fundamental feature of markets.

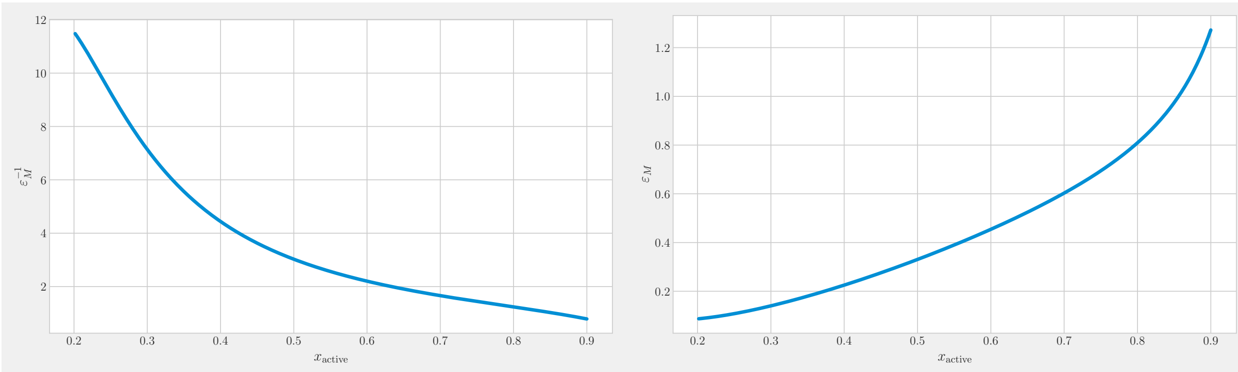

Figure 11. Price impact and aggregate elasticity.
The left panel plots the price impact and the right panel plots the aggregate elasticity.

The framework developed here helps explain several empirical patterns, including the countercyclical nature of the volatility multiplier and the relationship between passive investment growth and market dynamics. It also provides a foundation for analyzing how different market frictions interact to influence aggregate elasticity and market stability. Our results suggest that policies aimed at improving risk allocation may be more effective at reducing excess volatility than those focused on individual investor behavior.

Future research could extend this framework to study the implications of market inelasticity for asset pricing anomalies, monetary policy transmission, and financial stability. The role of market structure and trading mechanisms in determining aggregate elasticity also remains an important area for investigation.

References

Adrian, Tobias, and Hyun Song Shin, 2014, Procyclical leverage and value-at-risk, *Review of Financial Studies* 27, 373–403.

Ameriks, John, and Stephen P. Zeldes, 2004, How do household portfolio shares vary with age?, Working Paper, Columbia University.

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

Basak, Suleyman, and Domenico Cuoco, 1998, An equilibrium model with restricted stock market participation, *Review of Financial Studies* 11, 309–341.

Brunnermeier, Markus K., and Stefan Nagel, 2008, Do wealth fluctuations generate time-varying risk aversion? micro-evidence on individuals, *American Economic Review* 98, 713–36.

Brunnermeier, Markus K., and Lasse Heje Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22, 2201–2238.

Campbell, John Y., and Albert S. Kyle, 1993, Smart money, noise trading and stock price behaviour, *Review of Economic Studies* 60, 1–34.

Chabakauri, Georgy, 2013, Dynamic equilibrium with two stocks, heterogeneous investors, and portfolio constraints, *Review of Financial Studies* 26, 3104–3141.

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.

Chen, Hui, Winston Wei Dou, and Leonid Kogan, 2024, Measuring "dark matter" in asset pricing models, *The Journal of Finance* 79, 843–902.

Chinco, Alex, and Marco Sammon, 2024, The passive ownership share is double what you think it is, *Journal of Financial Economics* 157, 103860.

Cochrane, John H., 1992, Explaining the variance of price–dividend ratios, *Review of Financial Studies* 5, 243–280.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.

Cremers, KJ Martijn, and Antti Petajisto, 2009, How active is your fund manager? a new measure that predicts performance, *Review of Financial Studies* 22, 3329–3365.

Deuskar, Prachi, and Timothy C. Johnson, 2011, Market Liquidity and Flow-driven Risk, *Review of Financial Studies* 24, 721–753.

Duarte, Victor, Diogo Duarte, and Dejanir H. Silva, 2024, Machine learning for continuous-time finance, *Review of Financial Studies* 37, 3217–3271.

Duffie, Darrell, and Larry G. Epstein, 1992, Stochastic differential utility, *Econometrica* 353–394.

Gabaix, Xavier, and Ralph SJ Koijen, 2020, In search of the origins of financial fluctuations: The inelastic markets hypothesis, Working Paper, Harvard and Chicago Booth.

Garleanu, Nicolae, and Lasse Heje Pedersen, 2011, Margin-based asset pricing and deviations from the law of one price, *Review of Financial Studies* 24, 1980–2022.

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

Hartzmark, Samuel M., and David H. Solomon, 2021, Predictable price pressure, Working Paper, Chicago Booth and Boston College.

Hugonnier, Julien, 2012, Rational asset pricing bubbles and portfolio constraints, *Journal of Economic Theory* 147, 2260–2302.

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

Johnson, Timothy C., 2008, Volume, liquidity, and liquidity risk, *Journal of Financial Economics* 87, 388–417.

Johnson, Timothy C., 2009, Liquid capital and market liquidity, *The Economic Journal* 119, 1374–1404.

Kargar, Mahyar, 2021, Heterogeneous intermediary asset pricing, *Journal of Financial Economics* 141, 505–532.

Kargar, Mahyar, Juan Passadore, and Dejanir Silva, 2020, Liquidity and risk in OTC markets: A theory of asset pricing and portfolio flows, Working Paper, UIUC, EIEF, and Purdue.

Khorrami, Paymon, 2022, Entry and slow-moving capital: using asset markets to infer the costs of risk concentration, *Available at SSRN 2777747* .

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo, 2024, Which investors matter for equity valuations and expected returns?, *Review of Economic Studies* .

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 1315–1335.

LeRoy, Stephen F., and Richard D. Porter, 1981, The present-value relation: Tests based on implied variance bounds, *Econometrica* 555–574.

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

Panageas, Stavros, 2020, The implications of heterogeneity and inequality for asset pricing, *Foundations and Trends® in Finance* 12, 199–275.

Parker, Jonathan A., Antoinette Schoar, Allison T Cole, and Duncan Simester, 2022, Household portfolios and retirement saving over the life cycle, Technical report, National Bureau of Economic Research.

Parker, Jonathan A., Antoinette Schoar, and Yang Sun, 2023, Retail financial innovation and stock market dynamics: The case of target date funds, *Journal of Finance* 78, 2673–2723.

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

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

Schmitt-Grohé, Stephanie, and Martın Uribe, 2004, Solving dynamic general equilibrium models using a second-order approximation to the policy function, *Journal of economic dynamics and control* 28, 755–775.

Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *The American Economic Review* 71, 421–436.

Shiller, Robert J., 1992, *Market Volatility* (MIT press).

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

Silva, Dejanir, 2020, The risk channel of unconventional monetary policy, Working Paper, Purdue.

Veronesi, Pietro, 2019, Heterogeneous households under uncertainty, Working Paper 25448, National Bureau of Economic Research.

Appendix

A Derivations for Section 2

A.1 The bond demand view

The bond demand of investor j is given by

$$B_j = R_f(1-\alpha_j)(1-c_j(r+\pi))W_j. \tag{A.1}$$

Let $b_j \equiv \frac{B_j}{W_j}$ denote the bond-to-wealth ratio. The market clearing condition for bonds can be expressed as follows

$$\underbrace{x_a b_a}_{\textbf{active bond demand}} = \underbrace{-x_p b_p}_{\textbf{net bond supply}} \tag{A.2}$$

The linearized bond demand for a passive investor is given by

$$b_j - b_j^* = \left[r - r^* + \frac{c_j'(\mu - p^*)}{1 - c_j(\mu - p^*)}(p - p^*)\right] b_j^* - \alpha_j^* e^{r^*}(1 - c_j(\mu - p^*))\hat{\alpha}_j \tag{A.3}$$

For simplicity, focus on the case $\overline{\alpha}_p = 1$, so $b_p^* = b_a^* = 0$. The passive bond demand is then given by

$$x_p b_p = f^b, \tag{A.4}$$

where $f^b \equiv -x_p[1 - c_p(\mu - p^*)]\hat{\alpha}_p$. The active bond demand is given by

$$x_a b_a = -\zeta_p^b(p - p^*) - \zeta_r^b(r - r^*), \tag{A.5}$$

where $\zeta_p^b = \zeta_r^b = -x_a g_a'(\mu - p^* - r^*)[1 - c_a(\mu - p^*)]$.

The demand system can be written as follows:

$$\begin{bmatrix} \zeta_p^q & \zeta_r^q \\ \zeta_p^b & \zeta_r^b \end{bmatrix} \begin{bmatrix} p - p^* \\ r - r^* \end{bmatrix} = \begin{bmatrix} f^q \\ f^b \end{bmatrix}, \tag{A.6}$$

where we denote here $f^q \equiv f$ for symmetry.

Inverting the system above, we obtain

$$\begin{bmatrix} p - p^* \\ r - r^* \end{bmatrix} = \frac{1}{\zeta_p^q \zeta_r^b - \zeta_r^q \zeta_p^b} \begin{bmatrix} \zeta_r^b & -\zeta_r^q \\ -\zeta_p^b & \zeta_p^q \end{bmatrix} \begin{bmatrix} f^q \\ f^b \end{bmatrix}. \tag{A.7}$$

The price is given by

$$p - p^* = \frac{\zeta_r^b f^q - \zeta_r^q f^b}{\zeta_p^q \zeta_r^b - \zeta_r^q \zeta_p^b} = x_p \hat{\alpha}_p \frac{\zeta_r^b + \zeta_r^q (1 - c_p(\mu - p^*))}{\zeta_p^q \zeta_r^b - \zeta_r^q \zeta_p^b} = 0, \tag{A.8}$$

using the fact that $\zeta_r^q = x_a g_a'(\mu - p^* - r^*)$ and $c_a(\mu - p^*) = c_p(\mu - p^*)$.

B Derivations for Section 3

B.1 Investors' problem

The Hamilton-Jacobi-Bellman (HJB) equation for investor j can be written as

$$\begin{aligned} 0 = \max_{C_j, \alpha_j} & f_j(C_j, V_j) + V_{j,W}\left[rW_j + (\mu_R - r)\alpha_j W_j - C_j\right] + V_{j,X}\mu_X \\ & + \frac{1}{2} V_{j,WW} W_j^2 \alpha_j^2 \|\sigma_R\|^2 + V_{j,WX} W_j \sigma_X \sigma_R' \alpha_j + \frac{1}{2}\sum_{k=1}^{d} \sigma_{X,k}' V_{j,XX} \sigma_{X,k}, \end{aligned} \tag{B.1}$$

subject to $\alpha_j \in \Omega_j$. For ease of notation, we dropped time subscripts. Note that $V_{j,X}$ and $V_{j,WX}$ are $1 \times N$ vectors, $V_{j,XX}$ is a $N \times N$ matrix, and both $V_{j,W}$ and $V_{j,WW}$ are scalars. The drift μ_X is a $N \times 1$ vector, the diffusion σ_X is a $N \times d$ matrix, while σ_R is a $1 \times d$ vector. The notation $\sigma_{X,k}$ denotes the $k-$th column of σ_X, that is, the exposure to the $k-$th Brownian motion.

The optimal consumption is given by

$$C_j = \rho^\psi ((1 - \gamma_j) V_j)^{\frac{1 - \gamma_j \psi}{1 - \gamma_j}} V_{j,W}^{-\psi}. \tag{B.2}$$

The optimal portfolio share for an active investor is given by

$$\alpha_j = \min\left\{ -\frac{V_{j,W}(\mu_R - r)}{V_{j,WW} W \|\sigma_R\|^2} - \frac{V_{WX}}{V_{WW} W} \frac{\sigma_X \sigma_R'}{\|\sigma_R\|^2}, \frac{\overline{\sigma}}{\|\sigma_{R,t}\|} \right\}. \tag{B.3}$$

Given the homotheticity of preferences, the value function for investor j can be written as

$$V_{j,t} = \left(\frac{\xi_{j,t}}{\rho^\psi}\right)^{\frac{1-\gamma_j}{1-\psi}} \frac{W_{j,t}^{1-\gamma_j}}{1 - \gamma_j}. \tag{B.4}$$

This particular parametrization of the value function implies that the consumption-wealth ratio is given by

$$\frac{C_{j,t}}{W_{j,t}} = \xi_{j,t}. \tag{B.5}$$

The optimal portfolio share for active investors is given by

$$\alpha_{j,t} = \min\left\{\frac{\mu_{R,t} - r_t}{\gamma_j \|\sigma_{R,t}\|^2} - \frac{1-\gamma_j^{-1}}{1-\psi}\frac{\sigma_{\xi_j,t}\sigma'_{R,t}}{\|\sigma_{R,t}\|^2}, \frac{\overline{\sigma}}{\|\sigma_{R,t}\|}\right\}. \tag{B.6}$$

It is convenient to consider the investor's risk exposure $\sigma_j \equiv \alpha_j \|\sigma_R\|$, which is then given by

$$\sigma_{j,t} = \min\left\{\frac{\eta_t}{\gamma_j} - \frac{1-\gamma_j^{-1}}{1-\psi}\frac{\sigma_{\xi_j,t}\sigma'_{R,t}}{\|\sigma_{R,t}\|}, \overline{\sigma}\right\}, \tag{B.7}$$

where $\eta_t \equiv \frac{\mu_{R,t}-r_t}{\|\sigma_{R,t}\|}$ denotes the Sharpe ratio of the risky asset.

Plugging the consumption-wealth ratio into the HJB equation and using Equation (B.4), we obtain

$$\begin{aligned}0 = {} & \frac{\rho}{1-\psi^{-1}}\left[\rho^{-1}\xi_j - 1\right] + r + \eta\sigma_j - \xi_j + \frac{1}{1-\psi}\left[\frac{\xi_{j,X}}{\xi_j}\mu_X + \frac{1}{2}\sum_{k=1}^{d}\sigma'_{X,k}\frac{\xi_{j,XX}}{\xi_j}\sigma_{X,k}\right] \\ & - \frac{\gamma_j}{2}\sigma_j^2 + \frac{1-\gamma_j}{1-\psi}\frac{\xi_{j,X}}{\xi_j}\sigma_X\frac{\sigma'_R}{\|\sigma_R\|}\sigma_j + \frac{1}{2}\frac{\psi-\gamma_j}{(1-\psi)^2}\sum_{k=1}^{d}\sigma'_{X,k}\frac{\xi'_{j,X}}{\xi_j}\frac{\xi_{j,X}}{\xi_j}\sigma_{X,k}.\end{aligned} \tag{B.8}$$

Rearranging the expression above, we obtain

$$\xi_{j,t} = \psi\rho + (1-\psi)\left[r_t + \eta_t\sigma_{j,t} - \frac{\gamma_j}{2}\sigma_{j,t}^2\right] + \mu_{\xi_j,t} + (1-\gamma_j)\sigma_{\xi_j,t}\frac{\sigma'_{R,t}}{\|\sigma_{R,t}\|}\sigma_{j,t} + \frac{\psi-\gamma_j}{1-\psi}\frac{\|\sigma_{\xi_j,t}\|^2}{2}, \tag{B.9}$$

where the law of motion of $\xi_{j,t}$ is given by

$$\frac{d\xi_{j,t}}{\xi_{j,t}} = \mu_{\xi_j,t}dt + \sigma_{\xi_j,t}dZ_t, \tag{B.10}$$

and the drift and diffusion of $\xi_{j,t}$ are given by Ito's lemma:

$$\mu_{\xi_j,t} = \frac{\xi_{j,X}}{\xi_j}\mu_{X,t} + \frac{1}{2}\sum_{k=1}^{d}\sigma'_{X,k,t}\frac{\xi_{j,XX,t}}{\xi_{j,t}}\sigma_{X,k,t}, \qquad \sigma_{\xi_j,t} = \frac{\xi_{j,X}}{\xi_j}\sigma_{X,t}. \tag{B.11}$$

B.2 Pricing condition

Let $y_t \equiv Y_t/P_t$ denote the dividend yield on the risky asset. From Equation (7), we can write the expected return on the risky asset as:

$$r_t + \eta_t\|\sigma_{R,t}\| = y_t + \frac{1}{dt}\frac{d(Y_t/y_t)}{(Y_t/y_t)} = y_t + \mu - \mu_{y,t} + \|\sigma_{y,t}\|^2 - \sigma\sigma'_{y,t}. \tag{B.12}$$

Rearranging the expression above, we obtain

$$y_t = r_t + \eta_t \|\sigma_{R,t}\| - \mu + \mu_{y,t} - \|\sigma_{R,t}\|^2 + \sigma\sigma'_{R,t}, \tag{B.13}$$

where $\sigma_{R,t} = \sigma - \sigma_{y,t}$ and $(\mu_{y,t}, \sigma_{y,t})$ are given by Ito's lemma:

$$\mu_{y,t} = y_{X,t}\mu_{X,t} + \frac{1}{2}\sum_{k=1}^{d} \sigma'_{X,k,t}\, y_{XX,t}\, \sigma_{X,k,t}, \qquad \sigma_{y,t} = y_{X,t}\sigma_{X,t}. \tag{B.14}$$

B.3 Aggregate state variable

Define the share of wealth of type-j investors as follows

$$x_{j,t} \equiv \frac{\omega_j W_{j,t}}{P_t}. \tag{B.15}$$

We define the aggregate state variable as $X_t = (x_t, \overline{\alpha}_{p,t})$, where $x_t \equiv (x_{1,t}, x_{2,t}, \ldots, x_{J,t})$. The law of motion of $\overline{\alpha}_{p,t}$ is given by (9). To compute the law of motion of $x_{j,t}$, first, note that the law of motion of wealth for a type-j investor can be written as

$$\frac{dW_{j,t}}{W_{j,t}} = \left[r_t + \eta_t\sigma_{j,t} - \xi_{j,t}\right] dt + \sigma_{j,t}\frac{\sigma_{R,t}}{\|\sigma_{R,t}\|} dZ_t \tag{B.16}$$

From Ito's lemma, the law of motion of $x_{j,t}$ is given by

$$\begin{aligned} \frac{dx_{j,t}}{x_{j,t}} &= \left(r_t + \eta_t\sigma_{j,t} - \xi_{j,t} - \mu + \mu_{y,t} + \sigma\sigma'_{R,t} - \sigma_{j,t}\|\sigma_{R,t}\| + \kappa\frac{\omega_j - x_{j,t}}{x_{j,t}}\right) dt \\ &\quad + (\sigma_{j,t} - \|\sigma_{R,t}\|)\frac{\sigma_{R,t}}{\|\sigma_{R,t}\|} dZ_t, \end{aligned} \tag{B.17}$$

using $\mu_{P,t} = \mu - \mu_{y,t} + \|\sigma_{R,t}\|^2 - \sigma\sigma'_{R,t}$.

B.4 Asset prices

Let $\mathcal{J}_t^u \subset \{1, 2, \ldots, J\}$ denote the set of unconstrained active investors at period t, that is, the set of investors such that $\sigma_{j,t} < \overline{\sigma}$. Let $\mathcal{J}_t^c \subset \{1, 2, \ldots, J\}$ denote the set of constrained active investors, that is, the set of investors such that $\sigma_{j,t} = \overline{\sigma}$. From the market clearing condition for the risky asset, the second term in Equation (23), we obtain

$$\eta_t = \frac{\gamma_{u,t}}{x_{u,t}}\left[(1 - \overline{\alpha}_{p,t}x_{0,t})\|\sigma_{R_t}\| - \overline{\sigma}x_{c,t} + \sum_{j\in\mathcal{J}_t^u} x_{j,t}\frac{1-\gamma_j^{-1}}{1-\psi}\frac{\sigma_{\xi_j,t}\sigma'_{R,t}}{\|\sigma_{R,t}\|}\right], \tag{B.18}$$

where $x_{u,t} \equiv \sum_{j\in\mathcal{J}_t^u} x_{j,t}$, $x_{c,t} \equiv \sum_{j\in\mathcal{J}_t^c} x_{j,t}$, and $\gamma_{u,t} \equiv \left[\frac{1}{x_{u,t}} \sum_{j\in\mathcal{J}_t^u} \frac{x_{j,t}}{\gamma_j}\right]^{-1}$ is the aggregate risk aversion of the unconstrained investors.

From the market clearing condition for goods, the first term in Equation (23), we obtain

$$
\begin{aligned}
y_t = \psi\rho + (1-\psi)\left[r_t + \eta_t\|\sigma_{R,t}\| - \sum_{j=0}^{J} x_{j,t}\frac{\gamma_j}{2}\sigma_{j,t}^2\right] \\
+ \sum_{j=0}^{J} x_{j,t}\left[\mu_{\xi_j,t} + (1-\gamma_j)\sigma_{\xi_j,t}\frac{\sigma'_{R,t}}{\|\sigma_{R,t}\|}\sigma_{j,t} + \frac{\psi-\gamma_j}{1-\psi}\frac{\|\sigma_{\xi_j,t}\|^2}{2}\right].
\end{aligned} \tag{B.19}
$$

Using the pricing condition (B.13), we obtain the expression for the risk-free rate

$$
\begin{aligned}
r_t = \rho - \eta_t\|\sigma_{R,t}\| + \psi^{-1}(\mu - \mu_{y,t} + \|\sigma_{R,t}\|^2 - \sigma\sigma'_{R,t}) + \left(1-\psi^{-1}\right)\sum_{j=0}^{J} x_{j,t}\frac{\gamma_j}{2}\sigma_{j,t}^2 \\
+ \psi^{-1}\sum_{j=0}^{J} x_{j,t}\left[\mu_{\xi_j,t} + (1-\gamma_j)\sigma_{\xi_j,t}\frac{\sigma'_{R,t}}{\|\sigma_{R,t}\|}\sigma_{j,t} + \frac{\psi-\gamma_j}{1-\psi}\frac{\|\sigma_{\xi_j,t}\|^2}{2}\right].
\end{aligned} \tag{B.20}
$$

B.5 The system of PDEs

To compute the equilibrium, one needs to solve a system of $J+2$ partial differential equations (PDEs), involving the consumption-wealth ratio $\xi_j(X)$ for the $J+1$ type of investors and the dividend yield $y(X)$. These functions depend on $J+1$ state variables, the $J-$dimensional vector x_t and the portfolio share of passive investors $\overline{\alpha}_{p,t}$.

The differential equation for the consumption-wealth ratio is given by

$$
\begin{aligned}
\xi_{j,t} = \psi\rho + (1-\psi)\left[r_t + \eta_t\sigma_{j,t} - \frac{\gamma_j}{2}\sigma_{j,t}^2\right] + \frac{\xi_{j,X}}{\xi_j}\mu_{X,t} + \frac{1}{2}\sum_{k=1}^{d}\sigma'_{X,k,t}\frac{\xi_{j,XX,t}}{\xi_{j,t}}\sigma_{X,k,t} \\
+ (1-\gamma_j)\frac{\xi_{j,X}}{\xi_j}\sigma_{X,t}\frac{\sigma'_{R,t}}{\|\sigma_{R,t}\|}\sigma_{j,t} + \frac{\psi-\gamma_j}{1-\psi}\frac{1}{2}\left\|\frac{\xi_{j,X}}{\xi_j}\sigma_{X,t}\right\|^2.
\end{aligned} \tag{B.21}
$$

Plugging the expressions for interest rate and the Sharpe ratio (r_t, η_t), the risk exposure $\sigma_{j,t}$, the drift and diffusion of the aggregate state variables $(\mu_{X,t}, \sigma_{X,t})$, and the aggregate volatility $\|\sigma_{R,t}\|$, we can express the condition above in terms of $\xi_{j,t}$ and y_t and their derivatives.

Similarly, we can write the condition for the dividend yield:

$$
\begin{aligned}
y_t = r_t + \eta_t \left\| \sigma - \frac{y_{X,t}}{y_t}\sigma_{X,t} \right\| - \mu + \frac{y_{X,t}}{y_t}\mu_{X,t} + \frac{1}{2}\sum_{k=1}^{d} \sigma'_{X,k,t}\frac{y_{XX,t}}{y_t}\sigma_{X,k,t} \\
- \left\| \sigma - \frac{y_{X,t}}{y_t}\sigma_{X,t} \right\|^2 + \sigma \left(\sigma - \frac{y_{X,t}}{y_t}\sigma_{X,t} \right)',
\end{aligned}
\tag{B.22}
$$

which again can be expressed only in terms of $\xi_{j,t}$ and y_t and their derivatives.

C Derivations for Section 4

C.1 Proof of Lemma 1

Proof. The assumption $\epsilon = 0$ implies that there is no preference heterogeneity and passive investors are fully invested in the risky asset. We guess and verify that in this benchmark economy, there are no variation in expected returns. In particular, the wealth distribution plays no role in the economy. This implies that $\mu_{c_j,0}(X) = \sigma_{c_j,0}(X) = \mu_{p,0}(X) = \sigma_{p,0}(X) = 0$. In this case, the risk premium is given by

$$
\pi_0(X) = \frac{\gamma}{x_{u,t}} \left[1 - x_{0,t} - x_{c,t}\overline{\alpha}_{c,t}\right] \|\sigma_{R_t}\|^2,
\tag{C.1}
$$

using the fact that $\varsigma_t = 0$, as $\sigma_{c_j,t} = 0$, and $\overline{\alpha}_{p,t} = 1$.

Given that $\sigma_{y,t} = 0$, we have that $\sigma_{R,0}(X) = \sigma$. Using the fact that $\overline{\sigma} = \|\sigma\|$ and $x_{u,t} = 1 - x_{0,t} - x_{c,t}$, we obtain the risk premium

$$
\pi_0(X) = \gamma\|\sigma\|^2,
$$

using $\alpha_{c,t} = 1$. Using $\sigma_{c_j,t} = 0$ and the expression for $\pi_0(X)$, we obtain that $\alpha_{j,0}(X) = 1$, for $j = 1, \ldots, J$, from Equation (19).

The interest rate is given by

$$
r_0(X) = \rho + \psi^{-1}\mu - \gamma(1+\psi^{-1})\frac{\|\sigma\|^2}{2}.
\tag{C.2}
$$

The consumption-wealth ratio $c_{j,0}(X)$ is given by

$$
c_{j,0}(X) = \psi\rho + (1-\psi)\left[r_0(X) + \pi_0(X)\alpha_{j,0}(X) - \frac{\gamma}{2}\alpha_{j,0}(X)^2\|\sigma\|^2\right].
\tag{C.3}
$$

Plugging in the expression for $r_0(X)$, $\eta_0(X)$ and $\sigma_{j,0}$, we obtain

$$c_{j,0}(X) = \rho - \left(1 - \psi^{-1}\right)\left(\mu - \frac{\gamma\|\sigma\|^2}{2}\right), \tag{C.4}$$

where we assume $\rho > \left(1 - \psi^{-1}\right)\left(\mu - \frac{\gamma\|\sigma\|^2}{2}\right)$.

From the market clearing condition for goods, we obtain:

$$\frac{1}{p_0(X)} = \rho - \left(1 - \psi^{-1}\right)\left(\mu - \frac{\gamma\|\sigma\|^2}{2}\right). \tag{C.5}$$

The drift and diffusion of the wealth shares are given

$$\mu_{X,j,0}(X) = x_j\left[r_0(X) + \pi_0(X)\alpha_{j,0}(X) - c_{j,0}(X) - \mu\right] \tag{C.6}$$

$$\sigma_{X,j,0}(X) = x_j(\alpha_{j,0}(X) - 1)\sigma_{R,0}(X), \tag{C.7}$$

where $\mu_{X,j,0} = \sigma_{X,j,0} = 0$, using the expression for returns, portfolio share, and consumption-wealth ratio. The result $\mu_{X,j,0} = 0$ uses the fact that $\kappa = 0$. $\square$

C.2 Proof of Proposition 2

Proof. We consider next the first-order correction terms. Note that the diffusion terms for ξ_j and y are both equal to zero up to the first order, since $\sigma_{\xi_j,t} = O(\epsilon^2)$ and $\sigma_{y,t} = O(\epsilon^2)$. From Ito's lemma:

$$\sigma_{\xi_j,t} = \underbrace{\frac{\xi_{j,X}}{\xi_j}}_{O(\epsilon)}\underbrace{\sigma_{X,t}}_{O(\epsilon)} = O(\epsilon^2). \tag{C.8}$$

We have $\xi_{j,X} = O(\epsilon)$, because $\xi_{j,X,0} = 0$, as $\xi_{j,0}(X)$ does not depend on X. Also, $\sigma_{X,t} = O(\epsilon)$ because $\sigma_{X,0}(X) = 0$. This implies that $\sigma_{\xi_j,1}(X) = 0$. An analogous argument applies to y_t, so that we have $\sigma_{y,1}(X) = 0$ and $\|\sigma_{R,1}(X)\| = 0$.

Risk exposure of active investors. The risk exposure for active investors can be written as

$$\sigma_j(X,\epsilon) = \min\left\{\frac{\eta(X,\epsilon)}{\gamma_j} + \frac{1-\gamma_j^{-1}}{\psi - 1}\frac{\sigma_{\xi_j}(X,\epsilon)\sigma_R'(X,\epsilon)}{\|\sigma_R(X,\epsilon)\|}, \|\sigma\| + \hat{\sigma}\epsilon\right\}, \tag{C.9}$$

Expanding the first term inside brackets in ϵ, we obtain

$$\sigma_j(X,\epsilon) = \min\left\{\frac{\eta_0(X)}{\gamma} + \left(\frac{\eta_1(X)}{\gamma} - \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j\right)\epsilon + \frac{1-\gamma^{-1}}{\psi-1}\frac{\sigma_{\xi_j,2}(X)\sigma'}{\|\sigma\|}\epsilon^2 + O(\epsilon^3), \|\sigma\| + \hat{\sigma}\epsilon\right\}, \tag{C.10}$$

Adding and subtracting $\|\sigma\| + \hat{\sigma}\epsilon$, and using $\frac{\eta_0(X)}{\gamma} = \|\sigma\|$, we obtain

$$\sigma_j(X,\epsilon) = \min\left\{\left(\frac{\eta_1(X)}{\gamma} - \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j - \hat{\sigma}\right)\epsilon + \frac{1-\gamma^{-1}}{\psi - 1}\frac{\sigma_{\xi_j,2}(X)\sigma'}{\|\sigma\|}\epsilon^2 + O(\epsilon^3), 0\right\} + \|\sigma\| + \hat{\sigma}\epsilon. \quad \text{(C.11)}$$

Consider first the case where the following condition is satisfied:

$$\frac{\eta_1(X)}{\gamma} - \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j - \hat{\sigma} = \mathcal{O}(1), \quad \text{(C.12)}$$

If this is the case, then

$$\left(\frac{\eta_1(X)}{\gamma} - \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j - \hat{\sigma}\right)\epsilon \gg \left|\frac{1-\gamma^{-1}}{\psi - 1}\frac{\sigma_{\xi_j,2}(X)\sigma'}{\|\sigma\|}\right|\epsilon^2, \quad \text{(C.13)}$$

for small ϵ. So, the sign of the term inside the min operator in (C.11) is determined by the first term.

We can then write $\sigma_j(X,\epsilon)$ as follows:

$$\sigma_j(X,\epsilon) = \|\sigma\| + \min\left\{\frac{\eta_1(X)}{\gamma} - \frac{\eta_0(X)}{\gamma}\hat{\gamma}, \hat{\sigma}\right\}\epsilon + \mathcal{O}(\epsilon^2). \quad \text{(C.14)}$$

In the region of the state space where condition (C.12) holds, one can determine whether an investor is constrained or unconstrained only based on the first-order terms. Suppose now that the following condition holds

$$\frac{\eta_1(X)}{\gamma} - \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j - \hat{\sigma} = \mathcal{O}(\epsilon). \quad \text{(C.15)}$$

This condition states that, up to the first order, the leverage constraint is either always binding or slack by just a tiny amount parameterized by ϵ and ϵ^2 terms inside the min operator in (C.11). In this case, we can write $\sigma_j(X,\epsilon)$ as follows:

$$\sigma_j(X,\epsilon) = \|\sigma\| + \hat{\sigma}\epsilon + \min\left\{\left(\frac{\eta_1(X)}{\gamma} - \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j - \hat{\sigma}\right)\epsilon + \frac{1-\gamma^{-1}}{\psi - 1}\frac{\sigma_{\xi_j,2}(X)\sigma'}{\|\sigma\|}\epsilon^2, 0\right\} + \mathcal{O}(\epsilon^3). \quad \text{(C.16)}$$

In this region of the state space where condition (C.15) is satisfied, we need the second-order term to determine whether an investor is constrained. This distinction will be relevant when computing the second-order correction.

For the first-order correction terms here, we focus on the case where condition (C.12) holds.

Aggregate risk aversion. The aggregate risk aversion of unconstrained investors, defined above, is given by

$$\gamma_u(X,\epsilon) = \frac{x_u}{\sum_{j\in\mathcal{J}^u}\frac{x_j}{\gamma(1+\hat{\gamma}\epsilon)}} = \gamma + \gamma\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j\,\epsilon + O(\epsilon^2) \tag{C.17}$$

Market price of risk. The market price of risk can be written as

$$\begin{aligned}\eta(X,\epsilon) &= \frac{\gamma_u(X,\epsilon)}{1-x_0-x_c}\left[(1-(1+\hat{\alpha}_p\epsilon)x_0)\|\sigma\| - (\|\sigma\|+\hat{\sigma}\epsilon)x_c\right] + O(\epsilon^2),\\ &= \eta_0(X) + \gamma\|\sigma\|\left[\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j - \left(\frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\|\sigma\|}x_c}{1-x_0-x_c}\right)\right]\epsilon + O(\epsilon^2).\end{aligned} \tag{C.18}$$

In the region of the state space where all active investors are unconstrained, we have

$$\eta(X,\epsilon) = \eta_0(X) + \gamma\|\sigma\|\left(\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j - \frac{\hat{\alpha}_p\,x_0}{1-x_0}\right)\epsilon + O(\epsilon^2). \tag{C.19}$$

The expression above shows the impact of fluctuations in the aggregate risk aversion and the effect of portfolio inflows in the market price of risk. If the average risk aversion in state X is lower than its level at $\epsilon = 0$, then the market price of risk will be lower than its level at $\epsilon = 0$, everything else constant.

Interest rate. The interest rate is given by

$$\begin{aligned}r(X,\epsilon) &= r_0(X) + \left[-\eta_1(X)\|\sigma\| + \left(1-\psi^{-1}\right)\left(\sum_{j=0}^{J}x_{j,t}\frac{\gamma}{2}2\|\sigma\|\sigma_{j,1}(X) + \sum_{j=0}^{J}x_{j,t}\frac{\gamma}{2}\|\sigma\|^2\hat{\gamma}_j\right)\right]\epsilon + O(\epsilon^2)\\ &= r_0(X) + \gamma\|\sigma\|^2\left[-\frac{\eta_1(X)}{\gamma\|\sigma\|} + \left(1-\psi^{-1}\right)\sum_{j=0}^{J}x_{j,t}\frac{\hat{\gamma}_j}{2}\right]\epsilon + O(\epsilon^2)\\ &= r_0(X) - \gamma\|\sigma\|^2\left[\left(\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j - \sum_{j=0}^{J}x_{j,t}\frac{\hat{\gamma}_j}{2}\right) - \frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\|\sigma\|}x_c}{1-x_0-x_c} + \psi^{-1}\sum_{j=0}^{J}x_{j,t}\frac{\hat{\gamma}_j}{2}\right]\epsilon + O(\epsilon^2),\end{aligned} \tag{C.20}$$

where we use the fact that $\sum_{j=0}^{J}x_{j,t}\sigma_{j,1}(X) = 0$ from the first-order correction of the risky asset market clearing condition.

The term in the square brackets in Equation (C.20) captures the first-order effect of frictions on the interest rate, $r_1(X)$. First, we see that the interest rate is decreasing in the difference between the average risk aversion of unconstrained investors and the average risk aversion of all investors in the economy. These two averages can differ for two reasons. First, passive investors may have a risk aversion different from the average unconstrained investor, the first term in parentheses in $r_1(X)$. Second, constrained investors are exactly the ones with low risk aversion, so unconstrained investors are on average more risk averse than all investors in the economy, which include the low risk aversion ones.

Consumption-wealth ratio. The consumption-wealth ratio for investor j is given by

$$\begin{aligned}\xi_j(X,\epsilon) &= \xi_{j,0}(X) + (1-\psi)\left[r_1(X) + \eta_0(X)\sigma_{j,1}(X) + \eta_1(X)\sigma_{j,0}(X)\right.\\ &\qquad\left. -\gamma\|\sigma\|\sigma_{j,1}(X) - \frac{\gamma}{2}\hat{\gamma}_j\|\sigma\|^2\right]\epsilon + O(\epsilon^2)\\ &= \xi_{j,0}(X) + (1-\psi)\gamma\|\sigma\|^2\left[\frac{r_1(X)}{\gamma\|\sigma\|^2} + \frac{\eta_1(X)}{\gamma\|\sigma\|} - \frac{\hat{\gamma}_j}{2}\right]\epsilon + O(\epsilon^2)\\ &= \xi_{j,0}(X) + (1-\psi)\gamma\|\sigma\|^2\left[\left(1-\psi^{-1}\right)\sum_{j=0}^{J} x_{j,t}\frac{\hat{\gamma}_j}{2} - \frac{\hat{\gamma}_j}{2}\right]\epsilon + O(\epsilon^2).\end{aligned} \tag{C.21}$$

Dividend yield. The dividend yield is given by

$$y(X,\epsilon) = y_0(X) + \left(1-\psi^{-1}\right)\gamma\|\sigma\|^2\sum_{j=0}^{J} x_{j,t}\frac{\hat{\gamma}_j}{2}\epsilon + O(\epsilon^2). \tag{C.22}$$

Notice that portfolio flows do not affect the dividend yield up to first order. The reason is that the interest rate and risk premium effects exactly cancel each other out. To derive the effect of portfolio flows on asset prices, we need to consider the second-order correction.

Wealth dynamics. The diffusion of the wealth share of investor j, $j = 1, \ldots, J$, is given by

$$\begin{aligned}\sigma_{X,j}(X) &= x_j\sigma_{j,1}(X)\frac{\sigma}{\|\sigma\|}\epsilon + O(\epsilon^2)\\ &= x_j \min\left\{\sum_{k\in\mathcal{J}^u}\frac{x_k}{x_u}\hat{\gamma}_j - \left(\frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\|\sigma\|}x_c}{1-x_0-x_c}\right) - \hat{\gamma}_j, \frac{\hat{\sigma}}{\|\sigma\|}\right\}\sigma\epsilon + O(\epsilon^2).\end{aligned} \tag{C.23}$$

The drift of the wealth share of investor j, $j = 1, \ldots, J$, is given by

$$\begin{aligned}\mu_{X,j}(X,\epsilon) &= x_j\left[r_1(X) + \eta_1(X)\sigma_{j,0}(X) + \eta_0(X)\sigma_{j,1}(X) - \xi_{j,1}(X) - \sigma_{j,1}(X)\|\sigma\|\right]\epsilon + \kappa(\omega_j - x_j) + O(\epsilon^2)\\ &= x_j\left[(\psi-1)\gamma\|\sigma\|^2\left(\sum_{k=0}^{J} x_{k,t}\frac{\hat{\gamma}_k}{2} - \frac{\hat{\gamma}_j}{2}\right) + (\gamma-1)\|\sigma\|\sigma_{j,1}(X)\right]\epsilon + \kappa(\omega_j - x_j) + O(\epsilon^2)\end{aligned} \tag{C.24}$$

$\square$

C.3 Second-order correction

In Proposition 6, we compute the second-order correction for our economy.

Proposition 6 (Second-order correction)**.** *Suppose* $\rho > \left(1-\psi^{-1}\right)\left(\mu - \frac{\gamma\sigma^2}{2}\right)$*. Then,*

(i) The second-order correction for the consumption-wealth ratio and risk exposure are given by:

$$\xi_{j,2}(X) = (1-\psi)\frac{\gamma\sigma^2}{2}\left[\left(1-\psi^{-1}\right)\sum_{k=0}^{J} x_{k,t}\hat{\gamma}_k - \hat{\gamma}_j\right] \tag{C.25}$$

$$\sigma_{j,2}(X) = \frac{\eta_2(X)}{\gamma} - \frac{\eta_1(X)}{\gamma}\hat{\gamma}_j + \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j^2 - (1-\gamma^{-1})\sigma_{y,2}(X), \tag{C.26}$$

where

$$\sigma_{y,2}(X) = \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2y_0(X)}\sum_{k=1}^{J}(\hat{\gamma}_k - \hat{\gamma}_0)\, x_k \sigma_{k,1}(X). \tag{C.27}$$

(ii) The second-order correction for the Sharpe ratio, interest rate, and dividend yield are given by:

$$\eta_2(X) = -\frac{(\gamma-1)x_c + 1 - x_0}{1 - x_0 - x_c}\sigma_{y,2}(X) - \gamma\sigma\mathbb{E}^u[\hat{\gamma}_j]\frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c} - \gamma\sigma\mathrm{Var}^u[\hat{\gamma}_j] \tag{C.28}$$

$$\begin{aligned} r_2(X) &= -\eta_2(X)\sigma + \eta_0(X)\sigma_{y,2}(X) - \psi^{-1}(\mu_{y,2}(X) + \sigma\sigma_{y,2}(X)) \\ &\quad + \left(1-\psi^{-1}\right)\sum_{j=0}^{J} x_j\gamma\sigma\left[\sigma_{j,2}(X) + \frac{\sigma_{j,1}^2(X)}{2\sigma} + \hat{\gamma}_j\sigma_{j,1}\right] \\ &\quad + \psi^{-1}\sum_{j=0}^{J} x_j\left[\mu_{\xi_j,2}(X) + (1-\gamma)\sigma_{\xi_j,2}(X)\sigma\right] \end{aligned} \tag{C.29}$$

$$y_2(X) = \left(1-\psi^{-1}\right)\sum_{j=0}^{J} x_j\gamma\sigma\left(\frac{\sigma_{j,1}^2(X)}{2\sigma} + \hat{\gamma}_j\sigma_{j,1}(X)\right), \tag{C.30}$$

where

$$\mathbb{E}^u[\hat{\gamma}_j] \equiv \sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j, \quad \textit{and} \quad \mathrm{Var}^u[\hat{\gamma}_j] \equiv \sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j^2 - \left(\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j\right)^2.$$

Proof. **Step1: Laws of motion for ξ_j and y.**
We start by considering the diffusion terms for ξ_j and y. Given that we are abstracting from portfolio

shocks, we assume that $d = 1$ without further loss of generality, so that we can treat diffusion terms as scalars. The expansion of $\sigma_{\xi_j,t}$ in ϵ is given by

$$\begin{aligned}\sigma_{\xi_j}(X,\epsilon) &= \frac{\xi_{j,X}(X,\epsilon)}{\xi_j(X,\epsilon)}\sigma_X(X,\epsilon)\\ &= \frac{\xi_{j,X,1}(X)}{\xi_{j,0}(X)}\sigma_{X,1}(X)\epsilon^2 + O(\epsilon^3)\\ &= -(\psi-1)\left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2\xi_{j,0}(X)}\sum_{k=1}^{J}\left(\hat{\gamma}_k-\hat{\gamma}_0\right)x_k\sigma_{k,1}(X)\epsilon^2 + O(\epsilon^3).\end{aligned} \tag{C.31}$$

Notice that $\sigma_{\xi_j,2}(X)$ does not depend on j, that is, it is the same for all investors. Moreover, $\sigma_{\xi_j,2}(X) > 0$, as $\sigma_{k,1}(X)$ is inversely related to $\hat{\gamma}_k$.

Similarly, the diffusion for dividend yield y can be written as

$$\begin{aligned}\sigma_y(X,\epsilon) &= \frac{y_{X,1}(X)}{y_0(X)}\sigma_{X,1}(X)\epsilon^2 + O(\epsilon^2)\\ &= \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2y_0(X)}\sum_{k=1}^{J}\left(\hat{\gamma}_k-\hat{\gamma}_0\right)x_k\sigma_{k,1}(X)\epsilon^2 + O(\epsilon^3).\end{aligned} \tag{C.32}$$

The expression above is negative if $\psi > 1$, which implies that $\sigma_{R,2}(X) = -\sigma_{y,2}(X)$ is positive. In this case, a negative aggregate shock redistribute wealth to more risk averse investors, leading to a rise in the risk premium and a decline in the risk-free rate. If $\psi > 1$, the risk premium effect dominates, so the price-dividend ratio, $1/y$, falls in response to the shock. The movement in the price-dividend ratio amplifies the initial effect of the drop in dividends.

The drift of $\xi_{j,t}$ is given by

$$\begin{aligned}\mu_{\xi_j}(X,\epsilon) &= \frac{\xi_{j,X}(X,\epsilon)}{\xi_j(X,\epsilon)}\mu_X(X,\epsilon) + \frac{1}{2}\sigma_X'(Z,\epsilon)\frac{\xi_{j,XX}(X,\epsilon)}{\xi_j(X,\epsilon)}\sigma_X(X,\epsilon)\\ &= \frac{\xi_{j,X,1}(X)}{\xi_{j,0}(X)}\mu_{X,1}(X)\epsilon^2 + O(\epsilon^3)\\ &= -(\psi-1)\left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2\xi_{j,0}(X)}\sum_{k=1}^{J}\left(\hat{\gamma}_k-\hat{\gamma}_0\right)\mu_{X_k,1}(X)\epsilon^2 + O(\epsilon^3).\end{aligned} \tag{C.33}$$

The drift for y_t is given by

$$
\begin{aligned}
\mu_y(X,\epsilon) &= \frac{y_X(X,\epsilon)}{y(X,\epsilon)}\mu_X(X,\epsilon) + \frac{1}{2}\sigma'_X(Z,\epsilon)\frac{y_{XX}(X,\epsilon)}{y(X,\epsilon)}\sigma_X(X,\epsilon) \\
&= \frac{y_{X,1}(X)}{y_0(X)}\mu_{X,1}(X)\epsilon^2 + O(\epsilon^3) \\
&= \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2y_0(X)}\sum_{k=1}^{J}\left(\hat{\gamma}_k - \hat{\gamma}_0\right)\mu_{X_k,1}(X)\epsilon^2 + O(\epsilon^3). \qquad \text{(C.34)}
\end{aligned}
$$

Step 2: Risk exposures of investors.

We focus on the *inner region*, that is, the case where all investors are sufficiently far from the constraint boundary (on either side). For a constrained investor, the second-order term is zero. For an unconstrained investor, the second-order term is given by

$$
\sigma_{j,2}(X) = \frac{\eta_2(X)}{\gamma} - \frac{\eta_1(X)}{\gamma}\hat{\gamma}_j + \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j^2 + \frac{1-\gamma^{-1}}{\psi-1}\sigma_{\xi_j,2}(X), \qquad \text{(C.35)}
$$

where investor j is unconstrained if the following condition holds:

$$
\sum_{k\in\mathcal{J}^u}\frac{x_k}{x_u}\hat{\gamma}_k - \left(\frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\|\sigma\|}x_c}{1-x_0-x_c}\right) - \hat{\gamma}_j < \frac{\hat{\sigma}}{\|\sigma\|}. \qquad \text{(C.36)}
$$

Step 3: Aggregate risk aversion.

The aggregate risk aversion can be written as

$$
\gamma_u(X,\epsilon) = \gamma\left[1 + \mathbb{E}^u\left[\hat{\gamma}_j\right]\epsilon - \mathrm{Var}^u\left[\hat{\gamma}_j\right]\epsilon^2\right] + O(\epsilon^3), \qquad \text{(C.37)}
$$

where

$$
\mathbb{E}^u\left[\hat{\gamma}_j\right] \equiv \sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j, \quad \text{and} \quad \mathrm{Var}^u[\hat{\gamma}_j] \equiv \sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j^2 - \left(\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j\right)^2.
$$

Step 4: Market price of risk.

The market price of risk is given by

$$
\eta(X,\epsilon) = \frac{\gamma_u(X,\epsilon)}{x_u}\left[(1-(1+\hat{\alpha}_p\epsilon)x_{0,t})(\sigma-\sigma_y(X,\epsilon)) - (\sigma+\hat{\sigma}\epsilon)x_{c,t} + \sum_{j\in\mathcal{J}_t^u}x_{j,t}\frac{1-\gamma_j^{-1}}{1-\psi}\sigma_{\xi_j}(X,\epsilon)\right], \qquad \text{(C.38)}
$$

The second-order term is then given by

$$\eta_2(X) = \frac{\gamma_{u,0}(X)}{x_u}\left[-(1-x_{0,t})\sigma_{y,2}(X) + \sum_{j\in\mathcal{J}_t^u} x_{j,t}\frac{1-\gamma_j^{-1}}{1-\psi}\sigma_{\xi_j,2}(X)\right] + \\ + \frac{\gamma_{u,1}(X)}{x_u}\left[-\hat{\alpha}_p x_0\sigma - \hat{\sigma}x_c\right] + \frac{\gamma_{u,2}(X)}{x_u}\left[(1-x_0)\sigma - \sigma x_c\right]. \tag{C.39}$$

The expression above can be written as

$$\eta_2(X) = -\frac{(\gamma-1)x_c + 1 - x_0}{1-x_0-x_c}\sigma_{y,2}(X) - \gamma\sigma\mathbb{E}^u[\hat{\gamma}_j]\frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c} - \gamma\sigma\text{Var}^u[\hat{\gamma}_j]. \tag{C.40}$$

Step 5: Interest rate.
The interest rate is given by

$$r(X,\epsilon) = \rho - \eta(X,\epsilon)\sigma_R(X,\epsilon) + \psi^{-1}(\mu - \mu_y(X,\epsilon) + \sigma_{R,t}^2(X,\epsilon) - \sigma\sigma_R(X,\epsilon)) \\ + \left(1-\psi^{-1}\right)\sum_{j=0}^J x_j\frac{\gamma_j}{2}\sigma_j^2(X,\epsilon) \\ + \psi^{-1}\sum_{j=0}^J x_{j,t}\left[\mu_{\xi_j}(X,\epsilon) + (1-\gamma_j)\sigma_{\xi_j}(X,\epsilon)\sigma_j(X,\epsilon) + \frac{\psi-\gamma_j}{1-\psi}\frac{\sigma_{\xi_j,t}^2(X,\epsilon)}{2}\right]. \tag{C.41}$$

The second-order term is given by

$$r_2(X) = -\eta_2(X)\sigma + \eta_0(X)\sigma_{y,2}(X) - \psi^{-1}(\mu_{y,2}(X) + \sigma\sigma_{y,2}(X)) \\ + \left(1-\psi^{-1}\right)\sum_{j=0}^J x_j\gamma\sigma\left[\sigma_{j,2}(X) + \frac{\sigma_{j,1}^2(X)}{2\sigma} + \hat{\gamma}_j\sigma_{j,1}\right] \\ + \psi^{-1}\sum_{j=0}^J x_j\left[\mu_{\xi_j,2}(X) + (1-\gamma)\sigma_{\xi_j,2}(X)\sigma\right]. \tag{C.42}$$

Step 6: Dividend yield.
The dividend yield, y, is given by

$$y(X,\epsilon) = \psi\rho + (1-\psi)\left[r(X,\epsilon) + \eta(X,\epsilon)\sigma_R(X,\epsilon) - \sum_{j=0}^J x_j\frac{\gamma_j}{2}\sigma_j^2(X,\epsilon)\right] \\ + \sum_{j=0}^J x_j\left[\mu_{\xi_j}(X,\epsilon) + (1-\gamma_j)\sigma_{\xi_j}(X,\epsilon)\sigma_j(X,\epsilon) + \frac{\psi-\gamma_j}{1-\psi}\frac{\sigma_{\xi_j}^2(X,\epsilon)}{2}\right]. \tag{C.43}$$

The second-order term in the expansion of $y(X,\epsilon)$ is given by

$$y_2(X) = (1-\psi)\left[r_2(X) + \eta_2(X)\sigma + \eta_0(X)\sigma_{R,2}(X) - \sum_{j=0}^{J} x_j \frac{\gamma}{2}\left(\sigma_{j,1}^2(X) + 2\sigma\sigma_{j,2}(X) + 2\hat{\gamma}_j\sigma\sigma_{j,1}(X)\right)\right]$$
$$+ \sum_{j=0}^{J} x_j\left[\mu_{\xi_j,2}(X) + (1-\gamma)\sigma_{\xi_j,2}(X)\sigma\right]. \tag{C.44}$$

Using the expression for the interest rate, we obtain

$$y_2(X) = \left(1-\psi^{-1}\right)\left[\mu_{y,2}(X) + \sigma\sigma_{y,2}(X) + \sum_{j=0}^{J} x_j\gamma\sigma\left(\frac{\sigma_{j,1}^2(X)}{2\sigma} + \sigma_{j,2}(X) + \hat{\gamma}_j\sigma_{j,1}(X)\right)\right]$$
$$+ \psi^{-1}\sum_{j=0}^{J} x_j\left[\mu_{\xi_j,2}(X) + (1-\gamma)\sigma_{\xi_j,2}(X)\sigma\right]. \tag{C.45}$$

Given that $\mu_{\xi_j,2} = (1-\psi)\mu_{y,2}$ and $\sigma_{\xi_j,2} = (1-\psi)\sigma_{y,2}$, we obtain

$$\begin{aligned} y_2(X) &= \left(1-\psi^{-1}\right)\left[\gamma\sigma_{y,2}(X)\sigma + \sum_{j=0}^{J} x_j\gamma\sigma\left(\frac{\sigma_{j,1}^2(X)}{2\sigma} + \sigma_{j,2}(X) + \hat{\gamma}_j\sigma_{j,1}(X)\right)\right] \\ &= \left(1-\psi^{-1}\right)\sum_{j=0}^{J} x_j\gamma\sigma\left(\frac{\sigma_{j,1}^2(X)}{2\sigma} + \hat{\gamma}_j\sigma_{j,1}(X)\right), \end{aligned} \tag{C.46}$$

where in the second equality, we use the fact that, from the market clearing condition for the risky asset, we have $\sum_{j=0}^{J} x_j\sigma_{j,2}(X) = \sigma_{R,2}(X) = -\sigma_{y,2}(X)$.

Step 7: Risk premium.

Since the risk premium is given by $\pi_t = \eta_t\sigma_{R,t}$, we can write

$$\pi(X,\epsilon) = \eta_0(X)\sigma_{R,0}(X) + \eta_1(X)\sigma_{R,0}(X)\epsilon + \left(\eta_0(X)\sigma_{R,2}(X) + \eta_2(X)\sigma_{R,0}(X)\right)\epsilon^2 + O\left(\epsilon^3\right).$$

Given that $\sigma_{R,0} = \sigma$ and $\sigma_{R,2}(X) = -\sigma_{y,2}(X)$, we get

$$\pi_2(X) = -\eta_0(X)\sigma_{y,2}(X) + \eta_2(X)\sigma. \tag{C.47}$$

□

D Derivation of the Market Elasticity

Let $p(X,\epsilon) \equiv 1/y(X,\epsilon)$ denote the price-dividend ratio. The second-order expansion of $p(X,\epsilon)$ is given by

$$p(X,\epsilon) = \frac{1}{y_0(X)} - \frac{y_1(X)}{y_0^2(X)}\epsilon + \left[\frac{y_1^2(X)}{y_0^3(X)} - \frac{y_2(X)}{y_0^2(X)}\right]\epsilon^2 + O(\epsilon^3). \tag{D.1}$$

Let $F(X) \equiv \frac{W_0(1+\hat{\alpha}_p\epsilon)-W_0}{P} = \hat{\alpha}_p\epsilon x_0$ denote the flow into the risky asset relative to the benchmark economy.

Proof. From Equation (D.1), the first-order impact of flows on the price-dividend ratio can be written as

$$\frac{\partial p(X,\epsilon)}{\partial F(X,\epsilon)} = -\frac{1}{y_0^2(X)}\frac{\partial y_1(X)}{\partial \hat{\alpha}_p}\frac{1}{x_0} + O(\epsilon). \tag{D.2}$$

Since from Proposition 2, $y_1(X)$ does not depend on $\hat{\alpha}_p$, the right hand side of Equation (D.2) is zero, leading to an infinite aggregate elasticity to the first-order:

$$\varepsilon_M^{-1} = \frac{1}{p(X,\epsilon)}\frac{\partial p(X,\epsilon)}{\partial F(X,\epsilon)} = 0 + O(\epsilon).$$

Given that from Equation (17), we have $y_t = r_t + \pi_t - \mu_{P,t}$, we can write the first-order term for the dividend yield as

$$y_1(X) = r_1(X) + \pi_1(X) - \mu_{P,1}(X),$$

where $\mu_{P,t} = \mu - \mu_{y,t} - \sigma_{y,t}\sigma_{R,t}$ is the drift of the risky asset price P_t, and π_t is the risk premium. As shown in Proposition 2, the first-order term for the dividend yield is constant. This means $\mu_{y,1}(X) = \sigma_{y,1}(X) = 0$, leading to $\sigma_{R,1}(X) = 0$. Therefore, we have $\mu_{P,1}(X) = 0$, and

$$\frac{\partial y_1(X)}{\partial \hat{\alpha}_p} = \frac{\partial r_1(X)}{\partial \hat{\alpha}_p} + \frac{\partial \pi_1(X)}{\partial \hat{\alpha}_p}.$$

From Proposition 2, we have

$$\begin{aligned}
\frac{\partial r_1(X)}{\partial \hat{\alpha}_p} &= -\sigma\frac{\partial \eta_1(X)}{\partial \hat{\alpha}_p} = \frac{\gamma\sigma^2}{x_u}x_0, \\
\frac{\partial \pi_1(X)}{\partial \hat{\alpha}_p} &= \sigma\frac{\partial \eta_1(X)}{\partial \hat{\alpha}_p} = -\frac{\gamma\sigma^2}{x_u}x_0.
\end{aligned}$$

Thus, up to the first order, the effect of portfolio flows on the risk-free rate is the exact opposite of its impact on the risk premium and portfolio flows do not affect the price-dividend ratio up to first order. Therefore, up to the first-order, the aggregate market elasticity is infinite. $\square$

Using Equation (D.1), the derivative of the price-dividend ratio with respect to flows F is given by

$$\frac{\partial p(X,\epsilon)}{\partial F(X,\epsilon)} = -\frac{1}{y_0^2(X)}\frac{\partial y_2(X)}{\partial \hat{\alpha}_p}\frac{\epsilon}{x_0} + O(\epsilon^2), \tag{D.3}$$

where $y_1(X)$ does not depend on $\hat{\alpha}_p$, leading to no price impact (infinite elasticity) up to the first-order.

D.1 Proof of Proposition 3

Proof. Consider the case where there is no preference heterogeneity and active investors do not face leverage constraints. In this case, $y_2(X)$ simplifies to

$$y_2(X) = \left(1-\psi^{-1}\right)\gamma\sigma^2\sum_{j=0}^{J} x_j \frac{\sigma_{j,1}^2(X)}{2\sigma^2}, \tag{D.4}$$

using the fact that $\sigma_{y,2} = \sigma_{j,2} = 0$ when $\hat{\gamma}_j = 0$ for $j = 0, 1, \ldots, J$ when investors have the same preferences.

Using the expression for $\sigma_{j,1}$ in Proposition 2, the (inverse) aggregate market elasticity is given by

$$\frac{1}{p(X,\epsilon)}\frac{\partial p(X,\epsilon)}{\partial F(X,\epsilon)} = -\frac{1-\psi^{-1}}{2y_0(X)}\gamma\sigma^2\left(2x_0\hat{\alpha}_p + x_a\frac{2\hat{\alpha}_p x_0^2}{x_a^2}\right)\frac{\epsilon}{x_0} + O(\epsilon^2), \tag{D.5}$$

where $x_a \equiv 1 - x_0$ denotes the wealth share of active investors. This can be written as

$$\frac{1}{p(X,\epsilon)}\frac{\partial p(X,\epsilon)}{\partial F(X,\epsilon)} = \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{y_0(X)}\frac{1-\overline{\alpha}_p}{x_a} + O(\epsilon^2), \tag{D.6}$$

where we use $\overline{\alpha}_p = 1 + \hat{\alpha}_p\epsilon$ from Equation (25). $\square$

From Equation (C.47), the impact of flows of the risk premium can be written as

$$\frac{\partial \pi(X,\epsilon)}{\partial F(X,\epsilon)} = \sigma\frac{\partial \eta_1(X)}{\partial \hat{\alpha}_p}\frac{\epsilon}{x_0} + \frac{\partial \pi_2(X)}{\partial \hat{\alpha}_p}\frac{\epsilon}{x_0}$$

D.2 Proof of Proposition 4

Proof. Consider the case in which active investors have heterogeneous risk aversions, but face no leverage constraints. In this case, the expression for $y_2(X)$ in Equation (C.30) can be written as

$$y_2(X) = \left(1-\psi^{-1}\right)\gamma\sigma^2\left[\sum_{j=1}^{J} x_j\left(\frac{\sigma_{j,1}^2(X)}{2\sigma^2}\hat{\gamma}_j\frac{\sigma_{j,1}(X)}{\sigma}\right)\right] + \left(1-\psi^{-1}\right)\gamma\sigma^2 x_0\left(\frac{\hat{\alpha}_p^2}{2}+\hat{\gamma}_0\hat{\alpha}_p\right) \tag{D.7}$$

Note that with unconstrained active investors, the effect of endogenous volatility and hedging demand exactly cancel out. We first compute the derivative of the term involving $\sigma_{j,1}^2$:

$$\begin{aligned}
\sum_{j=0}^{J} x_j\frac{\partial\sigma_{j,1}^2}{\partial\hat{\alpha}_p} &= \sum_{j=0}^{J} 2x_j\sigma_{j,1}\frac{\partial\sigma_{j,1}}{\partial\hat{\alpha}_p} \\
&= 2x_0\sigma^2\hat{\alpha}_p + 2\sigma^2\sum_{j=1}^{J} x_j\left(\sum_{k\in\mathcal{J}^u}\frac{x_k}{x_u}\hat{\gamma}_k - \left(\frac{\hat{\alpha}_p x_0}{1-x_0}\right) - \hat{\gamma}_j\right)\left(-\frac{x_0}{1-x_0}\right) \\
&= 2\sigma^2\left[x_0\hat{\alpha}_p + \frac{\hat{\alpha}_p x_0^2}{1-x_0} + \left((1-x_0)\sum_{k=1}^{J}\frac{x_k}{x_u}\hat{\gamma}_k - \sum_{j=1}^{J}x_j\hat{\gamma}_j\right)\left(-\frac{x_0}{1-x_0}\right)\right] \\
&= 2\sigma^2\frac{\hat{\alpha}_p x_0}{1-x_0}.
\end{aligned} \tag{D.8}$$

The derivatives of the term involving $\sigma_{j,1}(X)$ with respect to $\hat{\alpha}_p$ are given by:

$$\frac{1}{\sigma}\sum_{j=0}^{J} x_j\hat{\gamma}_j\frac{\partial\sigma_{j,1}(X)}{\partial\hat{\alpha}_p} = \sum_{j=1}^{J} x_j\hat{\gamma}_j\left(-\frac{x_0}{1-x_0}\right) + x_0\hat{\gamma}_0 = x_0\left(\hat{\gamma}_0 - \mathbb{E}^u[\hat{\gamma}_k]\right) \tag{D.9}$$

Thus, the (inverse) aggregate elasticity is then given by:

$$\frac{1}{p(X,\epsilon)}\frac{\partial p(X,\epsilon)}{\partial F(X,\epsilon)} = \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{y_0(X)}\left[\frac{1-\overline{\alpha}_p}{x_a} - \frac{\gamma_0 - \mathbb{E}^u\left[\gamma_j\right]}{\gamma}\right] + O(\epsilon^2), \tag{D.10}$$

where $x_a \equiv 1-x_0$ is the wealth share of active investors, and we use $\gamma_j = \gamma\left(1+\hat{\gamma}_j\epsilon\right)$ from Equation (24). □

D.3 Proof of Proposition 5

Proof. Consider the case in which active investors have heterogeneous risk aversions and also face leverage constraints. We can then write $y_2(X)$ as follows

$$
\begin{aligned}
y_2(X) &= \left(1-\psi^{-1}\right)\gamma\sigma^2\sum_{j=0}^{J}x_j\left[\frac{\sigma_{j,1}^2(X)}{2\sigma^2}+\hat{\gamma}_j\frac{\sigma_{j,1}(X)}{\sigma}\right] \\
&= \left(1-\psi^{-1}\right)\gamma\sigma^2\sum_{j=1}^{J}x_j\left[\frac{\sigma_{j,1}^2(X)}{2\sigma^2}+\hat{\gamma}_j\min\left\{\sum_{k\in\mathcal{J}^u}\frac{x_k}{x_u}\hat{\gamma}_k-\frac{\hat{\alpha}_p x_0+\frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c}-\hat{\gamma}_j,\frac{\hat{\sigma}}{\sigma}\right\}\right] \\
&+\left(1-\psi^{-1}\right)\gamma\sigma^2 x_0\left[\frac{\hat{\alpha}_p^2}{2}+\hat{\gamma}_0\hat{\alpha}_p\right],
\end{aligned}
\tag{D.11}
$$

The derivative of $y_2(X)$ with respect to $\hat{\alpha}_p$ is given by

$$
\frac{\partial y_2(X)}{\partial\hat{\alpha}_p}=\left(1-\psi^{-1}\right)\gamma\sigma^2\left(\sum_{j=1}^{J}x_j\frac{\sigma_{j,1}(X)}{\sigma^2}\frac{\partial\sigma_{j,1}}{\partial\hat{\alpha}_p}-\mathbb{E}^u[\hat{\gamma}_k]x_0+x_0(\hat{\alpha}_p+\hat{\gamma}_0)\right),
$$

where

$$
\begin{aligned}
\frac{1}{\sigma^2}\sum_{j=0}^{J}x_j\frac{\partial\sigma_{j,1}^2}{\partial\hat{\alpha}_p} &= \frac{1}{\sigma^2}\sum_{j=0}^{J}x_j 2\sigma_{j,1}\frac{\partial\sigma_{j,1}}{\partial\hat{\alpha}_p} \\
&= 2x_0\hat{\alpha}_p+2\sum_{j\in\mathcal{J}^u}x_j\left(\sum_{k\in\mathcal{J}^u}\frac{x_k}{x_u}\hat{\gamma}_k-\left(\frac{\hat{\alpha}_p x_0+\frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c}\right)-\hat{\gamma}_j\right)\left(-\frac{x_0}{1-x_0-x_c}\right) \\
&= 2\left[x_0\hat{\alpha}_p+\frac{\hat{\alpha}_p x_0+\frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c}x_0+\left(x_u\sum_{k\in\mathcal{J}^u}\frac{x_k}{x_u}\hat{\gamma}_k-\sum_{j\in\mathcal{J}^u}x_j\hat{\gamma}_j\right)\left(-\frac{x_0}{1-x_0-x_c}\right)\right] \\
&= 2x_0\left[\frac{\hat{\alpha}_p}{x_u}+\left(\frac{\hat{\sigma}}{\sigma}-\hat{\alpha}_p\right)\frac{x_c}{x_u}\right].
\end{aligned}
\tag{D.12}
$$

The second derivative of $y(X,\epsilon)$ can then be written as

$$
\frac{\partial y_2(X)}{\partial\hat{\alpha}_p}=\left(1-\psi^{-1}\right)\gamma\sigma^2\left[\frac{\hat{\alpha}_p}{x_u}+\left(\frac{\hat{\sigma}}{\sigma}-\hat{\alpha}_p\right)\frac{x_c}{x_u}-\left(\mathbb{E}^u[\hat{\gamma}_k]-\hat{\gamma}_0\right)\right]x_0.
$$

Thus, the (inverse) aggregate market elasticity is then given by

$$\frac{1}{p(X,\epsilon)}\frac{\partial p(X,\epsilon)}{\partial F(X,\epsilon)} = \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{y_0(X)}\left[\frac{(1-\overline{\alpha}_p)(1-x_c)-\left(\frac{\overline{\sigma}}{\sigma}-1\right)x_c}{x_u} - \frac{\gamma_0 - \mathbb{E}^u\left[\gamma_j\right]}{\gamma}\right] + \mathcal{O}(\epsilon^2), \tag{D.13}$$

where we use $\overline{\sigma} = \sigma + \hat{\sigma}\epsilon$ from Equation (26). □

E Useful Formula

The following are useful for computing the derivatives above:

$$\gamma\sigma\frac{\partial\sigma_{y,2}(X)}{\partial\hat{\alpha}_p} = \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2y_0(X)}\sum_{k=1}^{J}\left(\hat{\gamma}_k - \hat{\gamma}_0\right)x_k\left(-\frac{\gamma\sigma^2 x_0}{1-x_0}\right). \tag{E.1}$$

Note that we can write $\sigma_{j,2}(X)$ as follows

$$\begin{aligned}
\frac{\sigma_{j,2}(X)}{\sigma} &= \frac{\eta_2(X)}{\gamma\sigma} - \frac{\eta_1(X)}{\gamma\sigma}\hat{\gamma}_j + \frac{\eta_0(X)}{\gamma\sigma}\hat{\gamma}_j^2 - (1-\gamma^{-1})\frac{\sigma_{y,2}(X)}{\sigma} \\
&= -\left(1+\frac{x_c}{x_u}\right)\frac{\sigma_{y,2}(X)}{\sigma} - \mathbb{E}^u[\hat{\gamma}_k]\frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c} - \mathrm{Var}^u[\hat{\gamma}_k] + \\
&- \hat{\gamma}_j\left[\sum_{k\in\mathcal{J}^u}\frac{x_k}{x_u}\hat{\gamma}_j - \frac{\hat{\alpha}_p x_0}{1-x_0}\right] + \hat{\gamma}_j^2.
\end{aligned} \tag{E.2}$$

$$\sigma_{j,1}(X) = \sigma\min\left\{\sum_{k\in\mathcal{J}^u}\frac{x_k}{x_u}\hat{\gamma}_k - \left(\frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\|\sigma\|}x_c}{1-x_0-x_c}\right) - \hat{\gamma}_j, \frac{\hat{\sigma}}{\sigma}\right\} \tag{E.3}$$

$$\sigma_{j,2}(X) = \frac{\eta_2(X)}{\gamma} - \frac{\eta_1(X)}{\gamma}\hat{\gamma}_j + \frac{\eta_0(X)}{\gamma}\hat{\gamma}_j^2 - (1-\gamma^{-1})\sigma_{y,2}(X) \tag{E.4}$$

$$\sigma_{y,2}(X) = \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2y_0(X)}\sum_{k=1}^{J}\left(\hat{\gamma}_k-\hat{\gamma}_0\right)x_k\sigma_{k,1}(X) \tag{E.5}$$

$$\eta_1(X) = \gamma\sigma\left[\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j - \frac{\hat{\alpha}_p x_0}{1-x_0}\right] \tag{E.6}$$

$$\eta_2(X) = -\frac{(\gamma-1)x_c + 1 - x_0}{1-x_0-x_c}\sigma_{y,2}(X) - \gamma\sigma\mathbb{E}^u[\hat{\gamma}_j]\frac{\hat{\alpha}_p x_0 + \frac{\hat{\sigma}}{\sigma}x_c}{1-x_0-x_c} - \gamma\sigma\mathrm{Var}^u[\hat{\gamma}_j], \tag{E.7}$$

where

$$
\begin{aligned}
\frac{\partial \sigma_{y,2}}{\partial \hat{\alpha}_p} &= \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2y_0(X)}\sum_{k=1}^{J}\left(\hat{\gamma}_k-\hat{\gamma}_0\right)x_k\left(-\frac{\sigma x_0}{1-x_0}\right)\\
&= \left(1-\psi^{-1}\right)\frac{\gamma\sigma^2}{2y_0(X)}\sigma\left(\hat{\gamma}_0-\mathbb{E}^u[\hat{\gamma}_j]\right)x_0 \qquad \text{(E.8)}\\
\sum_{j=0}^{J}x_j\gamma\sigma\hat{\gamma}_j\frac{\partial \sigma_{j,1}(X)}{\partial \hat{\alpha}_p} &= \gamma\sigma^2\sum_{j=0}^{J}x_j\hat{\gamma}_j\left(-\frac{x_0}{1-x_0}\right) \qquad \text{(E.9)}
\end{aligned}
$$

F Derivation of the perturbed solution

In this section, we compute the first-order and second-order correction of the equilibrium objects. It turns out that the system of equations determining the perturbed solution is block-recursive, so we are able to solve for the equilibrium objects one by one, provided we proceed in the appropriate order.

In contrast to the case considered in the text, we allow for portfolio-flow shocks. In particular, we assume that the portfolio share of the passive investor is given by $\alpha_{0,t} = 1+\epsilon(\overline{\alpha}_{p,t}-1)$, where $\overline{\alpha}_{p,t}$ follows the process

$$
d\overline{\alpha}_{p,t} = \theta_p(\overline{\alpha}-\overline{\alpha}_{p,t})\epsilon dt + \sigma_p\sqrt{\overline{\alpha}_{p,t}}\epsilon dZ_t. \qquad \text{(F.1)}
$$

Notice that $\alpha_{0,t} = 1$ and $\overline{\alpha}_{p,t}$ is constant when $\epsilon = 0$. Finally, we assume that the mortality parameter is given by $\kappa = \hat{\kappa}\epsilon$.

F.1 First-order correction

Diffusion and drift terms. The diffusion term for the price-dividend ratio is given by

$$
\sigma_{p,t} = \frac{p_x}{p}\sigma_x + \frac{p_{\overline{\alpha}_p}}{p}\sigma_p\sqrt{\overline{\alpha}_{p,t}}\epsilon = O(\epsilon^2). \qquad \text{(F.2)}
$$

Notice that $p_{x_j} = O(\epsilon)$ and $\sigma_{x_j} = O(\epsilon)$, as p and x_j are constant when $\epsilon = 0$, so the zeroth-order terms for p_{x_j}, $p_{\overline{\alpha}_p}$, and σ_{x_j} are equal to zero. This implies that the first-order correction for $\sigma_{p,t}$ is equal to zero. A similar argument shows that $\sigma_{c_j,t} = O(\epsilon^2)$.

The drift of p is given by

$$
\mu_{p,t} = \frac{p_x}{p}\mu_x + \frac{p_{\overline{\alpha}_p}}{p}\theta_p(\overline{\alpha}-\overline{\alpha}_{p,t})\epsilon + \frac{1}{2}\sum_{k=1}^{d}\left[\sigma_{x,k}'\frac{p_{xx}}{p}\sigma_{x,k} + 2\sigma_{p,k}\sqrt{\overline{\alpha}_{p,t}}\epsilon\frac{p_{x\alpha_p}}{p}\sigma_{x,k} + \frac{p_{\overline{\alpha}_p\overline{\alpha}_p}}{p}\sigma_{p,k}^2\overline{\alpha}_{p,t}\epsilon^2\right], \qquad \text{(F.3)}
$$

where $\sigma_{x,k}$ is the k-th column of the $J \times d$ matrix σ_x. As p_x and $p_{\overline{\alpha}_p}$ are first-order in ϵ, and the same goes for μ_x and σ_x, then $\mu_{p,t} = O(\epsilon^2)$. A similar argument shows that $\mu_{c_j,t} = O(\epsilon^2)$. Notice these facts imply that $\varsigma_j = O(\epsilon^2)$ and $\xi_{j,t} = O(\epsilon^2)$.

Risk premium. The risk premium is given by

$$\pi_1(X) = \gamma||\sigma||^2 \left[\sum_{j \in \mathcal{J}^u} \frac{x_j}{x_u} \hat{\gamma}_j - \frac{x_0 \hat{\alpha}_p + x_c \frac{\hat{\sigma}}{||\sigma||}}{1 - x_0 - x_c} \right], \tag{F.4}$$

using the fact that $||\sigma_{R,t}|| = ||\sigma|| + O(\epsilon^2)$, and $\hat{\alpha}_{p,t} \equiv \overline{\alpha}_{p,t} - 1$.

Portfolio share. The portfolio share of an unconstrained investor is given by

$$\alpha_j(X, \epsilon) = 1 + \left[\frac{\pi_1(X)}{\gamma||\sigma||^2} - \hat{\gamma}_j \right] \epsilon + O(\epsilon^2), \tag{F.5}$$

the portfolio share of a constrained investor is given by

$$\alpha_j(X, \epsilon) = 1 + \frac{\hat{\sigma}}{||\sigma||} \epsilon + O(\epsilon^2), \tag{F.6}$$

and the portfolio share of the passive investor is given by $\alpha_0(X, \epsilon) = 1 + \hat{\alpha}_{p,t}\epsilon$. Notice that $\sum_{j=0}^{J} x_j \alpha_{j,1}(X) = 0$, consistent with market clearing.

Interest rate. The first-order correction for the interest rate is given by

$$r_1(X) = \left(1 - \psi^{-1}\right) \gamma||\sigma||^2 \sum_{j=0}^{J} x_j \left[\frac{\hat{\gamma}_j}{2} + \alpha_{j,1}(X) \right] - \pi_1(X), \tag{F.7}$$

using the fact that $\xi_t = O(\epsilon^2)$, $\mu_{p,t} = O(\epsilon^2)$, and $\sigma_{p,t} = O(\epsilon^2)$. Given the market clearing for the risky asset, we can write:

$$r_1(X) = \left(1 - \psi^{-1}\right) \gamma||\sigma||^2 \sum_{j=0}^{J} x_j \frac{\hat{\gamma}_j}{2} - \pi_1(X), \tag{F.8}$$

so $r_1(X) + \pi_1(X)$ is independent of $\overline{\alpha}_p$.

Price-dividend ratio. From the pricing condition, we obtain

$$-\frac{1}{p_0(X)^2} p_1(X) = r_1(X) + \pi_1(X). \tag{F.9}$$

Rearranging the expression above, and using the expression for the interest rate, we obtain

$$p_1(X) = -p_0(X)^2 \left(1 - \psi^{-1}\right) \gamma ||\sigma||^2 \sum_{j=0}^{J} x_j \frac{\hat{\gamma}_j}{2}, \tag{F.10}$$

which is independent of $\overline{\alpha}_{p,t}$.

Consumption-wealth ratio. The consumption-wealth ratio is given by

$$c_{j,1}(X) = (1 - \psi) \left[r_1(X) + \pi_1(X) + \pi_0(X)\alpha_{j,1}(X) - \frac{1}{2}\gamma |||\sigma|||^2 (\hat{\gamma}_j + 2\alpha_{j,1}(X)) \right]. \tag{F.11}$$

Using the expression for $r_1(X)$, we can write the expression above as follows:

$$c_{j,1}(X) = (1 - \psi) \left[(1 - \psi^{-1}) \sum_{i=0}^{J} x_i \frac{\hat{\gamma}_i}{2} - \frac{\hat{\gamma}_j}{2} \right] \gamma |||\sigma|||^2. \tag{F.12}$$

Wealth dynamics. The diffusion term of x_j is given by

$$\sigma_{x_j}(X) = x_j \alpha_{j,1}(X) \epsilon \sigma + O(\epsilon^2). \tag{F.13}$$

The drift of x_j is given by

$$\mu_{x_j}(X) = x_j \left[r_1(X) + \pi_1(X) + \pi_0(X)\alpha_{j,1}(X) - c_{j,1}(X) - \alpha_{j,1}(X)||\sigma||^2 + \hat{\kappa}\frac{\omega_j - x_j}{x_j} \right] \epsilon + O(\epsilon^2). \tag{F.14}$$

We can write the first-order correction of μ_{x_j} as follows:

$$\mu_{x_j,1}(X) = x_j \left[(\psi - 1)\frac{\gamma |||\sigma|||^2}{2} \left(\sum_{i=0}^{J} x_i \hat{\gamma}_i - \hat{\gamma}_j \right) + (\gamma - 1)|||\sigma|||^2 \alpha_{j,1}(X) \right] + \hat{\kappa}(\omega_j - x_j). \tag{F.15}$$

F.2 Second-order correction

Diffusion and drift terms. The diffusion term for the price-dividend ratio is given by

$$\sigma_{p,2}(X) = \frac{p_{x,1}(X)}{p_0(X)} \sigma_{x,1}(X) + \frac{p_{\overline{\alpha}_p,1}(X)}{p_0(X)} \sigma_p \sqrt{\overline{\alpha}_p}. \tag{F.16}$$

We can write the expression above as follows:

$$\sigma_{p,2}(X) = -p_0(X)(1-\psi^{-1})\frac{\gamma\|\sigma\||^2}{2}\sum_{j=1}^{J}\left(\hat{\gamma}_j - \hat{\gamma}_0\right)x_j\alpha_{j,1}(X)\sigma, \tag{F.17}$$

where we used the fact that $p_{\overline{\alpha}_p,1}(X) = 0$.

Similarly, the diffusion for c_j is given by

$$\sigma_{c_j,2}(X) = -(\psi-1)\frac{\gamma\||\sigma\||^2}{2c_{j,0}(X)}(1-\psi^{-1})\sum_{i=1}^{J}(\hat{\gamma}_i - \hat{\gamma}_0)\,x_i\alpha_{i,1}(X)\sigma. \tag{F.18}$$

The second-order correction for the hedging demand is then given by $\varsigma_{j,2}(X) = \frac{1-\gamma^{-1}}{\psi-1}\frac{\sigma_{c_j,2}\sigma'}{\||\sigma\||^2}$.

The second-order correction of the drift of p and c_j are given by

$$\mu_{p,2}(X) = \frac{p_{x,1}(X)}{p_0(X)}\mu_{x,1}(X), \qquad \mu_{c_j,2}(X) = \frac{c_{j,x,1}(X)}{c_{j,0}(X)}\mu_{x,1}(X), \tag{F.19}$$

which can be written as

$$\mu_{p,2}(X) = -p_0(X)(1-\psi^{-1})\frac{\gamma\|\sigma\||^2}{2}\sum_{j=1}^{J}(\hat{\gamma}_j - \hat{\gamma}_0)\mu_{x_j,1}(X) \tag{F.20}$$

$$\mu_{c_i,2}(X) = (1-\psi)(1-\psi^{-1})\frac{\gamma\|\sigma\||^2}{2c_{i,0}(X)}\sum_{j=1}^{J}(\hat{\gamma}_j - \hat{\gamma}_0)\mu_{x_j,1}(X). \tag{F.21}$$

The second-order correction for $\xi_{j,t}$ is then given by $\xi_{j,2}(X) = \mu_{c_j,2}(X) + (1-\gamma)\sigma_{c_j,2}\sigma'$.

Risk premium. The risk premium is given by

$$\pi_2(X) = \gamma\|\sigma\||^2\left[\frac{\gamma_{u,2}(X)}{\gamma} - \frac{\gamma_{u,1}(X)}{\gamma}\frac{x_0\hat{\alpha}_p + x_c\frac{\hat{\sigma}}{\|\sigma\||}}{1-x_0-x_c} + 2\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}}{||\sigma||^2} - \frac{x_c}{1-x_0-x_c}\overline{\alpha}_{c,2}(X) - \varsigma_2(X)\right]. \tag{F.22}$$

Notice that we can write the aggregate risk aversion as follows:

$$\gamma_u(X) = \gamma\left[1 + \mathbb{E}^u[\hat{\gamma}_j]\epsilon - \delta^u[\hat{\gamma}_j]\epsilon^2\right] + O(\epsilon^3), \tag{F.23}$$

where $\mathbb{E}^u[\hat{\gamma}_j] \equiv \sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j$ and $\delta^u[\hat{\gamma}_j] \equiv \sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j^2 - \left(\sum_{j\in\mathcal{J}^u}\frac{x_j}{x_u}\hat{\gamma}_j\right)^2$, so $\gamma_{u,2}(X)/\gamma = -\delta^u[\hat{\gamma}_j]$.

Combining the previous two expressions, we obtain

$$
\begin{aligned}
\frac{\pi_2(X)}{\gamma\|\sigma\||^2} &= -\delta^u[\hat{\gamma}_j] - \mathbb{E}^u[\hat{\gamma}_j]\frac{x_0\hat{\alpha}_p + x_c\frac{\hat{\sigma}}{\|\sigma\||}}{1-x_0-x_c} + 2\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}}{\|\sigma\||^2} + \frac{x_c}{1-x_0-x_c}\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}}{\|\sigma\||^2} - \varsigma_2(X) \\
&= -\delta^u[\hat{\gamma}_j] - \mathbb{E}^u[\hat{\gamma}_j]\frac{x_0\hat{\alpha}_p + x_c\frac{\hat{\sigma}}{\|\sigma\||}}{1-x_0-x_c} + \left(1+\gamma^{-1}+\frac{x_c}{1-x_0-x_c}\right)\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}}{\|\sigma\||^2},
\end{aligned}
\tag{F.24}
$$

where we used the fact that $\varsigma_2(X) = \frac{1-\gamma^{-1}}{\psi-1}\frac{\sigma_{c_j,2}\sigma'}{\|\sigma\||^2}$, $\sigma_{p,2} = \frac{\sigma_{c_j,2}}{\psi-1}$, and $\overline{\alpha}_{c,2}(X) = -\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}}{\|\sigma\||^2}$.

Portfolio share. The portfolio share of an unconstrained investor is given by

$$
\alpha_{j,2}(X) = \frac{\pi_2(X)}{\gamma\||\sigma\||^2} - \frac{\pi_1(X)}{\gamma\||\sigma\||^2}\hat{\gamma}_j + \hat{\gamma}_j^2 - 2\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}(X)}{\|\sigma\|^2} + \varsigma_{j,2}(X), \tag{F.25}
$$

the portfolio share of a constrained investor is $\alpha_{j,2}(X) = -\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}}{\|\sigma\||^2}$, and the portfolio share of the passive investor satisfies $\alpha_{0,2} = 0$.

We can write the expression above as follows:

$$
\alpha_{j,2}(X) = \frac{\pi_2(X)}{\gamma\||\sigma\||^2} - \frac{\pi_1(X)}{\gamma\||\sigma\||^2}\hat{\gamma}_j + \hat{\gamma}_j^2 - (1+\gamma^{-1})\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}(X)}{\|\sigma\|^2}. \tag{F.26}
$$

Notice that $\sum_{j=0}^{J} x_j\alpha_{j,2}(X) = 0$, consistent with market clearing.

Interest rate. The interest rate is given by

$$
\begin{aligned}
r_2(X) &= \psi^{-1}(\mu_{p,2}(X) + \sigma\sigma_{p,2}(X)') + (1-\psi^{-1}).\frac{\gamma\|\sigma\||^2}{2}\sum_{j=0}^{J}x_j\left[2\hat{\gamma}_j\alpha_{j,1}(X) + \alpha_{j,1}^2(X) + 2\alpha_{j,2}(X)\right] \\
&\quad + (1-\psi^{-1})\gamma\|\sigma\|^2\sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,2,k}(X)}{\|\sigma\|^2} - \pi_2(X) + \psi^{-1}\xi_2(X),
\end{aligned}
\tag{F.27}
$$

where $\xi_2(X) = (\psi-1)\left[\mu_{p,2}(X) + (1-\gamma)\sigma_{p,2}\sigma'\right]$

We can write the expression for the portfolio of the unconstrained investor as follows:

$$
r_2(X) = \mu_{p,2}(X) + \sigma\sigma_{p,2}(X)' + (1-\psi^{-1})\gamma\|\sigma\||^2\sum_{j=0}^{J}x_j\left[\hat{\gamma}_j\alpha_{j,1}(X) + \frac{\alpha_{j,1}^2(X)}{2}\right] - \pi_2(X), \tag{F.28}
$$

Price-dividend ratio. The price-dividend ratio is given by

$$\frac{p_1^2(X)}{p_0^3(X)} - \frac{p_2(X)}{p_0^2(X)} = r_2(X) + \pi_2(X) - \mu_{p,2}(X) - \sigma\sigma_{p,2}'. \tag{F.29}$$

Rearranging the expression above, and using the expression for $r_2(X)$, we obtain

$$\frac{p_2(X)}{p_0(X)} = -p_0(X)(1-\psi^{-1})\gamma\|\sigma\||^2 \sum_{j=0}^{J} x_j \left[\hat{\gamma}_j \alpha_{j,1}(X) + \frac{\alpha_{j,1}^2(X)}{2}\right] + \left(\frac{p_1(X)}{p_0(X)}\right)^2. \tag{F.30}$$

Consumption-wealth ratio. The second-order correction for the consumption-wealth ratio is given by

$$c_{j,2}(X) = (1-\psi)\left[r_2(X) + \pi_2(X) + \pi_1(X)\alpha_{j,1}(X) + \pi_0(X)\alpha_{j,2}(X)\right] + \xi_{j,2}(X) \tag{F.31}$$

$$- (1-\psi)\gamma\|\sigma\||^2\left[\alpha_{j,1}(X)\hat{\gamma}_j + \alpha_{j,2}(X) + \frac{\alpha_{j,1}^2(X)}{2} + \sum_{k=1}^{d}\frac{\sigma_k\sigma_{p,k,2}(X)}{\|\sigma\||^2}\right]. \tag{F.32}$$

Wealth dynamics. The diffusion term of x_j is given by

$$\sigma_{x_j,2}(X) = x_j\alpha_{j,2}\sigma. \tag{F.33}$$

The drift of x_j is given by

$$\mu_{x_j,2} = x_j\left[r_2(X) + \pi_2(X) + \pi_1(X)\alpha_{j,1}(X) + \pi_0(X)\alpha_{j,2}(X) - c_{j,2}(X) - \mu_{p,2}(X) - \sigma\sigma_{p,2}(X)' - \alpha_{j,2}(X)\|\sigma\|^2\right]. \tag{F.34}$$

Aggregate market elasticity. The derivative of p with respect to $\overline{\alpha}_p$ is given by

$$\frac{1}{p(X,\epsilon)}\frac{\partial p(X,\epsilon)}{\partial\overline{\alpha}_p} = \frac{1}{p_0(X)}\frac{\partial p_2(X)}{\partial\overline{\alpha}_p}\epsilon^2 + O(\epsilon^3). \tag{F.35}$$

The market elasticity satisfies the condition

$$\frac{1}{p_0(X)}\frac{\partial p_2(X)}{\partial\overline{\alpha}_p} = -p_0(X)(1-\psi^{-1})\gamma\|\sigma\||^2\left[x_0(\hat{\gamma}_0 + \hat{\alpha}_p) + \sum_{j\in\mathcal{J}^u} x_j\left(\hat{\gamma}_j + \alpha_{j,1}(X)\right)\left(-\frac{x_0}{x_u}\right)\right]\epsilon^2. \tag{F.36}$$

From the market clearing for the risky asset, we have $x_0\hat{\alpha}_p + \sum_{j\in\mathcal{J}^u} x_j\alpha_{j,1}(X) + x_c\frac{\hat{\sigma}}{\|\sigma\|} = 0$, so we can write the expression above as follows:

$$\frac{1}{p_0(X)}\frac{\partial p_2(X)}{\partial \overline{\alpha}_p} = p_0(X)(1-\psi^{-1})\gamma\|\sigma\||^2\left[\hat{\gamma}_u(X)-\hat{\gamma}_0+\frac{1-x_c}{1-x_0-x_c}(1-\overline{\alpha}_p)-\frac{x_c\frac{\hat{\sigma}}{\|\sigma\||}}{1-x_0-x_c}\right]x_0\epsilon^2. \tag{F.37}$$

F.3 Third-order approximation

F.3.1 Passive demand

Suppose there is no preference heterogeneity and no leverage constraint. Without loss of generality, set $J=1$. In this case, the price-dividend ratio is given by

$$p(X,\epsilon) = p^* - (p^*)^2(1-\psi^{-1})\frac{\gamma\|\sigma\|^2}{2}\frac{x_0\hat{\alpha}_p^2}{x_1}\epsilon^2 + O(\epsilon^3) \tag{F.38}$$

$$\pi(X,\epsilon) = \pi_0(X) - \gamma\|\sigma\|^2\frac{x_0\hat{\alpha}_p}{x_1}\epsilon + O(\epsilon^3) \tag{F.39}$$

$$r(X,\epsilon) = r_0(X) + \gamma\|\sigma\|^2\frac{x_0\hat{\alpha}_p}{x_1}\epsilon + (1-\psi^{-1})\frac{\gamma\|\sigma\|^2}{2}\frac{x_0\hat{\alpha}_p^2}{x_1}\epsilon^2 + O(\epsilon^3) \tag{F.40}$$

$$\alpha_0(X,\epsilon) = 1 + \hat{\alpha}_p\epsilon + O(\epsilon^3) \tag{F.41}$$

$$\alpha_1(X,\epsilon) = 1 - \frac{x_0}{x_1}\hat{\alpha}_p\epsilon + O(\epsilon^3) \tag{F.42}$$

$$c_j(X,\epsilon) = c_{j,0}(X) + (1-\psi^{-1})\frac{\gamma\|\sigma\|^2}{2}\frac{x_0\hat{\alpha}_p^2}{x_1}\epsilon^2 + O(\epsilon^3) \tag{F.43}$$

$$\sigma_{x_1}(X,\epsilon) = -\frac{x_0}{x_1}\hat{\alpha}_p\sigma\epsilon + O(\epsilon^3) \tag{F.44}$$

$$\mu_{x_1}(X,\epsilon) = \left[(1-\gamma)\|\sigma\|^2(1-x_1)\hat{\alpha}_p + \hat{\kappa}(\omega_j - x_1)\right]\epsilon + O(\epsilon^3), \tag{F.45}$$

where $\mu_p, \sigma_p, \mu_{c_j}$, and σ_{c_j} are all equal to zero up to second order, and $x_0 = 1 - x_1$.

The law of motion of $x_{1,t}$ can be written as

$$\frac{dx_{1,t}}{x_{1,t}} = \left[x_{0,t}(c_{0,t}-c_{1,t}) + x_{0,t}(\alpha_{0,t}-\alpha_{1,t})\left(\|\sigma_{R,t}\|^2 - \pi_t\right) + \kappa\frac{\omega_1 - x_{1,t}}{x_{1,t}}\right]dt + (\alpha_{1,t}-1)\sigma_{R,t}dZ_t. \tag{F.46}$$

Diffusion and drift terms. The derivatives of $p(X,\epsilon)$ with respect to x_1 and $\overline{\alpha}_p$ are given by

$$\frac{p_{x_1}(X,\epsilon)}{p^*} = (1-\psi^{-1})\frac{\gamma\|\sigma\|^2}{2y^*}\frac{1}{x_1^2}\hat{\alpha}_p^2\epsilon^2 + O(\epsilon^3), \qquad \frac{p_{\overline{\alpha}_p}(X,\epsilon)}{p^*} = -(1-\psi^{-1})\frac{\gamma\|\sigma\|^2}{y^*}\frac{x_0}{x_1}\hat{\alpha}_p\epsilon^2 + O(\epsilon^3) \tag{F.47}$$

The diffusion term for $p(X, \epsilon)$ is then given by

$$\sigma_{p,3}(X) = -(1 - \psi^{-1})\frac{\gamma\|\sigma\|^2}{y^*}\frac{x_0}{x_1}\left[\frac{\hat{\alpha}_p^3}{2x_1^2}\sigma + \hat{\alpha}_p\sigma_p\sqrt{\overline{\alpha}_p}\right]. \tag{F.48}$$

and $\sigma_{c_j,3}(X) = \sigma_{p,3}(X)$. Notice that excess volatility depends on the market elasticity times the volatility of portfolio flows.

The drift of p is given by

$$\mu_{p,3}(X) = (1 - \psi^{-1})\frac{\gamma\|\sigma\|^2}{2y^*}\frac{1}{x_1^2}\hat{\alpha}_p^2\mu_{x_1,1} - (1 - \psi^{-1})\frac{\gamma\|\sigma\|^2}{y^*}\frac{x_0}{x_1}\hat{\alpha}_p\theta_p(\overline{\alpha} - \overline{\alpha}_{p,t}), \tag{F.49}$$

and $\mu_{c_j,3}(X) = \mu_{p,3}(X)$.

Risk premium. The risk premium is given by

$$\pi(X) = \gamma\|\sigma\|^2\left[1 - \frac{x_0(\overline{\alpha}_p - 1)}{x_1} - \frac{1 - \gamma^{-1}}{\psi - 1}\frac{\gamma\|\sigma\|^2}{y^*}\frac{x_0}{x_1}\left[\frac{(1 - \overline{\alpha}_p)^3}{2x_1^2}\sigma + \hat{\alpha}_p\sigma_p\sqrt{\overline{\alpha}_p}\right]\right] \tag{F.50}$$

G Higher-order perturbations

Suppose we have the $(n-1)$-th order perturbation of $c_j(X, \epsilon) = \sum_{k=0}^{n-1} c_{j,k}(X)\epsilon^k$ and the law of motion of X. Let $lc_j(X, \epsilon) = \sum_{k=0}^{n-1} lc_{j,k}(X)\epsilon^k$ denote the expansion of log $c_j(X, \epsilon)$. Then we can compute $\sigma_j(X)$ up to order n:

$$\sigma_{c_j,n}(X) = \sum_{k=1}^{n-1} lc_{j,k,X}(X)\sigma_{X,n-k}(X), \tag{G.1}$$

which is independent of the n-th order term in $c_j(X, \epsilon)$ and $\sigma_X(X, \epsilon)$, as $lc_{j,0,X}(X) = 0$. Similarly, we can compute $\sigma_{y,n}(X)$. A similar argument gives $\mu_{p,n}(X)$ and $\mu_{c_j,n}(X)$. We can then compute $\pi_n(X)$ and $\alpha_{j,n}(X)$. The n-th term of the consumption-wealth ratio satisfies the condition

$$c_{j,t} = \psi\rho + (1 - \psi)\left[\pi_t(\alpha_{j,t} - 1) + \mu + \mu_{p,t} + \sigma\sigma_{p,t}' - \frac{\gamma_j}{2}\|\sigma_{R,t}\|^2\alpha_{j,t}^2 + \sum_{j=0}^{J} x_j c_{j,t}\right] + \xi_{j,t} \tag{G.2}$$

We can rewrite the system above in matrix form as follows:

$$[I - (1 - \psi)\mathbf{1}_{J+1}x_t']c_t = \zeta_t, \tag{G.3}$$

where $c_t = [c_{0,t}, \ldots, c_{J,t}]'$, $x_t = [x_{0,t}, \ldots, x_{J,t}]'$, $\mathbf{1}_{J+1}$ is a $(J+1)$-th dimensional vector of ones, and $\zeta_{j,t} \equiv \psi\rho + (1-\psi)\left[\pi_t(\alpha_{j,t}-1) + \mu + \mu_{p,t} + \sigma\sigma'_{p,t} - \frac{\gamma_j}{2}\|\sigma_{R,t}\|^2\alpha^2_{j,t}\right] + \xi_{j,t}$. Applying the Sherman-Morrison formula, we obtain

$$c_t = [I - (1-\psi^{-1})\mathbf{1}_{J+1}x'_t]\zeta_t, \tag{G.4}$$

or $c_{j,t} = \zeta_{j,t} - (1-\psi^{-1})x'_t\zeta_t$. Notice that ζ_t can be computed at order n based on the coefficients of order $n-1$ and their derivatives.

Computing the derivatives. The derivation above shows that, given the order $n-1$ expansion of $\zeta_j(X,\epsilon)$ and its derivatives, we can compute the expansion of order n. Suppose the expansion of $\zeta_j(X,\epsilon)$ is given by

$$\zeta_j(X,\epsilon) = \sum_{k=0}^{n-1} \zeta_{j,k}(X)\epsilon^k, \tag{G.5}$$

where $\zeta_{j,k}(X)$ takes the form:

$$\zeta_{j,k}(X) = A_{j,k} + B'_{j,k}(X-\overline{X}) + \frac{1}{2}(X-\overline{X})'C_{j,k}(X-\overline{X}), \tag{G.6}$$

where $\overline{X}$ is a reference point, $A_{j,k}$ is a scalar, $B_{j,k}$ is a vector, and $C_{j,k}$ is a matrix. Notice that . $\zeta_{j,k}(\overline{X}) = A_{j,k}$, $\zeta_{j,k,X}(\overline{X}) = B_{j,k}$ and $\zeta_{j,k,XX} = C_{j,k}$. Given this expansion, we can compute $c_j(X,\epsilon) = \sum_{k=0}^{n-1} c_{j,k}(X)\epsilon^k$.

G.1 Inner region

Consider the case of no preference heterogeneity and no leverage constraints. Consider the following change of variables: $x_1 = \epsilon\tilde{x}_1$. Define $\tilde{c}_j(\tilde{x}_1, \overline{\alpha}_p) = c_j\left(x_1, \overline{\alpha}_p\right)$, so $c_{j,x_1} = \frac{1}{\epsilon}\tilde{c}_{j,\tilde{x}_1}$ and $c_{j,x_1x_1} = \frac{1}{\epsilon^2}\tilde{c}_{j,\tilde{x}_1\tilde{x}_1}$. This implies the following is true:

$$\sigma_{c_j} = \frac{1}{\epsilon}\frac{\tilde{c}_{j,\tilde{x}_1}}{\tilde{c}_j}\sigma_{x_1}, \tag{G.7}$$

where $\sigma_{x_1} = -\frac{1-\epsilon\tilde{x}_1}{\tilde{x}_1}\hat{\alpha}_p\sigma_R$. Similarly, we can write σ_y

$$\sigma_y = \frac{1}{\epsilon}\frac{\tilde{y}_{x_1}}{y}\sigma_{x_1}. \tag{G.8}$$

The drift of y is given by

$$\mu_y = \frac{1}{\epsilon}\frac{\tilde{y}_{\tilde{x}_1}}{y}\mu_{\tilde{x}_1} + \frac{1}{2\epsilon^2}\frac{\tilde{y}_{\tilde{x}_1\tilde{x}_1}}{y}\sigma^2_{\tilde{x}_1}. \tag{G.9}$$

The term of order 0 is the same as before.