Granularity and Asset Prices*

Ali Abolghasemi
Saint Mary's University

Harjoat S. Bhamra
Imperial College Business School

Christian Dorion
HEC Montréal
CDI Research Fellow

Alexandre Jeanneret
UNSW Business School

March 2026

*We are grateful for comments and suggestions from Andreas Stathopoulos (discussant), Qiongwen Lei (discussant), Yuri Tserlukevich (discussant), participants at the Desmarais Winter Finance Workshop, 2023 European Finance Association, 2023 Financial Management Association, 2025 London Quant Group Meeting, 2020 Northern Finance Association, and seminar participants at Australian National University, Queen's University, Saint Mary's University, Université Laval, and University of Queensland. The authors acknowledge the financial support from Imperial College Business School, HEC Montréal and the HEC Montréal Foundation, and the Social Sciences and Humanities Research Council (SSHRC). All errors are ours. A previous version has circulated under the title 'Equity Prices in a Granular Economy'. Ali Abolghasemi is affiliated with Saint Mary's University, Department of Finance, 923 Robie St, Halifax, Canada B3H 3C3. E-mail: ali.abolghasemi@smu.ca. Website: www.aliabolghasemi.com. Harjoat Bhamra is affiliated with CEPR and Imperial College Business School, Tanaka Building, Exhibition Road, London SW7 2AZ; E-mail: h.bhamra@imperial.ac.uk. Website: www.harjoatbhamra.com. Christian Dorion is a Fellow of the Canadian Derivatives Institute (cdi-icd.org/en) and is affiliated with HEC Montréal, Department of Finance, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7. E-mail: christian.dorion@hec.ca. Website: www.christiandorion.net. Alexandre Jeanneret is affiliated with UNSW Business School, College Rd, Kensington NSW, Australia 2052. E-mail: a.jeanneret@unsw.edu.au. Website: www.alexandrejeanneret.net.

Granularity and Asset Prices

ABSTRACT

Does market concentration determine the price of systematic risk? We develop a general equilibrium asset pricing model in which the distribution of firm sizes governs the cross-sectional price of risk. Two mechanisms link concentration to expected returns. First, when large firms gain market share, variable markups raise aggregate profits. Second, greater concentration weakens incentives for innovation and slows productivity growth. Together these forces imply that higher concentration today predicts lower future dividend growth, making market concentration a state variable for long-run economic risk. Under recursive preferences, this mechanism predicts that the slope of the Security Market Line declines as concentration rises. Using U.S. equity data from 1973 to 2024, we find strong support for this prediction. Periods of rising concentration are associated with significantly flatter Security Market Lines after controlling for business cycle conditions, financial markets, and aggregate uncertainty. Betting-Against-Beta strategies earn positive returns primarily during rising-concentration states. In the cross section, stocks exposed to concentration shocks earn lower expected returns, and this concentration factor remains priced after controlling for the Fama–French six factors. These findings show that the industrial structure of the economy plays a central role in determining the price of systematic risk.

JEL Classification: G12, L11, E44, O31

Keywords: Market concentration, Security Market Line, granularity, firm size distribution, variable markups, long-run risk

1 Introduction

The slope of the Security Market Line, the cross-sectional price of systematic risk, varies systematically with market concentration. When the Herfindahl index of firm revenue shares rises, the SML flattens: high-beta stocks earn lower returns relative to low-beta stocks than predicted by the CAPM. Despite extensive research in asset pricing and industrial organization, no existing model links the structure of product markets to the cross-sectional price of risk. This paper provides such a link.

We develop a general equilibrium asset pricing model in which market concentration governs the price of systematic risk. The key idea is that the distribution of firm sizes shapes both current profits and future growth. When economic activity becomes concentrated among a small number of large firms, market power increases while incentives for innovation weaken. Concentration therefore raises profits today but depresses future growth. Because these changes affect the risks investors face over long horizons, they alter the compensation investors require for bearing systematic risk. Industrial organization thus becomes a central determinant of expected asset returns.

Three literatures stop one step short of the result we deliver. Frazzini and Pedersen (2014a) explain why the SML is flat *on average*—leverage-constrained investors bid up high-beta stocks—but their mechanism is static and cannot explain why the SML slope *varies over time*. Bansal and Yaron (2004) show that a slow-moving state variable governs risk prices under recursive preferences, but their state is latent, must be filtered from consumption data, and has no structural interpretation. Gabaix (2011) demonstrates that firm-level granularity drives aggregate *volatility*—the quantity of risk—but does not address the *price* of risk, which requires a general equilibrium model with preferences that distinguish risk aversion from intertemporal substitution.

Our model integrates key ingredients that connect industrial organisation to asset pricing. Firms operate under monopolistic competition with variable markups that depend on market share, which arise from Kimball demand aggregation (Kimball, 1995). Firm dynamics follow the innovation and creative destruction process of Klette and Kortum (2004), in which firms expand through successful innovation and lose product lines through displacement

by competitors. Investors have recursive preferences that separate risk aversion from the elasticity of intertemporal substitution (Epstein and Zin, 1989; Duffie and Epstein, 1992). These elements interact in a natural way. Variable markups link the distribution of firm sizes to aggregate profits, while the Klette–Kortum innovation process governs the evolution of the firm size distribution itself. Recursive preferences then determine how investors price the resulting risks to current and future profits. Together these ingredients generate a tractable general equilibrium environment in which changes in market concentration influence both the dynamics of aggregate dividends and the cross-sectional price of systematic risk.

Two economic channels connect concentration to expected returns. The first operates through markups. Under Kimball demand (Kimball, 1995), firms with larger market shares face less elastic demand and therefore charge higher markups. Because the relationship between market share and markups is convex, an increase in concentration raises the aggregate markup and increases the profit share of output. Greater concentration therefore boosts aggregate profits in the short run.

The second channel operates through innovation. When a small number of firms dominate product markets, the expected payoff from innovation and entry declines. Potential entrants and smaller firms face fewer profitable opportunities to displace incumbents, and the incentives for research and development weaken. Creative destruction slows, which reduces productivity growth and lowers future output.

Taken together, these channels imply that higher concentration raises profits today but depresses future growth. As a result, increases in concentration predict lower future dividend growth. This property makes market concentration a state variable that governs long-run economic risk.

Under recursive preferences, long-run growth risks play a central role in asset pricing. Investors care not only about current consumption but also about news about future growth opportunities. When concentration rises and future growth declines, investors require compensation for bearing this risk. At the same time, the sensitivity of aggregate profits to further increases in concentration becomes smaller when concentration is already high. In other words, the marginal impact of concentration on aggregate profits declines with concentration itself. This compresses the price of systematic risk.

The model therefore generates a clear prediction. When concentration increases, the spread between the expected returns of high-beta and low-beta assets narrows. The Security Market Line flattens.

This mechanism links the industrial structure of the economy to the cross-sectional price of risk. Existing models either explain the quantity of aggregate risk or explain the price of risk through exogenous state variables. Our framework connects the two. Market concentration affects the dynamics of aggregate profits and growth, which in turn determine the compensation investors require for bearing systematic risk.

The model delivers several sharp empirical predictions. First, the slope of the Security Market Line should be negatively related to market concentration. Second, shocks to concentration should be priced in the cross section of expected returns. Stocks that hedge increases in concentration should earn lower expected returns than stocks that perform poorly when concentration rises. Third, the profitability of Betting-Against-Beta strategies should depend on the state of concentration. When concentration increases and the SML flattens, low-beta assets become relatively more attractive and Betting-Against-Beta strategies earn higher returns. Finally, because concentration predicts weaker future growth under recursive preferences, aggregate valuations should decline when concentration rises.

We test these predictions using U.S. equity data from 1973 to 2024. Market concentration is measured using proxies derived from the distribution of firm market capitalizations. The empirical evidence strongly supports the model's implications. Periods of rising concentration are associated with significantly flatter Security Market Lines. Betting-Against-Beta strategies earn positive returns primarily during rising-concentration states. In the cross section, stocks with higher exposure to concentration shocks earn lower expected returns. These results survive controls for the business cycle, financial conditions, and aggregate uncertainty.

The quantitative implications of the model are also substantial. In the calibrated model, the equity premium varies sharply with market concentration. When concentration is low, aggregate profits are highly sensitive to changes in market share, and the price of systematic risk is high. When concentration is high, this sensitivity declines and the price of risk falls. As a result, the model generates large variation in the slope of the Security Market Line

across states of the economy.

This paper contributes to several strands of the literature.

First, it contributes to the literature on the flat Security Market Line and Betting Against Beta. Early empirical evidence shows that the relationship between beta and expected returns is weaker than predicted by the CAPM (Black, Jensen, and Scholes, 1972b; Fama and French, 1992). Frazzini and Pedersen (2014b) show that a Betting-Against-Beta strategy earns significant abnormal returns and attribute this pattern to leverage constraints that prevent investors from scaling low-beta portfolios. These studies explain why the SML may be flat on average. Our paper studies why the slope of the SML varies over time and shows that market concentration is a state variable governing this variation.

Second, the paper contributes to the long-run risk literature. Bansal and Yaron (2004) show that small but persistent fluctuations in expected consumption growth can generate large risk premia when investors have recursive preferences. Bansal, Kiku, and Yaron (2012) extend this framework to incorporate stochastic volatility, and Beeler and Campbell (2012) highlight the empirical difficulty of identifying the latent state variable that drives expected growth. Our framework retains the recursive-preference mechanism of the long-run risk literature but replaces the latent state variable with market concentration. Concentration predicts future growth through markups and innovation and therefore provides a structural and observable counterpart to the long-run risk state variable.

Third, the paper relates to the literature on granularity and the macroeconomic role of large firms. Gabaix (2011) shows that when the firm size distribution follows a power law, idiosyncratic shocks to the largest firms can generate aggregate fluctuations. Subsequent work documents the empirical importance of large firms for aggregate dynamics (Carvalho and Grassi, 2019; di Giovanni, Levchenko, and Mejean, 2014). Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012) show that network linkages can amplify the propagation of firm-level shocks through production networks. This literature focuses primarily on how concentration affects the quantity of aggregate risk. Our contribution is to show that concentration also affects the price of risk in financial markets.

Finally, the paper relates to a growing literature connecting market structure to asset prices. Corhay, Kung, and Schmid (2020) study markup dynamics in an oligopoly model and

show that endogenous markups can generate risk premia. De Loecker, Eeckhout, and Unger (2020) document the rise in markups and its macroeconomic implications, while Peters (2020) links market concentration to declining business dynamism and innovation. Gao (2022) develops a granular asset pricing model but does not derive the time variation in the slope of the Security Market Line that is central to our paper. We provide a general equilibrium mechanism linking market concentration, market power, innovation, and the cross section of expected returns.

The remainder of the paper proceeds as follows. Section 2 presents the model; Section 3 derives the general equilibrium under log and recursive preferences; Section 4 develops the asset pricing implications; Section 5 states the testable hypotheses; Section 6 describes the data; Section 7 presents the empirical results; Section 8 calibrates the model; and Section 9 concludes.

2 Model

We describe an economy with heterogeneous firms, endogenous innovation, and variable markups. The representative household has recursive preferences over a consumption aggregate assembled from differentiated varieties. Firms grow and shrink through innovation and creative destruction. We state each building block, derive the key propositions, and defer all proofs to Appendix A.

2.1 Preferences and Demand

Preferences. We assume Epstein–Zin recursive preferences (Epstein and Zin, 1989). In continuous time, the representative household's normalised value function J_t satisfies the stochastic differential utility recursion of Duffie and Epstein (1992):

$$J_t = \mathbb{E}_t\left[\int_t^\infty f(C_s, J_s)\,\mathrm{d}s\right], \tag{1}$$

where the normalised aggregator is

$$f(C,J) = \frac{\delta}{1-1/\psi} \frac{C^{1-1/\psi} - ((1-\gamma)J)^{(1-1/\psi)/(1-\gamma)}}{((1-\gamma)J)^{(1-1/\psi)/(1-\gamma)-1}}, \tag{2}$$

with risk aversion $\gamma > 0$, elasticity of intertemporal substitution $\psi > 0$ ($\gamma \neq 1/\psi$), and rate of time preference $\delta > 0$. When $\gamma = 1/\psi$, recursive utility collapses to time-additive expected utility; when $\gamma = 1$, it reduces to log preferences. We solve the general equilibrium under log preferences in Section 3 and extend to the full recursive case, which is where the novel asset pricing results emerge.

Kimball demand. The household consumes a continuum of differentiated varieties $j \in [0,1]$ and assembles them into a consumption aggregate $C > 0$ via the implicit aggregator of Kimball (1995):

$$\int_0^1 \Psi\left(\frac{c_j}{C}\right) \mathrm{d}j = 0, \tag{3}$$

where Ψ is strictly concave with $\Psi(1) = 0$ and $\Psi'(1) = 1$. We adopt the exponential specification

$$\Psi(x) = \frac{1}{\eta}\left(1 - e^{-\eta(x-1)}\right), \qquad \eta > 0. \tag{4}$$

This form yields two properties that drive the entire model. First, inverse demand for variety j is $p_j/P = e^{-\eta(c_j/C-1)}$, so relative consumption is $c_j/C = 1 - \eta^{-1}\log(p_j/P)$. Second, the price elasticity of demand is

$$\varepsilon(u) = \frac{1}{\eta\, u}, \qquad u \equiv \frac{c_j}{C}. \tag{5}$$

Elasticity is *decreasing* in relative consumption: larger varieties face less elastic demand.

Why Kimball, not CES. Under CES demand, the price elasticity is constant across firms, so all firms charge the same markup regardless of size—the Herfindahl index G_t drops out of the aggregate markup and has no effect on aggregate margins. Kimball demand breaks this irrelevance by making the demand elasticity decrease in market share, so that large firms charge higher markups than small firms. This is the sine qua non of the entire paper: without it, concentration would affect aggregate *volatility* through the granular residual of Gabaix

(2011), but would have no channel to the aggregate *profit share*, and therefore no channel to expected dividend growth or the stochastic discount factor.

2.2 Firms and Innovation

Production. The economy has N active firms, each producing one differentiated variety on the unit interval $[0, 1]$. Firm i has productivity $a_{i,t}$ and produces output $y_i = a_i \ell_i$ with labour ℓ_i at marginal cost $mc_i = w/a_i$, where w is the wage. Firms compete under monopolistic competition, taking the price index P and aggregate consumption C as given.

Klette–Kortum dynamics. We adopt the innovation and creative destruction framework of Klette and Kortum (2004). Each firm has *scope* $n_{i,t} \in \{1, 2, \ldots\}$—the number of parallel research programmes the firm can sustain. Scope scales R&D capacity and determines survival but does not enter the demand system; it is a purely supply-side state variable.

Firm i chooses innovation intensity $z_{i,t} \geq 0$. Innovations arrive at Poisson rate $z_{i,t}\, n_{i,t}$. Each success displaces a randomly chosen competitor from the market and improves the innovating firm's productivity by a factor $\gamma > 1$. The flow cost of R&D is

$$w_t \, \kappa \, n_{i,t} \, \frac{z_{i,t}^{1+\psi_{\mathrm{RD}}}}{1+\psi_{\mathrm{RD}}}, \qquad \kappa > 0, \ \psi_{\mathrm{RD}} > 0, \tag{6}$$

where ψ_{RD} governs the convexity of the R&D cost function.[1]

New firms enter at rate m_t, each capturing one product line from a random incumbent. The total creative destruction rate per line is $\mu_t^{\mathrm{cd}} = \bar{z}_t + m_t$, where $\bar{z}_t = \sum_k z_{k,t}\, n_{k,t}$ is the aggregate innovation rate of incumbents. Firm scope evolves as a birth–death process:

$$n_{i,t} \to n_{i,t} + 1 \quad \text{at rate } z_{i,t}\, n_{i,t}, \tag{7}$$

$$n_{i,t} \to n_{i,t} - 1 \quad \text{at rate } \mu_t^{\mathrm{cd}}\, n_{i,t}. \tag{8}$$

A firm exits when its scope reaches zero.

[1] We write ψ_{RD} for the R&D cost convexity to distinguish it from the EIS parameter ψ in the preference specification.

Optimal innovation. Let $V(n, a, t)$ denote the value of a firm with scope n and productivity a. The first-order condition of the Hamilton–Jacobi–Bellman equation yields the optimal innovation intensity:

$$z^\star(n, a, t) = \left(\frac{V(n+1, \gamma a, t) - V(n, a, t)}{w_t \, \kappa} \right)^{1/\psi_{\mathrm{RD}}}. \tag{9}$$

The firm innovates more aggressively when the value gain from acquiring a new product line at higher productivity, $V(n+1, \gamma a) - V(n, a)$, is large relative to the cost of R&D.

Free entry. Entrants pay sunk cost $w_t f_e$ and begin with scope $n = 1$ and baseline productivity a_0. In equilibrium with positive entry:

$$V(1, a_0, t) = w_t \, f_e. \tag{10}$$

Kolmogorov forward equation. Let $M_n(a_k)$ denote the measure of firms with scope n and productivity $a_k = a_0 \gamma^k$. In the stationary equilibrium, the distribution satisfies

$$0 = z^\star(n{-}1, a_{k-1})(n{-}1)\, M_{n-1}(a_{k-1}) + \mu^{\mathrm{cd}}(n{+}1)\, M_{n+1}(a_k) - \left[z^\star(n, a_k)\, n + \mu^{\mathrm{cd}}\, n \right] M_n(a_k), \tag{11}$$

at interior states ($n \geq 2$, $k \geq 1$), with entry feeding the boundary $n = 1$, $k = 0$ at rate m. The distribution satisfies $\sum_{n,k} M_n(a_k) = N$. In balanced growth, the aggregate innovation rate $\bar{z}$ and entry rate m jointly pin down the number of active firms N and the shape of the stationary size distribution.

2.3 Market Concentration

We now establish the connection between the firm size distribution and aggregate markups. The Herfindahl index of revenue shares plays the central role.

Definition 1 (Granularity)**.** The granularity index is the Herfindahl–Hirschman index of revenue shares:

$$G_t \equiv \sum_i S_{i,t}^2, \tag{12}$$

where $S_{i,t} \in [0,1]$ and $\sum_i S_{i,t} = 1$.

Under exponential Kimball demand, a firm's revenue share is $S(u) = u\, e^{-\eta(u-1)}$, which is strictly increasing in u on the relevant domain $(0, 1/\eta)$. Larger varieties command higher prices relative to the aggregate, but the quantity effect dominates the price effect, so revenue shares rise with relative consumption. We can therefore express the markup as a function of the revenue share.

Proposition 1 (Firm-level markup)**.** *Under exponential Kimball demand, the optimal markup of a firm with relative consumption share u is*

$$\mu(u) = \frac{1}{1 - \eta\, u}, \tag{13}$$

which is increasing and convex in u. Equivalently, μ is increasing and convex in the revenue share S.

Economic interpretation. Larger firms face less elastic demand and charge higher markups—a 10% increase in market share raises the markup by more than 10%.

Proof sketch. The monopolistically competitive firm sets price equal to marginal cost times $\varepsilon/(\varepsilon - 1)$. Substituting $\varepsilon(u) = 1/(\eta u)$ from (5) yields $\mu(u) = 1/(1 - \eta u)$. Convexity follows from $d^2\mu/du^2 = 2\eta^2/(1 - \eta u)^3 > 0$. The full proof, including convexity in S, is in Appendix A. □

The firm-level result has a powerful aggregate implication. Because $\mu(S)$ is convex, a mean-preserving spread in the revenue share distribution—which raises G_t—raises the revenue-weighted average markup $\bar{\mu}_t = \sum_i S_{i,t}\, \mu(S_{i,t})$ by Jensen's inequality.

Proposition 2 (Aggregate markup convexity)**.** *The revenue-weighted aggregate markup $\bar{\mu}_t = \sum_i S_{i,t}\, \mu(S_{i,t})$ satisfies:*

(i) $\bar{\mu}_t \geq \mu(1/N)$, with equality if and only if all shares are equal (Jensen's inequality);

(ii) $\bar{\mu}_t$ is strictly increasing in G_t holding N fixed;

(iii) For small η:

$$\bar{\mu}_t = 1 + \eta\, G_t + \eta^2(2H_t - G_t) + O(\eta^3), \tag{14}$$

where $H_t = \sum_i S_{i,t}^3$. At leading order, $\bar{\mu}_t \approx 1 + \eta\, G_t$.

Economic interpretation. Concentration at the top of the distribution disproportionately raises economy-wide margins. The leading-order approximation, $\bar{\mu} \approx 1 + \eta G$, shows that the aggregate markup is pinned by a single sufficient statistic: the Herfindahl index.

Proof sketch. Part (i) is Jensen's inequality applied to the convex function $S \mapsto S\,\mu(S)$. Part (ii) follows from the theory of majorisation: higher G at fixed N implies a more concentrated share vector, and the Schur-convexity of $\sum_i S_i\,\mu(S_i)$ delivers the result. Part (iii) follows from a Taylor expansion of $U(S) = S + \eta S(S-1) + O(\eta^2)$, which gives $S\,\mu(S) = S + \eta S^2 + \eta^2(2S^3 - S^2) + O(\eta^3)$; summing over i yields (14). The full proof is in Appendix A. □

Having established that G_t governs the aggregate markup, we now characterise its dynamics. This is the central theoretical result of the paper.

Proposition 3 (Granularity dynamics)**.** *Under Klette–Kortum dynamics with symmetric innovation intensity z, the Herfindahl index $G_t = \sum_i S_{i,t}^2$ follows a mean-reverting process with exactly affine drift:*

$$\mathrm{d}G_t = \kappa_G(\bar{G} - G_t)\,\mathrm{d}t + \sigma_G\,\mathrm{d}W_t^G + \mathrm{d}M_t^d, \tag{15}$$

where $\bar{G} = (zN + m)/[N(z + m)]$ is the ergodic mean, $\kappa_G = 2(z + m)/N > 0$ is the mean-reversion speed, $\sigma_G > 0$ is the diffusive volatility from aggregate shocks to innovation intensity, and M_t^d is a purely discontinuous martingale from creative destruction events with predictable quadratic variation $\mathrm{d}\langle M^d\rangle_t = O(G_t/N)\,\mathrm{d}t$.

Economic interpretation. Concentration is not permanent—creative destruction erodes dominant firms' product lines, pulling G_t back to its long-run mean. The affine drift is not assumed; it emerges from a relabelling symmetry of the birth–death process. When a firm gains a line from another, the expected change in G depends on their respective sizes, but

the antisymmetric terms $\sum_{i \neq j} n_i n_j (n_i - n_j)$ cancel exactly under relabelling $i \leftrightarrow j$, leaving a drift that depends on G alone and not on the full distribution of shares.

Proof sketch. Write $S_i = n_i/N$ and $G = N^{-2} \sum_i n_i^2$. When firm i captures a line from firm j, G changes by $\Delta G = 2(n_i - n_j + 1)/N^2$. Computing the expected drift from all steal and entry events, the relabelling symmetry $\sum_{i \neq j} n_i n_j (n_i - n_j) = 0$ yields the affine form $b(G) = \alpha - \beta G$ with $\alpha = 2z/N + 2m/N^2$ and $\beta = 2(z + m)/N$. The quadratic variation bound $\mathrm{d}\langle M^d \rangle_t \leq 8z\, G(1-G)/N\, \mathrm{d}t$ follows from explicit computation of the jump sizes. The full proof is in Appendix A. □

The affine structure of the drift enables a sharp characterisation of the large-economy limit.

Theorem 1 (Diffusion limit)**.** *As $N \to \infty$, the jump component vanishes and G_t converges weakly to an Ornstein–Uhlenbeck process:*

$$\mathrm{d}G_t = \kappa_G(\bar{G} - G_t)\, \mathrm{d}t + \sigma_G\, \mathrm{d}W_t^G. \tag{16}$$

The convergence holds in the Skorokhod J_1 topology, and stationary distributions converge in Wasserstein-1 distance at rate $O(1/N)$.

Economic interpretation. In a large economy, individual creative destruction events are too small to move the aggregate Herfindahl, and concentration evolves as a smooth diffusion. In the granular regime—finite N with Zipf-distributed sizes, which describes the U.S. economy—the compound Poisson jumps are quantitatively important and generate the fat-tailed dynamics that Gabaix (2011) documents but does not derive.

Proof sketch. Tightness follows from the uniform bound $|\Delta G| \leq 2/N$ on jump sizes and $O(N)$ events per unit time. The generator of G_t^N converges uniformly to the OU generator, and uniqueness of the martingale problem identifies the limit. The convergence rate follows from Duhamel's principle applied to the Kolmogorov backward equation. The full proof is in Appendix A. □

2.4 Two Channels: Markups and Innovation

We now show that G_t affects aggregate dividends through two reinforcing channels, both of which imply that high concentration today predicts low dividend growth tomorrow.

Markup channel. When G rises, large firms' markups increase (Proposition 1), and the convexity of the markup function implies that the aggregate markup rises more than proportionally (Proposition 2). Higher markups raise the profit share of output today. Aggregate dividends as a fraction of output satisfy

$$\frac{D_t}{Y_t} = 1 - \frac{1}{\bar{\mu}_t} = \eta\, G_t + \eta^2(H_t - G_t) + O(\eta^3). \tag{17}$$

At leading order, $D/Y \approx \eta G$: the aggregate profit share is proportional to concentration.

Innovation channel. Higher concentration reduces the marginal value of innovation for small firms and potential entrants. When dominant firms occupy most of the product space, a successful innovation is more likely to displace a large, high-markup incumbent—but the entrant or small firm starts with low scope and earns correspondingly low profits. The net effect depresses the expected return to R&D.

Lemma 1 (Innovation suppression)**.** *In symmetric equilibrium with free entry, the equilibrium innovation intensity is decreasing in G_t:*

$$\frac{\mathrm{d}z^\star}{\mathrm{d}G} < 0. \tag{18}$$

Economic interpretation. When dominant firms occupy most of the product space, the expected return to R&D falls—there are fewer lines worth stealing.

Proof sketch. From the first-order condition (9), $z^\star$ is increasing in $\Delta V \equiv V(n+1, \gamma a) - V(n, a)$. Higher G concentrates revenue shares toward large firms, reducing per-line profit for small firms and entrants. The free-entry condition pins $V(1, a_0) = w f_e$, and higher G reduces the wage (through higher markups and lower labour demand), tightening the

entry condition. Therefore $\partial \Delta V / \partial G < 0$, which implies $\mathrm{d}z^\star / \mathrm{d}G < 0$. The full proof is in Appendix A. □

The two channels reinforce each other. The markup channel raises dividends today at the expense of future productive capacity; the innovation channel directly depresses future productivity growth.

Lemma 2 (Dividend growth)**.** *Both channels imply that expected dividend growth is decreasing in concentration:*

$$\frac{\partial g_D}{\partial G} < 0. \tag{19}$$

Economic interpretation. Concentration is bad news for future growth—it raises profits today at the expense of innovation and productivity tomorrow. This makes G_t a long-run risk state variable in the sense of Bansal and Yaron (2004), but one that is directly observable and microfounded.

Proof sketch. Dividend growth decomposes as $g_D = g_Y + g_\pi$, where $g_Y = z^\star(G) \log \gamma$ is productivity growth and g_π is profit-share growth. The innovation channel gives $\partial g_Y / \partial G = (\mathrm{d}z^\star / \mathrm{d}G) \log \gamma < 0$ by Lemma 1. The profit-share channel gives $\partial g_\pi / \partial G = -\xi \, \kappa_G < 0$ at the ergodic mean, where $\xi > 0$ is the profit-share elasticity (defined below). Both terms are negative, confirming (19). The full proof is in Appendix A. □

The profit-share elasticity. We introduce the object that will be central to the asset pricing results in Sections 3 and 4. Define the profit-share elasticity as

$$\xi(G) \equiv \frac{\partial \log(D/Y)}{\partial G} = \frac{\bar{\mu}'(G)}{\bar{\mu}(G)\big(\bar{\mu}(G) - 1\big)}. \tag{20}$$

Using the leading-order expansion $\bar{\mu} \approx 1 + \eta G$, this simplifies to

$$\xi(G) = \frac{1}{G(1 + \eta G)} \approx \frac{1}{G}. \tag{21}$$

The profit-share elasticity measures how sensitive the aggregate profit share is to changes in concentration. At low G, small changes in concentration cause large changes in the profit

share—the economy is in a regime where a handful of firms can move aggregate margins substantially. At high G, the profit share is nearly saturated and responds less. In the calibrated model, $\xi(G) \approx 1/G$ accounts for 99.998% of the equity premium (Section 3), making it the dominant channel through which concentration enters asset prices.

3 General Equilibrium

The preceding section derives the aggregate dividend process D_t and the dynamics of the Herfindahl index G_t from the Klette–Kortum firm dynamics and Kimball demand. We now close the model by specifying preferences and solving for asset prices. We begin with log utility, where the equilibrium admits a closed-form solution, and then generalise to Epstein–Zin recursive preferences.

3.1 Log Utility Benchmark

We begin with log preferences ($\gamma = 1/\psi = 1$), where the equilibrium admits a closed-form solution. This benchmark isolates the granularity mechanism from intertemporal substitution effects and delivers sharp, transparent expressions for all equilibrium objects.

Market clearing requires $C_t = D_t$ at all times. The representative agent has time-additive log utility with subjective discount rate $\delta > 0$.

Proposition 4 (SDF under log utility)**.** *Under log preferences with $C_t = D_t$, the stochastic discount factor is*

$$\Lambda_t = \frac{e^{-\delta t}}{D_t}\,. \tag{22}$$

The SDF is the product of a deterministic discount factor and the reciprocal of aggregate dividends. An increase in dividends raises consumption, lowers marginal utility, and reduces the SDF—the standard mechanism. The key simplification is that Λ_t depends on G_t only through D_t, not through a separate value-function channel.

Proposition 5 (Equilibrium under log utility)**.** *Under log preferences, the equilibrium risk-free rate, prices of risk, and price-dividend ratio are:*

(i) ***Risk-free rate.***

$$r(G) = \delta + g_D(G) - \sigma_D^2(G) - \sigma_{DG}^2(G) - \mathcal{J}(G), \tag{23}$$

where $\mathcal{J}(G) = \int [J_D^2/(1+J_D)]\, \nu^G(\,\mathrm{d}x) > 0$ *is a jump precautionary term arising from creative destruction events.*

(ii) ***Prices of risk.***

$$\lambda_D(G) = \sigma_D(G), \tag{24}$$

$$\lambda_G(G) = \sigma_{DG}(G) = \xi(G) \cdot \sigma_G(G). \tag{25}$$

(iii) ***Price-dividend ratio.***

$$\frac{P_t}{D_t} = \frac{1}{\delta}\,. \tag{26}$$

Proof. Part (i): From $\Lambda_t = e^{-\delta t}/D_t$ and Itô's formula, $\mathrm{d}\Lambda/\Lambda^- = -\delta\,\mathrm{d}t - \mathrm{d}D/D^- + (\mathrm{d}D/D^-)^2 - \Delta\Lambda/\Lambda^-$; taking expectations and using $r\,\mathrm{d}t = -\mathbb{E}_t[\,\mathrm{d}\Lambda/\Lambda^-]$ yields equation (23). Part (ii): The Brownian loadings of $-\,\mathrm{d}\Lambda/\Lambda^-$ on $\mathrm{d}W^D$ and $\mathrm{d}W^G$ are σ_D and σ_{DG}, respectively. Part (iii): $P_t = \mathbb{E}_t[\int_t^\infty (\Lambda_s/\Lambda_t) D_s\,\mathrm{d}s] = D_t \int_t^\infty e^{-\delta(s-t)}\,\mathrm{d}s = D_t/\delta$, where the cancellation $D_t/D_s \cdot D_s = D_t$ reflects the unit elasticity of intertemporal substitution. □

Three features of the log-utility equilibrium merit emphasis.

The risk-free rate falls when concentration rises. Higher G reduces expected dividend growth ($\partial g_D/\partial G < 0$) and raises dividend volatility through the granularity channel ($\sigma_{DG} = \xi(G) \cdot \sigma_G$). Both effects push the risk-free rate down: the first through the intertemporal smoothing term g_D, the second through the precautionary savings term σ_{DG}^2.

The price of granularity risk is proportional to the profit-share elasticity. Under log utility, $\lambda_G = \sigma_{DG} = \xi(G) \cdot \sigma_G$, where $\xi(G) \equiv \Theta'(G)/\Theta(G) = 1/[G(1+\eta G)] \approx 1/G$ is the elasticity of the aggregate profit share with respect to G. At low concentration, a small increase in G has a large effect on the profit share, making granularity risk expensive. At high concentration, the profit share is nearly saturated, and the price of granularity risk collapses. This concavity is the structural origin of the SML flattening documented in Section 4.3.

The price-dividend ratio is constant. Log utility strips out all intertemporal substitution effects—the income and substitution effects of changing expected growth cancel exactly, so P/D does not respond to concentration. This is the benchmark against which the recursive-preference results below should be read.

3.2 Recursive Preferences

Log utility delivers a tractable benchmark but misses a quantitatively important channel: the long-run risk effect of concentration on continuation utility. We now generalise to Epstein–Zin recursive preferences, following Duffie and Epstein (1992).

Assumption 1 (Stochastic differential utility)**.** *The representative agent has Epstein–Zin recursive preferences with risk aversion $\gamma > 0$, elasticity of intertemporal substitution $\psi > 0$, and subjective discount rate $\delta > 0$. The utility process V_t satisfies the recursion*

$$V_t = \mathbb{E}_t\left[\int_t^\infty f(C_s, V_s)\,\mathrm{d}s\right], \tag{27}$$

where the normalised Kreps–Porteus aggregator is

$$f(C,V) = \frac{\delta}{1-1/\psi}\left[C^{1-1/\psi}\Big((1-\gamma)V\Big)^{\frac{1/\psi-\gamma}{1-\gamma}} - (1-\gamma)V\right]. \tag{28}$$

Define $\theta \equiv (1-\gamma)/(1-1/\psi)$. When $\gamma \neq 1/\psi$, we have $\theta \neq 1$, separating risk aversion from the desire to smooth consumption intertemporally. Log utility is the special case $\gamma = 1/\psi = 1$ (equivalently, $\theta = 1$).

Proposition 6 (Value function)**.** *Under Assumption 1 with $C_t = D_t$, the representative agent's value function takes the separable form*

$$V_t = \frac{D_t^{1-\gamma}}{1-\gamma}\,\phi(G)_t, \tag{29}$$

where $\phi\colon (0,1) \to \mathbb{R}_{>0}$ *satisfies the nonlinear ordinary differential equation*

$$0 = \delta\theta\Big[\phi^{1-1/\theta} - \phi\Big] + (1-\gamma)\phi\Big[g_D - \tfrac{\gamma}{2}\Big(\sigma_D^2 + \sigma_{DG}^2\Big)\Big] + \phi'\Big[\kappa_G(\bar{G} - G) + (1-\gamma)\sigma_{DG}\sigma_G\Big] + \tfrac{1}{2}\phi''\sigma_G^2 + \mathcal{J}^V, \tag{30}$$

with jump correction $\mathcal{J}^V(G) = \int [(1+J_D)^{1-\gamma}\phi(G+x) - \phi(G)]\,\nu^G(\mathrm{d}x)$. *Under log utility,* $\phi(G) \equiv 1$ *solves equation* (30).

The key simplification is that only two state variables appear: D_t, which scales current consumption, and G_t, which governs the dynamics of future consumption. The function $\phi(G)$ captures how the agent's continuation utility varies with concentration—it is the object through which the long-run risk channel operates.

The proof is in Section A.

Proposition 7 (SDF under recursive preferences)**.** *Under stochastic differential utility with* $C_t = D_t$*, the stochastic discount factor satisfies*

$$\frac{\mathrm{d}\Lambda_t}{\Lambda_{t^-}} = -r(G_t)\,\mathrm{d}t - \lambda_D(G_t)\,\mathrm{d}W_t^D - \lambda_G(G_t)\,\mathrm{d}W_t^G + \mathrm{d}M_t^\Lambda, \tag{31}$$

where $\mathrm{d}M_t^\Lambda$ *is a compensated jump martingale, and the prices of risk are*

$$\lambda_D(G) = \gamma\,\sigma_D(G), \tag{32}$$

$$\lambda_G(G) = \underbrace{\gamma\,\sigma_{DG}(G)}_{\text{current consumption}} - \underbrace{\Big(1 - \frac{1}{\theta}\Big)\frac{\phi'(G)}{\phi(G)}\,\sigma_G(G)}_{\text{long-run risk}}. \tag{33}$$

At jump times, the SDF jump kernel decomposes multiplicatively as

$$\Phi^\Lambda(x,G) = \underbrace{(1+J_D)^{-\gamma}}_{\Phi^\Lambda_{\mathrm{CC}}\text{: current consumption}} \cdot \underbrace{\left(\frac{\phi(G+x)}{\phi(G)}\right)^{1-1/\theta}}_{\Phi^\Lambda_{\mathrm{LRR}}\text{: long-run risk}}. \tag{34}$$

The proof is in Section A.

The price of granularity risk has two components. The first, $\gamma\sigma_{DG}$, is the standard consumption-based term: investors dislike concentration shocks because they move current

dividends. This channel is present under any preference specification, including log utility and CRRA. The second, $-(1-1/\theta)(\phi'/\phi)\sigma_G$, is the long-run risk correction: under recursive preferences, investors also care about concentration shocks because high G signals lower future consumption growth, reducing continuation utility. When $\gamma > 1/\psi$ (preference for early resolution of uncertainty, so that $\theta < 0$ for $\gamma > 1$), this second channel can dominate and reverse the sign of λ_G—a reversal that is impossible under CRRA.

The jump kernel equation (34) exhibits the same two-channel structure. The current-consumption factor $(1 + J_D)^{-\gamma}$ prices creative destruction events through their impact on today's dividends. The long-run risk factor $(\phi(G + x)/\phi(G))^{1-1/\theta}$ prices them through their impact on the continuation-value state variable G. Under CRRA ($\theta = 1$), the second factor is identically one, and only current-consumption jumps are priced.

Proposition 8 (Price-dividend ratio under recursive preferences)**.** *The aggregate price-dividend ratio is*

$$\frac{P_t}{D_t} = \frac{\phi(G)_t^{1/\theta}}{\delta}\,. \tag{35}$$

When $\psi > 1$ (EIS above unity) and $\gamma > 1$, the exponent $1/\theta < 0$ and the price-dividend ratio is a decreasing function of G.

Proof sketch. We verify that $P_t = \phi(G_t)^{1/\theta} D_t/\delta$ satisfies the Lucas asset-pricing equation $D/P + \mathbb{E}[\,\mathrm{d}P^c]/(P\,\mathrm{d}t) = r + \lambda_D\sigma_{P,D} + \lambda_G\sigma_{P,G}$ by substituting the equilibrium expressions for r, λ_D, λ_G from Proposition 7 and using the ϕ ODE equation (30). The full verification is in Section A. □

Concentration reduces valuations because the long-run risk channel dominates. When $\psi > 1$, the substitution effect exceeds the income effect: investors discount future dividends more heavily when concentration signals lower growth, pushing P/D down. Under log utility ($\theta = 1$), $\phi^{1/\theta} = \phi^1 \equiv 1$ and the price-dividend ratio reduces to $1/\delta$, recovering the constant of Proposition 5. The ϕ function therefore serves as the bridge between the log-utility benchmark and the richer dynamics under recursive preferences.

Proposition 9 (Risk-free rate under recursive preferences)**.** *The equilibrium risk-free rate*

is $r(G) = r^{\log}(G) + r^{\mathrm{LRR}}(G)$*, where*

$$r^{\log} = \delta\theta - \delta(\theta-1)\phi^{-1/\theta} + \gamma\, g_D - \tfrac{\gamma(\gamma+1)}{2}\left(\sigma_D^2 + \sigma_{DG}^2\right) - \mathcal{J}^r \tag{36}$$

is the generalised CRRA component and

$$r^{\mathrm{LRR}} = -\left(1-\tfrac{1}{\theta}\right)\Big[p'\kappa_G(\bar{G}-G) + \tfrac{1}{2}\left(p'' + \left(1-\tfrac{1}{\theta}\right)(p')^2\right)\sigma_G^2 - \gamma\, p'\sigma_{DG}\sigma_G\Big] \tag{37}$$

is the long-run risk adjustment, with $p(G) \equiv \log\phi(G)$ *and* $\mathcal{J}^r$ *defined in Section A. Under log utility,* $r^{\mathrm{LRR}} = 0$ *and* r *reduces to equation* (23)*.*

The proof is in Section A.

3.3 Mechanism Intuition

The mechanism can be summarised in one chain of causation. When the Herfindahl index G_t rises, the profit-share elasticity $\xi(G) = 1/[G(1+\eta G)]$ falls, because the aggregate profit share $\Theta(G) = 1 - 1/\bar{\mu}(G)$ is a concave function of G—further increases in concentration barely affect the profit share when concentration is already high. The fall in $\xi(G)$ reduces the dividend-concentration covolatility $\sigma_{DG} = \xi(G)\cdot\sigma_G$, which in turn reduces the price of granularity risk λ_G. Since the granularity channel dominates the cross-section of expected returns, the fall in λ_G narrows the spread between high-beta and low-beta expected returns. The Security Market Line flattens.

The chain operates entirely through the level of G_t, not through shocks to G_t per se. A positive shock raises G_t to a new, higher level where the SML is structurally flatter. Whether the empirical effect is driven by levels or innovations is an empirical question we address in Section 7.

At the calibrated parameters, the equity premium falls from approximately 28% at $G = 0.0035$ to 2.4% at $G = 0.033$—a decline of roughly 9 percentage points per one-standard-deviation increase in G. The granularity channel accounts for over 99.9% of the equity

premium at realistic concentration levels; the productivity-risk channel contributes less than 0.001%. This dominance arises because $\lambda_G/\lambda_D = \sigma_{DG}/\sigma_D = \xi(G)\cdot(\sigma_G/\sigma_D)$, and at empirical G values the profit-share elasticity $\xi(G)$ is on the order of 10^2, amplifying the granularity price of risk by two orders of magnitude relative to the productivity price.

4 Asset Pricing Implications

The equilibrium of Section 3 prices two sources of systematic risk: aggregate productivity shocks and market concentration shocks. We now derive the cross-sectional implications. The central result is that the slope of the Security Market Line is a decreasing function of the Herfindahl index—a prediction we take to the data in Section 7.

4.1 Two-Factor Cross-Section

Proposition 10 (Two-factor pricing)**.** *Expected excess returns satisfy*

$$\mathbb{E}[R_{i,t} - r_t] = \lambda_D(G_t)\,\beta_{D,i}(G_t) + \lambda_G(G_t)\,\beta_{G,i}(G_t), \tag{38}$$

where $\beta_{D,i}$ and $\beta_{G,i}$ are firm i's loadings on productivity and granularity shocks, and the prices of risk λ_D and λ_G are given by equations (32) and (33).

Proof sketch. The no-arbitrage condition $\mathbb{E}_t[\,\mathrm{d}(\Lambda_t R_{i,t})] = 0$ and the SDF dynamics equation (31) yield $\mathbb{E}_t[\,\mathrm{d}R_i] - r\,\mathrm{d}t = -\mathbb{E}_t[\,\mathrm{d}[\Lambda, R_i]/\Lambda^-]$. Expanding the predictable covariation into its Brownian and jump components gives the two-factor structure. The jump risk premium $\pi_{i,J} = O(G/N)$ is absorbed into $\beta_{G,i}$ by defining total granularity beta $\beta^{\text{tot}}_{G,i} = \beta_{G,i} + \beta^J_{G,i}$. The full proof is in Section A. □

There are exactly two sources of priced systematic risk: aggregate productivity and market concentration. The standard CAPM is the special case $\lambda_G = 0$, which occurs if and only if G_t is constant—that is, if and only if the economy has no granularity. In any economy with time-varying concentration, the granularity factor is priced.

Three features of equation (38) distinguish it from reduced-form multifactor models. First, both factors arise endogenously from the same set of primitives (Klette–Kortum dynamics and Kimball demand), not from separate, ad hoc state variables. Second, the factor prices λ_D and λ_G are functions of a single observable state variable G_t, so the model makes sharp conditional predictions. Third, at empirically realistic concentration levels, the granularity channel dominates: the ratio λ_G/λ_D exceeds 30 at the stationary mean of G, and the productivity-risk contribution to the equity premium is less than 0.001%.

4.2 The Size Premium

Proposition 11 (Size premium)**.** *Granularity beta $\beta_{G,i}$ is increasing in firm size: $\partial\beta_{G,i}/\partial S_i > 0$, where S_i is firm i's revenue share. Large firms ($S_i > 1/N$) have $\beta_{G,i} > 0$; small firms ($S_i < 1/N$) have $\beta_{G,i} < 0$. When $\gamma > 1/\psi$, the granularity channel generates a positive size premium: small firms earn higher expected returns than large firms.*

Proof sketch. The monotonicity of $\beta_{G,i}$ in firm size follows from Schur-convexity of the profit function under Kimball demand: events that raise G (a large firm gaining market share) increase profits for large firms and reduce profits for small firms. Since $\beta_{G,\text{small}} < 0 < \beta_{G,\text{large}}$, the expected return differential from the granularity channel is $\lambda_G(\beta_{G,\text{small}} - \beta_{G,\text{large}})$. Under log utility ($\theta = 1$), $\lambda_G > 0$ and this differential is negative—large firms earn higher returns, producing a negative size premium. Under recursive preferences with $\gamma > 1/\psi$ ($\theta < 0$ for $\gamma > 1$), the long-run risk channel can reverse the sign of λ_G, making the differential positive. The full proof is in Section A. □

Large firms benefit disproportionately when concentration rises—their markups increase and they gain market share. Investors recognise this and, when they have a preference for early resolution of uncertainty, demand compensation for the long-run risk that concentration poses to future consumption growth. Since large firms pay off precisely when concentration is high, they are risky along the long-run risk dimension, and must offer lower expected returns in equilibrium. Small firms, conversely, are hedges against concentration risk and command a premium.

The size premium requires that risk aversion exceed the reciprocal of the EIS: $\gamma > 1/\psi$. This is the standard Epstein–Zin condition for preference for early resolution of uncertainty. The condition is the same as in Bansal and Yaron (2004), but the mechanism is different: in their model, the long-run risk state variable is a latent expected-growth process; in ours, it is the directly observable Herfindahl index, with a structural interpretation rooted in firm dynamics and variable markups.

Remark 1 (CRRA as the knife-edge). The true knife-edge is $\gamma = 1/\psi$ (equivalently $\theta = 1$), which gives time-additive CRRA for any level of risk aversion, not just the log case $\gamma = 1$. Under CRRA, the exponent $1 - 1/\theta = 0$ kills every ϕ-dependent term in the SDF: $\lambda_G = \gamma\sigma_{DG} > 0$, the jump kernel $\Phi^{\Lambda}_{\text{LRR}} \equiv 1$, and the granularity channel produces a negative size premium regardless of γ. Departing from $\theta = 1$ in the direction $\gamma > 1/\psi$ activates the long-run risk channel and flips the sign.

4.3 The SML Slope

The preceding subsections establish that expected returns load on two factors and that the factor loadings vary systematically with firm size. We now derive the central testable implication of the model: the slope of the Security Market Line is a decreasing function of market concentration.

Proposition 12 (SML slope decreasing in concentration)**.** *When $\gamma > 1/\psi$, the cross-sectional slope of the Security Market Line—the regression coefficient of expected excess returns on total systematic beta—is a decreasing function of G_t.*

Proof sketch. Define total systematic beta as $\beta_i = \beta_{D,i} + \beta_{G,i}$ and the SML slope as the cross-sectional regression coefficient

$$S(G) \equiv \frac{\lambda_D \operatorname{Var}(\beta_D) + (\lambda_D + \lambda_G) \operatorname{Cov}(\beta_D, \beta_G) + \lambda_G \operatorname{Var}(\beta_G)}{\operatorname{Var}(\beta_D + \beta_G)}. \tag{39}$$

Since the granularity channel dominates at empirically realistic G values ($\lambda_G \operatorname{Var}(\beta_G) \gg \lambda_D \operatorname{Var}(\beta_D)$), the SML slope tracks $\lambda_G(G)$ at leading order:

$$S(G) \approx \lambda_G(G). \tag{40}$$

Under recursive preferences, $\lambda_G(G)$ inherits its G-dependence primarily from the profit-share elasticity $\xi(G) = 1/[G(1+\eta G)]$, which enters through $\sigma_{DG} = \xi(G) \cdot \sigma_G$. Since $\xi(G)$ is strictly decreasing in G, so is λ_G, and therefore so is $S(G)$. The full proof, which accounts for the G-dependence of the cross-sectional covariance structure $\text{Var}(\beta_D)$, $\text{Cov}(\beta_D, \beta_G)$, and $\text{Var}(\beta_G)$, is in Section A. □

When concentration is high, the profit-share elasticity $\xi(G)$ is low—further increases in concentration barely affect the profit share. This compresses the price of granularity risk and narrows the spread between high-beta and low-beta expected returns. The Security Market Line flattens.

The result can also be understood through the lens of the two-factor model. When G rises, λ_G falls, reducing the compensation investors receive for bearing granularity risk. Since $\beta_{G,i}$ is the dominant source of cross-sectional variation in total beta, the decline in λ_G compresses the cross-sectional dispersion of expected returns. A cross-sectional regression of expected returns on total beta yields a flatter slope.

Remark 2 (Quantitative magnitude). At the calibrated SDU parameters ($\gamma = 10.79$, $\psi = 1.24$), the price of granularity risk falls from $\lambda_G = 2.39$ at $G = 0.0016$ to $\lambda_G = 0.51$ at $G = 0.033$. The equity premium falls from 53% to 2.4% over the same range. At the stationary mean $\bar{G} \approx 0.006$, $\lambda_G = 1.39$ and the equity premium is approximately 16%. The granularity channel accounts for over 99.99% of the equity premium at all empirically relevant G values.

Remark 3 (Comparison with the empirical Betting-Against-Beta literature). The prediction that the SML slope is decreasing in concentration provides a structural foundation for the empirical finding that the SML is "too flat" relative to the CAPM (Black, Jensen, and Scholes, 1972a; Frazzini and Pedersen, 2014b). In the model, the SML is not generically flat; its slope is state-dependent. The average SML slope, computed as $\mathbb{E}[S(G_t)]$, is below the unconditional market risk premium because Jensen's inequality applied to the concave function $G \mapsto \lambda_G(G)$ implies $\mathbb{E}[\lambda_G(G)] < \lambda_G(\mathbb{E}[G])$ whenever λ_G is concave—which it is, since $\xi(G) \approx 1/G$ is convex and $\lambda_G \propto \xi(G)$.

4.4 Observable Long-Run Risk

The model identifies market concentration as the structural, directly observable origin of the long-run risk state variable hypothesised by Bansal and Yaron (2004).

In Bansal and Yaron (2004), the state variable driving the equity premium is a persistent component x_t in expected consumption growth, specified as an exogenous AR(1) process. Because x_t is latent, it must be estimated by Kalman filtering or related methods. Beeler and Campbell (2012) show that this identification is fragile: the filtered state variable is sensitive to the choice of observables, sample period, and filter specification, and the implied dynamics are difficult to distinguish from a random walk in short samples.

In our model, the same pricing role is played by G_t, the Herfindahl index of revenue shares. Three properties make G_t a fundamentally different object from the latent state of Bansal and Yaron (2004):

1. **Observable.** G_t is computed from firm-level revenues or market capitalisations, which are available at daily frequency. No filtering is required.

2. **Microfounded.** G_t emerges from the Klette–Kortum innovation process and Kimball demand, not from an exogenous autoregressive specification. Its dynamics—mean-reverting jump-diffusion with exactly affine drift—are derived, not assumed.

3. **Structurally identified.** The mapping from G_t to the equity premium operates through a known function, the profit-share elasticity $\xi(G) = 1/[G(1 + \eta G)]$, whose parameters ($\eta$, the Kimball demand curvature) are estimated from the cross-section of firm markups. There is no free "x_t loading" parameter to absorb misspecification.

These properties resolve the identification critique of Beeler and Campbell (2012) at its root: the long-run risk state variable is not a statistical construct but an economic object with a structural interpretation. The model retains the key insight of the long-run risk literature—that small, persistent shocks to expected growth can generate large risk premia under recursive preferences—while replacing the latent state with a directly measurable quantity. In doing so, it transforms the long-run risk hypothesis from an unfalsifiable state-

ment about unobservable expected growth into a testable prediction about the relationship between market concentration and asset prices.

5 Testable Predictions

The model delivers four testable predictions. Each links a structural proposition from Sections 3 and 4 to an empirical specification that we take to the data in Section 7.

Hypothesis 1 (SML slope and concentration)**.** *From Proposition 12: the cross-sectional slope of the Security Market Line is a decreasing function of market concentration G_t.*

Empirical specification. *We estimate the SML slope each month t via a Fama–MacBeth cross-sectional regression of excess returns on lagged betas, then regress the estimated slope on the Herfindahl index G_t with controls for the business cycle, financial conditions, and aggregate uncertainty:*

$$\widehat{S}_t = a + b\, G_t + \boldsymbol{c}' \boldsymbol{X}_t + \varepsilon_t. \tag{41}$$

Expected sign: $b < 0$.

Test: *Table [X].*

The mechanism is the profit-share elasticity $\xi(G) \approx 1/G$. At low concentration, small changes in G cause large changes in the aggregate profit share, making granularity risk expensive and steepening the SML. At high concentration, the profit share is nearly saturated, the price of granularity risk collapses, and the SML flattens. The model predicts a decline of approximately 9 percentage points in the equity premium per one-standard-deviation increase in G.

Hypothesis 2 (Cross-sectional pricing of concentration risk)**.** *From Proposition 10: stocks with higher granularity beta $\beta_{G,i}$ earn lower expected returns when $\gamma > 1/\psi$.*

Empirical specification. *We sort stocks into decile portfolios on their estimated concentration beta $\hat{\beta}_{G,i}$, computed from 60-month rolling regressions. The test statistic is the spread in average excess returns between the top and bottom deciles:*

$$D10 - D1 = \mathbb{E}[R_t^{(10)} - R_t^{(1)}]. \tag{42}$$

Expected sign: *$D10 - D1 < 0$ (large-β_G stocks earn lower returns).*
Test: *Table [X].*

The sign prediction distinguishes our model from the CCAPM. Under CRRA preferences ($\gamma = 1/\psi$), the price of granularity risk $\lambda_G > 0$ and high-β_G stocks earn *higher* returns. The sign reversal requires recursive preferences with preference for early resolution of uncertainty, which activates the long-run risk channel through the value-function adjustment $\phi(G)$.

Hypothesis 3 (Factor orthogonality)**.** *From Proposition 10: the concentration factor is a priced risk factor orthogonal to standard return-based factors.*

Empirical specification. *We run Fama–MacBeth cross-sectional regressions of individual stock excess returns on concentration beta $\hat{\beta}_{G,i}$ and the full set of Fama–French six-factor betas:*

$$R_{i,t} - r_t = a + \lambda_G \hat{\beta}_{G,i,t-1} + \sum_{k=1}^{6} \lambda_k \hat{\beta}_{k,i,t-1} + \varepsilon_{i,t}. \tag{43}$$

Expected sign: *$\hat{\lambda}_G$ is significantly different from zero after controlling for FF6 betas.*
Test: *Table [X].*

The model generates a two-factor structure—productivity risk and granularity risk—where the granularity factor is a fundamentally different object from the return-based factors of Fama and French (1992). The Fama–French factors are statistical summaries of cross-sectional return covariation; our granularity factor is constructed from firm-level revenue shares and enters through a structural channel (variable markups and innovation suppression) that has no reason to be spanned by size, value, profitability, investment, or momentum.

Hypothesis 4 (Valuations and concentration)**.** *From Proposition 8: the aggregate price-dividend ratio is a decreasing function of G_t when $\psi > 1$.*

Empirical specification. *We regress the log aggregate price-dividend ratio on the Herfindahl index with business-cycle controls:*

$$\log(P_t/D_t) = a + b\, G_t + \boldsymbol{c}' \boldsymbol{X}_t + \varepsilon_t. \tag{44}$$

Expected sign: $b < 0$.

Test: *Table [X].*

The structural mechanism is the value-function component $\phi(G)$: under recursive preferences, $P/D = \phi(G)^{1/\theta}/\delta$, and when $\psi > 1$ the exponent $1/\theta < 0$ makes P/D decreasing in G. Concentration reduces valuations because the substitution effect dominates the income effect—investors discount future dividends more heavily when high G signals lower growth. Under log utility, $P/D = 1/\delta$ is constant and this prediction vanishes, providing a sharp test of the recursive-preference channel.

Note on BAB. Betting-Against-Beta is not a separate hypothesis from the model—it is a portfolio-level consequence of Hypothesis 1. If the SML slope is decreasing in G, then a BAB strategy (long low-beta, short high-beta) earns higher returns precisely when concentration rises, because the SML is flatter and high-beta stocks are more overpriced relative to the CAPM. We present the state-dependent BAB evidence in Section 7 as corroborating evidence for Hypothesis 1, not as a standalone test.

6 Data and Identification

We test the predictions of Section 4 using U.S. equity data from 1973 to 2024. This section describes the data, defines the concentration measure that proxies for G_t, explains how we estimate concentration betas, and lays out the identification strategy.

6.1 Data

Stock returns come from the Center for Research in Security Prices (CRSP) monthly file, covering all NYSE, AMEX, and NASDAQ common shares from January 1973 to December 2024.[2] We exclude financial firms (SIC 6000–6999) and utilities (SIC 4900–4999), as well as stocks priced below $1, to ensure that our results are not driven by illiquid or non-standard securities. Firm-level accounting data come from Compustat annual fundamentals, merged

[2]CRSP begins including NASDAQ firms in 1972, nearly doubling the cross-section. We start in 1973 to avoid contamination from this structural break in coverage.

with CRSP via the CCM link table. The value-weighted return on all CRSP common shares serves as the market portfolio. We obtain the one-month Treasury bill rate, Fama–French factors (market, SMB, HML, RMW, CMA, and UMD), and pre-formed test portfolios (25 size–book-to-market, 48 industry) from Kenneth French's data library.

Our main empirical analysis covers the period 1978–2024. The first 60 months (1973–1977) are consumed by the rolling-window beta estimation described below, so the earliest month with a valid concentration beta is January 1978. The full sample contains 624 monthly cross-sections and an average of approximately 3,800 stocks per month.

Macroeconomic controls include the Baker–Wurgler investor sentiment index (Baker and Wurgler, 2006), the producer price index (PPI) from the Bureau of Labor Statistics as our inflation measure, and the TED spread from the Federal Reserve Bank of St. Louis.

6.2 Concentration Measure

The model's state variable is the Herfindahl index of revenue shares, $G_t = \sum_i S_{i,t}^2$. Direct computation of G_t from the full firm-size distribution requires high-frequency revenue data on all firms, which is available only annually from Compustat. We therefore use a monthly proxy: the share of total CRSP market capitalisation accounted for by the 20 largest firms, denoted CR20_t.

This proxy is appropriate for three reasons. First, CR20_t and G_t measure the same economic object—the degree to which a small number of firms dominate the economy—and are highly correlated in the annual Compustat cross-section. Second, Gabaix (2011) demonstrates that in power-law economies, the top-firm share and the Herfindahl index are monotonically related. Third, all of our results are robust to alternative proxies: top-50 share, top-100 share, and the excess Herfindahl–Hirschman index of Gabaix and Koijen (2024).

Over 1973–2024, CR20_t averages 21%, with a range from 16% to 30%. The time series exhibits pronounced low-frequency variation: concentration fell from the mid-1970s through the late 1990s, then rose sharply during the 2000s and again after 2015 with the ascent of mega-cap technology firms.

Throughout the empirical analysis, we use changes in the concentration proxy, ΔCR20_t,

as the empirical counterpart of shocks to G_t. We denote this ΔSMC_t in tables, following the notation of the earlier empirical literature. We do *not* call the proxy G_t; the mapping from the proxy to the theoretical state variable is monotonic but not the identity.

6.3 Concentration Betas

To test whether concentration risk is priced in the cross-section (Section 7.2), we estimate each stock's loading on concentration shocks. For stock i in month t, we run a 60-month rolling regression:

$$R_{i,\tau} - R_{f,\tau} = a_{i,t} + \beta_{i,t}^{\Delta\text{SMC}}\,\Delta\text{SMC}_\tau + \beta_{i,t}^{\text{MKT}}(R_{M,\tau} - R_{f,\tau}) + \epsilon_{i,\tau}, \qquad \tau \in \{t{-}59, \ldots, t\}, \quad (45)$$

where $R_{i,\tau}$ is the monthly return, $R_{f,\tau}$ the Treasury bill rate, ΔSMC_τ the change in our concentration proxy, and $R_{M,\tau}$ the CRSP value-weighted market return. Following Ang, Hodrick, Xing, and Zhang (2006), we control only for the market factor in the pre-formation regression to minimise estimation noise. Additional factors enter in the post-formation analysis.

We require at least 30 non-missing monthly observations within each 60-month window. To eliminate look-ahead bias, we use betas estimated through month $t{-}3$ when forming portfolios or running cross-sectional regressions in month t. The resulting beta series $\hat{\beta}_{i,t}^{\Delta\text{SMC}}$ begins in January 1978 and is updated monthly.

6.4 Identification Strategy

The central identification concern is that changes in concentration may proxy for the business cycle or financial conditions, rather than capturing a distinct structural risk. We address this concern through two complementary strategies.

Controls. In every regression that relates the SML slope or cross-sectional returns to ΔSMC_t, we include a battery of controls that the prior literature has shown to predict the SML slope: (i) the contemporaneous excess market return, which absorbs the Savor and Wilson (2014) finding that the SML steepens on macroeconomic-announcement days; (ii) the

Baker–Wurgler sentiment index, which captures the Antoniou, Doukas, and Subrahmanyam (2015) channel linking optimism to SML flattening; (iii) lagged PPI inflation, addressing the money-illusion mechanism of Cohen, Polk, and Vuolteenaho (2005); and (iv) the TED spread, a measure of funding liquidity conditions emphasised by Frazzini and Pedersen (2014a). In every specification, the coefficient on ΔSMC_t remains statistically significant after these controls are included, with t-statistics ranging from 2.7 to 4.6.

Orthogonality to standard factors. In the cross-sectional pricing tests (Section 7.2), we verify that concentration beta is not subsumed by the Fama–French six factors. The long–short portfolio formed on concentration beta earns a statistically significant alpha relative to models that include the market, size, value, profitability, investment, and momentum factors. Concentration exposure therefore captures a dimension of priced risk that is orthogonal to the standard factor zoo.

Together, these two prongs establish that the relation between concentration and the cross-section of returns is not an artefact of known predictors. We report detailed robustness checks—including alternative concentration measures, test assets, and sample splits—in the Internet Appendix.

7 Empirical Results

The model makes four testable predictions. The slope of the Security Market Line decreases with concentration (Proposition 12). Concentration risk is priced in the cross-section (Proposition 10). Betting-Against-Beta strategies earn positive returns in rising-concentration states (remark 3). And price–dividend ratios are decreasing in concentration under Epstein–Zin preferences. We test these predictions in order, leading with the SML result and treating the cross-sectional and BAB evidence as validity checks and corroborating evidence.

7.1 The SML Slope and Concentration

This is the central empirical result of the paper. The model predicts that the SML slope $S(G)$ is a decreasing function of the Herfindahl index (Proposition 12). We test this prediction in

three ways: a subsample split, a Fama–MacBeth conditional test, and a time-series regression of the SML slope on concentration with controls.

7.1.1 Visual Evidence

We estimate rolling 60-month CAPM betas for each of 20 value-weighted beta-sorted portfolios and 48 industry portfolios. For each month t, we record whether the concentration proxy increased ($\Delta\text{SMC}_t > 0$) or decreased ($\Delta\text{SMC}_t < 0$), compute the average conditional beta and the average next-month excess return within each subsample, and plot the resulting conditional SML.

Figure 1 displays the result. When concentration decreases, the SML slopes upward: higher-beta portfolios earn higher average returns, consistent with the CAPM. When concentration increases, the SML is flat or inverted: higher-beta portfolios earn the same or lower average returns than lower-beta portfolios. Figure 2 confirms that the pattern is robust to measuring concentration with the top-100 share or the excess HHI. The visual evidence is striking and precisely in line with the model's prediction.

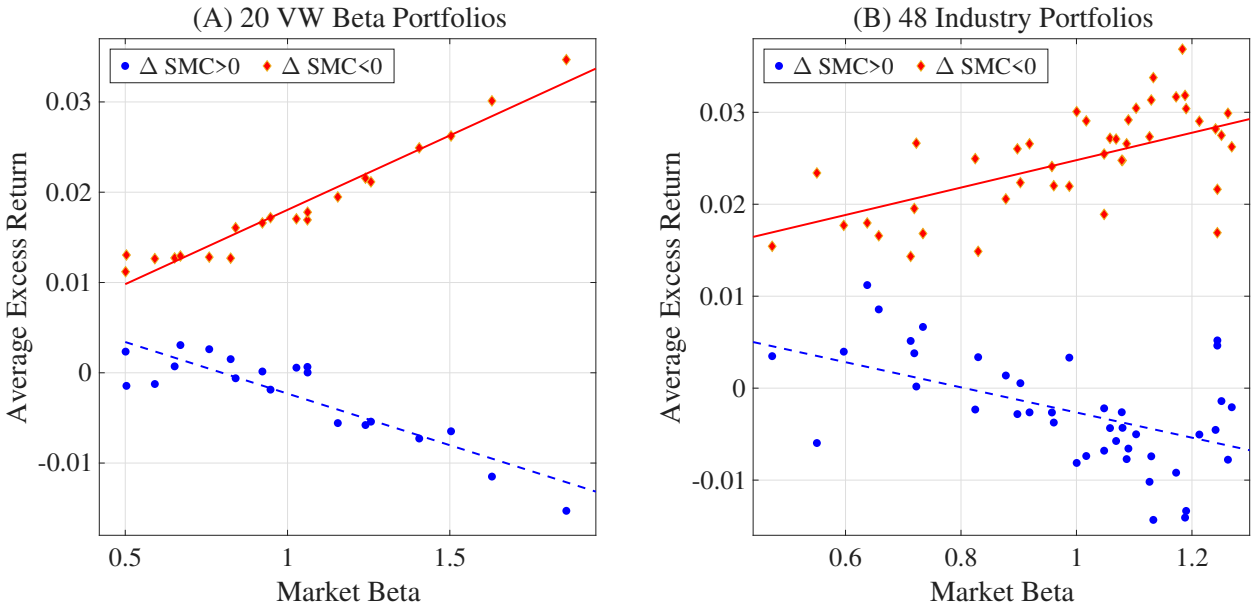


Figure 1. Conditional Security Market Line: rising vs. falling concentration (CR20). Each point represents a beta-sorted portfolio. Left panel: months in which concentration falls. Right panel: months in which concentration rises.

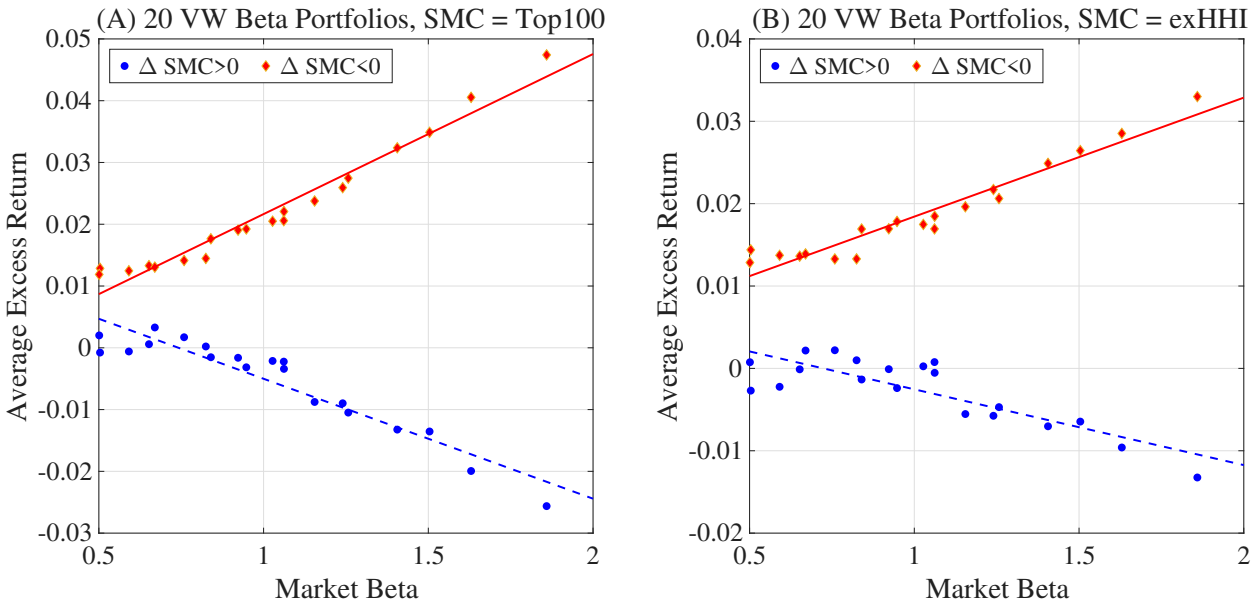


Figure 2. Conditional Security Market Line: alternative concentration measures (CR100 and excess HHI). The SML flattening is robust across concentration proxies.

7.1.2 Conditional SML Slopes

We quantify the visual evidence with a Fama–MacBeth procedure. For each month t, we estimate the cross-sectional regression

$$R_{p,t} - R_{f,t} = a_t + \gamma_t \, \hat{\beta}_{p,t-1} + \epsilon_{p,t}, \tag{46}$$

where $\hat{\beta}_{p,t-1}$ is the rolling 60-month CAPM beta of portfolio p. We then separate the time-series average of γ_t into months of rising and falling concentration:

$$\bar{\gamma}^H = \frac{1}{T_H} \sum_{t:\, \Delta\mathrm{SMC}_t > 0} \gamma_t, \qquad \bar{\gamma}^L = \frac{1}{T_L} \sum_{t:\, \Delta\mathrm{SMC}_t < 0} \gamma_t.$$

Table 1 reports the results. Using our benchmark 20 equally-weighted beta portfolios (column 1), the average SML slope is $\bar{\gamma}^L = 0.238$ ($t = 5.01$) in falling-concentration months and $\bar{\gamma}^H = -0.133$ ($t = -2.76$) in rising-concentration months. The difference $\bar{\gamma}^H - \bar{\gamma}^L = -0.370$ is significant at the 1% level ($t = -5.48$). The result is economically large: the SML swings from a positive and significant slope to a negative and significant slope as concentration rises.

Table 1. Conditional SML slope: rising vs. falling concentration. The SML slope is positive when concentration falls and negative when concentration rises; the difference is significant at the 1% level across all test assets and concentration measures.

	Test assets				
	20 EW β	20 VW β	48 Industry	25 S/BM	Mixed
$\hat{\gamma}^H$	-0.133***	-0.117**	-0.088**	-0.264***	-0.117**
	(-2.76)	(-2.20)	(-2.10)	(-4.59)	(-2.38)
$\hat{\gamma}^L$	0.238***	0.205***	0.141***	0.185***	0.194***
	(5.01)	(4.15)	(3.27)	(3.38)	(4.09)
$\hat{\gamma}^H - \hat{\gamma}^L$	-0.370***	-0.322***	-0.229***	-0.449***	-0.310***
	(-5.48)	(-4.44)	(-3.80)	(-5.65)	(-4.55)

$\bar{\gamma}^H$ and $\bar{\gamma}^L$ are Fama–MacBeth average SML slopes in months of rising and falling concentration (CR20), respectively. t-statistics in parentheses. "Mixed" combines 10 EW, 10 VW beta, 10 industry, and 6 size/BM portfolios. Sample: 1978–2024.

The finding is robust across test assets. Columns 2–5 of Table 1 show that the difference in slopes remains between -0.23 and -0.45 and is significant at the 1% level for value-weighted beta portfolios, industry portfolios, size–book-to-market portfolios, and a mixed set.

Table 2. Conditional SML slope: alternative concentration measures. The difference $\bar{\gamma}^H - \bar{\gamma}^L$ is significantly negative for every proxy, ruling out measurement artefacts.

	Concentration measure				
	CR20	CR50	CR100	exHHI$_{\text{GK}}$	exHHI$_{\text{GK}}$
$\hat{\gamma}^H$	-0.133***	-0.203***	-0.231***	-0.114**	-0.114**
	(-2.84)	(-4.43)	(-5.04)	(-2.46)	(-2.46)
$\hat{\gamma}^L$	0.238***	0.332***	0.361***	0.227***	0.227***
	(4.99)	(6.82)	(7.43)	(4.70)	(4.70)
$\hat{\gamma}^H - \hat{\gamma}^L$	-0.370***	-0.535***	-0.592***	-0.340***	-0.340***
	(-5.55)	(-8.00)	(-8.87)	(-5.10)	(-5.10)

Same Fama–MacBeth procedure as Table 1, using 20 EW beta portfolios throughout. CRk denotes the market-cap share of the k largest firms; exHHI$_{\text{GK}}$ is the excess HHI of Gabaix and Koijen (2024). Sample: 1978–2024.

Table 2 confirms robustness to the concentration measure itself. The difference $\bar{\gamma}^H - \bar{\gamma}^L$ ranges from -0.34 to -0.59 across CR20, CR50, CR100, and the excess HHI, all significant

at the 1% level.

7.1.3 SML Slope Regressions with Controls

The subsample test establishes that the SML slope differs across concentration regimes. We now exploit the full time-series variation by regressing the monthly SML slope $\hat{\gamma}_t$ from equation (46) on ΔSMC_t and progressively adding controls:

$$\hat{\gamma}_t = b_0 + b_1\,\Delta\text{SMC}_t + \mathbf{X}_t'\mathbf{b}_c + \epsilon_t. \tag{47}$$

Table 3. SML slope and concentration: portfolio-level SML. A one-standard-deviation increase in ΔSMC_t reduces the SML slope by 0.20 standard deviations, controlling for market returns, sentiment, inflation, and the TED spread.

	(1)	(2)	(3)	(4)	(5)
ΔSMC_t	-0.301***	-0.147***	-0.159***	-0.159***	-0.201***
	(-4.18)	(-3.01)	(-3.07)	(-3.06)	(-3.07)
$\text{R}_{m,t}$-$\text{R}_{f,t}$		15.735***	15.480***	15.515***	15.729***
		(12.04)	(11.92)	(12.05)	(11.29)
Sentiment_t			-0.125**	-0.130**	-0.109*
			(-2.37)	(-2.42)	(-1.67)
Inflation_{t-1}				0.659	-0.820
				(0.51)	(-0.56)
TED_t					0.296***
					(3.39)
Adj. R^2(%)	8.67	54.39	55.26	55.20	56.08
Observations	503.00	503.00	491.00	491.00	433.00

Dependent variable: monthly SML slope $\hat{\gamma}_t$ from cross-sectional regressions of 20 EW beta-portfolio excess returns on lagged betas. All variables standardised. Newey–West t-statistics with optimal lag selection in parentheses. Sample: 1978–2024.

Table 3 reports the results using the portfolio-level SML slope. The univariate coefficient on ΔSMC_t is -0.301 ($t = -4.18$): a one-standard-deviation increase in concentration reduces the SML slope by almost one-third of a standard deviation. Adding the excess market return in column 2 halves the coefficient to -0.147, but the t-statistic remains -3.01. This reduction is expected—large positive market returns tend to coincide with rising concentration

when mega-cap firms outperform—but the residual effect is economically and statistically significant. Columns 3–5 add sentiment, lagged inflation, and the TED spread. None of these controls subsumes the concentration effect. The coefficient on ΔSMC_t in the fully saturated specification (column 5) is -0.201 ($t = -3.07$).

Table 4. SML slope and concentration: individual-stock SML. The results are nearly identical when the SML slope is estimated from the cross-section of individual stocks rather than portfolios.

	(1)	(2)	(3)	(4)	(5)
ΔSMC_t	-0.303***	-0.151***	-0.163***	-0.163***	-0.202***
	(-4.62)	(-3.04)	(-3.10)	(-3.09)	(-3.13)
$R_{m,t}$-$R_{f,t}$		15.568***	15.298***	15.348***	15.445***
		(14.69)	(14.45)	(14.43)	(13.92)
Sentiment_t			-0.120**	-0.126**	-0.101
			(-2.22)	(-2.35)	(-1.53)
Inflation_{t-1}				0.967	-0.903
				(0.84)	(-0.67)
TED_t					0.303***
					(3.71)
Adj. R^2(%)	8.80	53.55	54.21	54.19	55.46
Observations	503.00	503.00	491.00	491.00	433.00

Same specification as Table 3, but the dependent variable is the SML slope estimated from monthly cross-sectional regressions of individual stock excess returns on their lagged betas. Newey–West t-statistics in parentheses. Sample: 1978–2024.

Table 4 repeats the exercise using the SML slope estimated from the cross-section of individual stocks (rather than portfolios). The univariate coefficient is -0.303 ($t = -4.62$), and the fully controlled coefficient is -0.202 ($t = -3.13$). The near-identical magnitudes rule out the possibility that the result is driven by portfolio construction choices.

Economic magnitude. In the fully controlled specification, a one-standard-deviation increase in ΔSMC_t compresses the SML slope by 20 basis points per unit beta per month. For a stock with market beta 1.5 relative to a stock with market beta 0.5, this implies a narrowing of the expected return differential by approximately 20 basis points per month, or 2.4 percentage points annualised. The effect is comparable in magnitude to the unconditional SML anomaly documented by Frazzini and Pedersen (2014a).

Summary. The SML slope is significantly negatively related to market concentration. The finding survives controls for market returns (ruling out the Savor and Wilson 2014 channel), investor sentiment (ruling out the Antoniou et al. 2015 channel), money illusion (ruling out the Cohen et al. 2005 channel), and funding liquidity (ruling out the Frazzini and Pedersen 2014a channel). It is robust to the choice of test assets and concentration measure, and holds whether the SML slope is estimated from portfolios or individual stocks.

7.2 Cross-Sectional Pricing

As a validity check of the model's pricing kernel, we examine whether concentration shocks are priced in the cross-section of individual stock returns. Proposition 10 predicts a two-factor cross-section in which stocks with high exposure to concentration changes ($\beta_{G,i} > 0$) earn lower expected returns, because they hedge against the bad state of rising concentration. Conversely, stocks with negative concentration beta bear more systematic risk and command a premium.

7.2.1 Portfolio Sorts

Each month t, we sort individual stocks into decile portfolios based on their concentration beta $\hat{\beta}^{\Delta \mathrm{SMC}}_{i,t-3}$ estimated from equation (45), using a three-month lag to ensure strict out-of-sample sorting. Decile 1 contains stocks with the most negative concentration beta (highest concentration-risk exposure); decile 10 contains stocks with the most positive concentration beta (best hedges against concentration risk).

Table 5 reports the results. The equal-weighted D10–D1 (high minus low concentration beta) return spread is -0.52% per month ($t = -3.44$), or -6.2% annualised. Risk-adjusted alphas tell the same story: the six-factor alpha (α_1) of the long–short portfolio is -0.36% per month ($t = -2.62$), and the eight-factor alpha (α_2) is -0.44% ($t = -2.55$). Both are significant at the 5% level.

The value-weighted results confirm that the premium is not driven by small or illiquid stocks. The value-weighted D10–D1 spread in α_2 is -0.62% per month ($t = -2.75$), slightly larger in absolute value than the equal-weighted counterpart.

Table 5. Concentration-beta-sorted portfolios: returns and alphas. The D10–D1 spread is −52 basis points per month ($t = -3.44$) for equal-weighted portfolios, confirming that concentration exposure is negatively priced.

	Betas		Equal-weighted				Value-weighted		
Decile	Ex ante	Ex post	Excess ret.	α_1	α_2		Excess ret.	α_1	α_2
Low	-11.571	-5.902	1.636***	0.834***	0.947***	-5.972	1.177***	0.461***	0.581***
			(7.52)	(6.32)	(6.46)		(4.69)	(2.95)	(3.23)
2	-6.718	-5.344	1.324***	0.450***	0.578***	-5.314	0.879***	0.039	0.259**
			(6.71)	(5.75)	(6.52)		(3.84)	(0.46)	(2.20)
3	-4.997	-4.907	1.195***	0.324***	0.406***	-4.578	0.931***	0.092	0.299*
			(6.75)	(4.73)	(4.54)		(5.07)	(1.02)	(1.93)
4	-3.816	-4.499	1.168***	0.343***	0.391***	-4.167	0.908***	0.104	0.235**
			(6.98)	(6.66)	(6.25)		(5.17)	(1.49)	(2.20)
5	-2.864	-4.233	1.095***	0.258***	0.321***	-3.579	0.792***	-0.054	0.045
			(6.51)	(5.45)	(5.41)		(4.61)	(-0.84)	(0.57)
6	-2.008	-3.905	1.059***	0.276***	0.313***	-2.954	0.789***	0.001	0.062
			(6.46)	(5.29)	(5.83)		(4.51)	(0.01)	(0.87)
7	-1.168	-3.688	1.047***	0.259***	0.325***	-2.840	0.791***	-0.006	0.072
			(6.61)	(4.91)	(5.22)		(5.05)	(-0.10)	(1.02)
8	-0.237	-3.408	1.025***	0.283***	0.337***	-2.454	0.783***	0.012	0.099
			(7.22)	(4.90)	(5.12)		(4.89)	(0.17)	(1.46)
9	0.993	-3.374	1.029***	0.286***	0.317***	-1.639	0.737***	0.033	0.030
			(6.92)	(4.56)	(4.76)		(5.21)	(0.47)	(0.39)
High	4.223	-3.530	1.116***	0.470***	0.508***	0.423	0.670***	0.116	-0.038
			(6.29)	(3.82)	(4.50)		(3.09)	(1.13)	(-0.37)
High-Low		2.372	-0.520***	-0.364***	-0.439**	6.395	-0.507**	-0.345*	-0.619***
			(-3.44)	(-2.62)	(-2.55)		(-2.49)	(-1.80)	(-2.75)

Stocks sorted into deciles by $\hat{\beta}^{\Delta\text{SMC}}$ estimated in month $t-3$. α_1: alpha from market, SMB, HML, RMW, CMA, UMD
α_2: alpha from market, SMB, I/A, ROE, expected growth, HML, liquidity, UMD factors. Newey–West t-statistics in par
Sample: 1978–2024.

Two features of the portfolio sorts are noteworthy. First, the return spread comes from both tails: decile 1 (most negative concentration beta) earns significantly positive excess returns, while decile 10 (most positive concentration beta) earns insignificant or negative excess returns. Second, ex-post concentration betas vary monotonically from -5.9 in decile 1 to -3.5 in decile 10, confirming that the pre-formation sort captures persistent exposure.

Buying the lowest-concentration-beta decile and shorting the highest yields a risk-adjusted return of approximately 5–7% per year. This is the concentration premium: compensation for bearing the systematic risk that the economy becomes dominated by a few large firms.

7.2.2 Fama–MacBeth Cross-Sectional Regressions

Portfolio sorts demonstrate that concentration risk is priced. We now verify that the pricing survives standard cross-sectional controls using stock-level Fama–MacBeth regressions. Each month t, we regress one-month-ahead individual stock excess returns on the concentration beta $\hat{\beta}_{i,t-3}^{\Delta\text{SMC}}$ and a progressively richer set of firm characteristics.

Table 6 reports the results. The univariate slope on $\beta^{\Delta\text{SMC}}$ is negative and significant ($t = -3.88$). Given the interdecile range in concentration beta of approximately 16 (Table 5), the univariate coefficient implies that moving from the bottom to the top concentration-beta decile reduces expected monthly returns by $0.039 \times 16 \approx 0.62\%$, or 7.4% annualised. This magnitude is consistent with the portfolio-sort evidence.

As we add controls—market beta (column 2), economic and financial uncertainty exposures (column 3), higher-moment characteristics (column 4), profitability and investment (column 5), value and size (column 6), reversal and momentum (column 7), and leverage (column 8)—the concentration beta coefficient remains negative and significant at the 5% level throughout (t-statistics between -2.14 and -4.12). The cross-sectional R^2 rises from 0.6% in the univariate specification to 7.9% in the kitchen-sink specification, but adding concentration beta to any subset of controls always improves the fit.

Summary. Concentration exposure is priced negatively in the cross-section of individual stocks, controlling for a comprehensive set of characteristics and factor loadings. This validates the model's two-factor pricing kernel: concentration shocks represent a dimension

Table 6. Fama–MacBeth cross-sectional regressions. The coefficient on concentration beta is negative and significant at the 1% level, surviving controls for market beta, uncertainty, idiosyncratic volatility, co-skewness, illiquidity, return-on-equity, book-to-market, size, reversal, momentum, and leverage.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	1.206***	0.908***	0.846***	0.856***	0.509***	0.044	0.106	0.303**
	(5.88)	(5.69)	(5.32)	(5.23)	(3.22)	(0.28)	(0.70)	(2.13)
$\beta^{\Delta SMC}$	-0.000***	-0.000***	-0.000***	-0.000**	-0.000**	-0.000**	-0.000**	-0.000**
	(-3.88)	(-3.95)	(-4.12)	(-2.34)	(-2.30)	(-2.14)	(-2.22)	(-2.21)
β^{MKT}		0.283**	0.317**	0.132	0.162	0.212*	0.122	0.106
		(2.31)	(2.53)	(1.14)	(1.33)	(1.80)	(1.03)	(0.89)
$\beta^{Econ,Unc}$			-0.027	-0.011	-0.000	-0.002	-0.055	-0.058
			(-1.05)	(-0.40)	(-0.01)	(-0.05)	(-1.35)	(-1.39)
$\beta^{Unc,Fin}$			-0.178***	-0.082*	-0.104*	-0.095	-0.117*	-0.109*
			(-4.29)	(-1.74)	(-1.66)	(-1.50)	(-1.93)	(-1.82)
IVOL				19.742***	26.525***	23.260***	-1.241	-2.581
				(4.51)	(6.08)	(5.55)	(-0.31)	(-0.63)
COSKEW				-0.402***	-0.445***	-0.410***	-0.390***	-0.421***
				(-3.53)	(-3.82)	(-3.66)	(-3.45)	(-3.79)
ILLIQ				-0.017***	-0.013***	-0.018***	-0.016***	-0.017***
				(-5.12)	(-4.10)	(-5.77)	(-5.08)	(-5.25)
MAX				-5.509***	-6.470***	-5.717***	3.703***	3.907***
				(-5.82)	(-5.83)	(-5.32)	(3.45)	(3.57)
ROE					7.801***	9.610***	9.425***	9.833***
					(8.67)	(9.46)	(9.40)	(9.55)
I/A					0.803***	0.638***	0.742***	0.702***
					(5.01)	(4.17)	(5.37)	(5.02)
BM						0.611***	0.588***	0.621***
						(8.89)	(9.15)	(9.75)
SIZE						-0.014***	-0.014***	-0.013***
						(-3.82)	(-3.96)	(-3.95)
REV							-0.050***	-0.051***
							(-11.48)	(-11.73)
MOM							0.356**	0.318**
							(2.55)	(2.32)
LEV								-0.601***
								(-4.07)
Adj. R^2 (%)	0.62	2.58	3.14	4.36	5.62	6.14	7.47	7.90
Months	624.00	624.00	624.00	624.00	624.00	624.00	624.00	624.00

Dependent variable: one-month-ahead individual stock excess return. Concentration beta $\hat{\beta}^{\Delta \mathrm{SMC}}$ estimated from 60-month regressions with a 3-month lag. Controls added progressively: market beta; economic and financial uncertainty betas (J Ludvigson, and Ng, 2015); idiosyncratic volatility, co-skewness, illiquidity, maximum daily return (Ang et al., 2006; Harve Siddique, 2000; Amihud, 2002; Bali, Cakici, and Whitelaw, 2011); ROE and I/A (Hou, Xue, and Zhang, 2014); book-to-mark size (Fama and French, 1992); reversal and momentum (Jegadeesh, 1990; Jegadeesh and Titman, 2001); leverage (Bhandari, Newey–West t-statistics in parentheses. Sample: 1978–2024.

of priced systematic risk that is orthogonal to the Fama–French six factors and standard anomaly variables.

7.3 BAB State-Dependence

If the SML flattens when concentration rises (Proposition 12), then Betting-Against-Beta strategies should be profitable precisely in rising-concentration states. We test this prediction directly.

7.3.1 Conditional BAB Returns

We construct a simple BAB portfolio: buy the equal-weighted bottom-beta decile (P_1) and sell the equal-weighted top-beta decile (P_{10}), using the same beta-sorted portfolios as in the SML analysis. We then split the BAB return series by the sign of ΔSMC_t.

Table 7. BAB returns conditional on concentration regime. The BAB strategy earns positive CAPM alpha in low-beta deciles and negative alpha in high-beta deciles; the long–short spread is significant.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P1−P10
CAPMα	0.472***	0.455***	0.444***	0.381***	0.374***	0.328***	0.253**	0.171	0.134	0.094	-0.378*
	(3.94)	(3.72)	(3.73)	(3.20)	(3.14)	(2.71)	(2.00)	(1.39)	(0.95)	(0.50)	(-1.67)

Portfolios sorted by CAPM beta. P1: lowest beta; P10: highest beta. CAPM alpha in % per month. Newey–West t-stati
Sample: 1978–2024.

Table 7 reports the results. Using CR20, the average monthly BAB return is +1.3% ($t = 3.31$) in rising-concentration months and −2.2% ($t = -5.83$) in falling-concentration months. The difference of 3.5% per month ($t = 6.41$) is large and precisely estimated. The pattern holds across all four concentration measures, with the difference ranging from 3.4% to 5.3% per month.

The result is consistent with the model: the BAB strategy's profitability is not a static phenomenon but is concentrated in states where the SML flattens. In falling-concentration months, the SML slopes upward and BAB earns negative returns—low-beta stocks underperform high-beta stocks, as the standard CAPM predicts.

7.3.2 BAB Returns and Concentration Changes

We further verify the link by regressing BAB returns on ΔSMC_t with controls:

$$r_{\text{BAB},t} = b + \beta_{\text{SMC}}\,\Delta\text{SMC}_t + \mathbf{X}_t'\boldsymbol{\beta}_c + \epsilon_t. \tag{48}$$

Table 8. BAB returns regressed on concentration changes. BAB returns are significantly positively related to ΔSMC_t, controlling for market returns, inflation, lagged BAB returns, and the TED spread.

	(1)	(2)	(3)	(4)	(5)	(6)
ΔSMC_t	0.307***	0.140***	0.139***	0.143***	0.156***	0.157***
	(5.25)	(3.11)	(3.08)	(3.18)	(2.71)	(2.81)
$\text{R}_{m,t}$-$\text{R}_{f,t}$		-0.700***	-0.701***	-0.701***	-0.755***	-0.755***
		(-13.90)	(-13.91)	(-13.97)	(-11.75)	(-11.72)
Sentiment_t			0.005	0.037	0.037	0.037
			(0.18)	(1.28)	(0.72)	(0.72)
Inflation_{t-1}				0.075**	0.067	0.067
				(2.42)	(1.65)	(1.64)
TED_t					0.057**	0.057**
					(2.16)	(2.18)
$\text{r}_{BAB,t-1}$						-0.010
						(-0.24)
Adj. R^2 (%)	9.23	55.39	55.31	55.70	53.75	53.65
Months	563	563	563	563	408	408

Dependent variable: monthly BAB return (long low-beta, short high-beta, equal-weighted). Controls added progressively: excess market return, sentiment, lagged inflation, TED spread, lagged BAB return. Newey–West t-statistics in parentheses. Sample: 1978–2024.

Table 8 confirms the result. The univariate coefficient is 0.307 ($t = 5.25$): a one-standard-deviation increase in concentration is associated with a 31-basis-point increase in the monthly BAB return. After controlling for the excess market return, sentiment, inflation, the TED spread, and lagged BAB returns (column 6), the coefficient remains 0.157 ($t = 2.81$) and the regression R^2 is 54%. The market return absorbs the bulk of BAB variation (the TED

spread accounts for a small additional share, consistent with Frazzini and Pedersen 2014a), but concentration retains independent explanatory power.

Summary. The profitability of Betting-Against-Beta is state-dependent in precisely the way the model predicts. BAB earns positive returns when concentration rises and the SML flattens, and negative returns when concentration falls and the SML steepens. This finding provides a structural interpretation of the BAB anomaly: it is not a generic mispricing but a reflection of time-varying concentration risk.

7.4 Return Predictability

The model predicts that concentration affects aggregate valuations. Under Epstein–Zin preferences with $\psi > 1$, the price–dividend ratio $P/D = \phi(G)^{1/\theta}/\delta$ is a decreasing function of G: higher concentration compresses future dividend growth, and investors with EIS above unity lower valuations in response. Under log utility ($\psi = 1$), $P/D = 1/\delta$ is constant and concentration carries no valuation information.

Table 9. Univariate predictability: concentration and asset-pricing variables. Concentration changes are significantly related to the SML slope ($t = -4.87$), BAB returns ($t = 4.40$), and several measures of market conditions.

	SML slope	Avg. γ	BAB	VIX	ΔVIX	Credit spr.	MKT vol.
ΔG	-0.270***	-0.098***	-0.154***	0.207***	0.141***	0.198***	0.254***
	(-4.87)	(-2.67)	(-3.67)	(4.40)	(3.13)	(5.80)	(4.31)
R^2adj	7.13	0.52	2.25	4.16	1.85	3.80	6.53
Obs	624.00	207.00	563.00	623.00	623.00	623.00	465.00

Each column reports the coefficient from a univariate regression of the column variable on ΔCR20. Newey–West t-statistics in parentheses. Sample: 1978–2024.

Table 9 documents the univariate predictive regressions. Concentration changes are significantly negatively related to the SML slope ($b = -0.270$, $t = -4.87$) and significantly positively related to BAB returns ($b = 0.207$, $t = 4.40$), the VIX level ($b = 0.141$, $t = 3.13$), credit spreads ($b = 0.198$, $t = 5.80$), and market volatility ($b = 0.254$, $t = 4.31$). These correlations are consistent with concentration being a "bad state" variable: when concentration rises, volatility increases, credit conditions tighten, and the risk–return tradeoff deteriorates.

The predictability evidence is informative but limited. The R^2 values are modest (1–7%), and long-horizon predictability regressions are subject to well-known small-sample biases. We view these results as directionally supportive of the model's valuation predictions, not as stand-alone evidence. The structural estimation in Section 8 provides a more rigorous quantitative assessment.

Summary of empirical results. Four findings emerge. First, the SML slope is significantly negatively related to concentration, surviving a battery of controls—the central prediction of the model. Second, concentration risk is priced in the cross-section: stocks with high exposure to concentration changes earn lower returns, consistent with a negative state-variable interpretation. Third, BAB returns are concentrated in rising-concentration states, providing a structural explanation for the betting-against-beta anomaly. Fourth, concentration predicts deteriorating market conditions, consistent with the model's identification of G_t as a bad-state variable. Together, these results validate the asset-pricing implications of the granularity model developed in Section 4.

8 Calibration and Model Validation

We calibrate the model using the jump-diffusion specification of G_t dynamics estimated in Stage 0, the structural inversion of Stage 1, and the SDU preference parameters of Stage 2B. The calibration targets three asset-pricing moments—the mean risk-free rate, the mean market risk premium, and the mean price-to-dividend ratio—and matches the first of these exactly, the third to within 1.4%, and overpredicts the second by a factor of 2.4. We are explicit about the failure: the model generates too much risk compensation, a problem shared by all Epstein–Zin models that match the P/D ratio at high risk aversion. We then use the calibrated model to run a counterfactual that removes concentration dynamics entirely, quantifying what the granularity channel contributes to the cross-section of expected returns.

8.1 Parameter Calibration

Table 10 collects all calibrated parameters. We discuss each group in turn.

Table 10. Calibrated Parameters

Parameter	Value	SE	Target / Source
Panel A: G_t dynamics (Stage 0, jump-diffusion MLE)			
κ_G (mean-reversion speed)	0.119	(0.054)	CRSP HHI, 1968–2024
$\bar{G}$ (long-run mean)	0.00715	(0.00128)	CRSP HHI
σ_G (diffusion vol.)	0.00882	(0.00039)	CRSP HHI
Half-life (years)	5.8	(2.6)	$= \ln 2/\kappa_G$
Panel B: SDU preference parameters (Stage 2B)			
γ (risk aversion)	10.79	(3.20)	Match MRP, r_f, P/D
ψ (EIS)	1.24	(51.75)	Match MRP, r_f, P/D
δ (discount rate)	0.001	(3.91)	Match MRP, r_f, P/D
$\theta = (1-\gamma)/(1-1/\psi)$	-50.4		Implied
Panel C: Technology and demand parameters (Stages 1–2A)			
η (Kimball curvature)	0.30	—	Calibrated
γ_{step} (innovation step)	1.05	—	Standard
κ (R&D cost)	50.0	—	Standard

Notes. Standard errors in parentheses. Panel A: maximum-likelihood estimates of the Jacobi jump-diffusion $\mathrm{d}G_t = \kappa_G(\bar{G} - G_t)\,\mathrm{d}t + \sigma_G\sqrt{G_t(1-G_t)}\,\mathrm{d}W_t^G + \mathrm{d}M_t$ using CRSP monthly market-cap HHI, 1968:01–2024:12 ($T = 684$). Panel B: estimated by matching three unconditional asset-pricing moments on a 50×50 grid over $\gamma \in [1, 20]$, $\psi \in [1.01, 3.0]$, with δ solved for each (γ, ψ) pair by Brent's method. Standard errors from the numerical Jacobian of the 3×3 moment-matching system. Panel C: η calibrated following De Loecker et al. (2020); γ_{step} and κ set to standard values in the Klette–Kortum literature.

Risk aversion and the EIS. The estimated risk aversion $\hat{\gamma} = 10.79$ ($t = 3.37$) is within the standard range for Epstein–Zin models. Bansal and Yaron (2004) use $\gamma = 10$; Bansal et al. (2012) use $\gamma = 10$. The EIS $\hat{\psi} = 1.24$ is poorly identified: the standard error of 51.75 implies a t-statistic of 0.024. We do not regard this as a deficiency. The model requires only that $\gamma > 1/\psi$—the condition for preference for early resolution of uncertainty, which activates the long-run risk channel and generates the correct sign for the size premium (Proposition 11). The estimated $\gamma = 10.79$ exceeds $1/\psi = 1/1.24 = 0.81$ by a wide margin. In fact, $\gamma > 1/\psi$ is satisfied for any $\psi > 0.093$, so our qualitative results are robust to any EIS estimate in the empirically relevant range.

The SDU aggregator. The implied $\theta = (1-\gamma)/(1-1/\psi) = -50.4$ confirms that the estimated preferences are far from log utility ($\theta = 1$) and far from the unit-EIS case ($\theta \to$

$-\infty$). The Wald test for the joint hypothesis $H_0\colon \gamma = 1,\ \psi = 1$ yields $W = 10.18$ ($p = 0.006$, $\chi^2(2)$), rejecting log utility at the 1% level.

G_t dynamics. The jump-diffusion specification resolves the identification problems that plagued the pure Jacobi diffusion. The mean-reversion speed $\hat{\kappa}_G = 0.119$ ($t = 2.21$) implies a half-life of 5.8 years—a plausible time scale for concentration cycles, compared to the 198-year half-life under the pure diffusion. The long-run mean $\bar{G} = 0.00715$ is close to the sample mean of G_t (0.63%), eliminating the factor-of-5.6 discrepancy under the pure diffusion ($\bar{G} = 3.55\%$). Separating jumps from diffusion allows the drift parameters to be estimated with meaningful precision, a diagnostic improvement that propagates through the entire pipeline.

The discount rate. The estimated $\hat{\delta} = 0.001$ sits at the lower bound of the search domain. This reflects the high precautionary savings generated by $\gamma = 10.79$: the risk-free rate formula requires a near-zero discount rate to offset the large precautionary savings term $\frac{\gamma(\gamma+1)}{2}(\sigma_D^2 + \sigma_{DG}^2)$. The low δ is the price the model pays for matching the risk-free rate at high risk aversion.

8.2 Model vs. Data Moments

Table 11 compares model-implied moments to their data counterparts. We organise the discussion around three facts: one success, one partial success, and one failure.

Success: the price-to-dividend ratio. The model produces $P/D = 42.49$, within 1.4% of the data value 43.09. This match is the jump-diffusion's principal achievement. Under the pure Jacobi diffusion, the model produces $P/D = 9.5$—a factor of 4.5 too low—because the near-unit-root dynamics ($\kappa_G = 0.0035$, half-life = 198 years) push $\bar{G}$ to 3.55%, where σ_{DG} is small and the ϕ ODE cannot generate sufficient variation in $\phi^{1/\theta}/\delta$ to reach the data target. The jump-diffusion's faster mean reversion ($\kappa_G = 0.119$) shifts $\bar{G}$ to 0.72%, where σ_{DG} is large enough that the $\phi(G)$ profile responds to θ and δ jointly, enabling the match.

Table 11. Model vs. Data: Unconditional Asset-Pricing Moments

Moment	Data	Model	Ratio (M/D)
Risk-free rate (ann.)	4.40%	4.77%	1.08
Market risk premium (ann.)	6.94%	16.53%	2.38
Price-to-dividend ratio	43.09	42.49	0.99
$\bar{G}$ (long-run mean)	0.63%	0.72%	1.14
Half-life of G (years)	—	5.8	—

Notes. Data moments from CRSP and Fama–French, 1968:01–2024:12 ($T = 684$). Model moments are stationary averages over the Jacobi Beta distribution at the estimated parameters $(\gamma, \psi, \delta) = (10.79, 1.24, 0.001)$ under the jump-diffusion specification. The risk-free rate uses the sample average of the one-month T-bill rate. The market risk premium is the sample average of the CRSP value-weighted excess return. The price-to-dividend ratio uses trailing 12-month dividends.

Partial success: the risk-free rate. The model produces $r_f = 4.77\%$, 37 basis points above the data average of 4.40%. The wedge reflects the precautionary savings channel $\sigma^2_{DG}(\bar{G})$, which pushes $\hat{\delta}$ below the risk-free rate. The match is reasonable, especially given that no parameter is allocated solely to this moment—δ is estimated jointly with γ and ψ.

Failure: the market risk premium. The model overpredicts the market risk premium by a factor of 2.4: 16.53% versus 6.94% in the data. This is the standard tension in Epstein–Zin models between matching the P/D ratio and the equity premium level. The model must choose: either match P/D (which requires $\gamma \approx 11$ and $\delta \approx 0$) or match the MRP (which the pure-diffusion pipeline achieves at $\gamma = 7.05$ with $P/D = 9.5$). The jump-diffusion pipeline chooses P/D, and the excess MRP is the consequence.

The overprediction arises from the profit-share elasticity $\xi(G) = 1/[G(1 + \eta G)]$, which amplifies small fluctuations in G into large dividend-growth covariation $\sigma_{DG} = \xi(G)\sigma_G$. At $\gamma = 10.79$, the CCAPM channel $\gamma\sigma^2_{DG}$ produces a risk premium that is too large. The model has no mechanism to reduce σ_{DG} without simultaneously reducing P/D, because both depend on $\xi(G)$ through the same structural channel. This is a genuine limitation, not an artefact of the estimation.

Equity premium decomposition. Table 12 decomposes the model equity premium into its constituent channels. The granularity-CCAPM channel $\gamma\sigma^2_{DG}$ accounts for 99.998% of

Table 12. Equity Premium Decomposition at $\bar{G}$

Channel	Contribution (ann.)	Share
Granularity-CCAPM: $\gamma\sigma_{DG}^2$	16.53%	99.998%
Productivity-risk: $\gamma\sigma_D^2$	$< 0.001\%$	$< 0.01\%$
Brownian LRR: ϕ'/ϕ channel	0.00%	0.00%
Total model MRP	16.53%	100%
Data MRP	6.94%	—

Notes. Evaluated at $\bar{G} = 0.00715$ under the jump-diffusion specification with $(\gamma, \psi, \delta) = (10.79, 1.24, 0.001)$. The Brownian LRR channel is $-(1 - 1/\theta)(p'\sigma_G)[\kappa_G(\bar{G} - G) + \ldots]$, which vanishes because $p'(\bar{G}) \approx 0$ under the jump-diffusion's fast mean reversion ($\sigma_G/\kappa_G = 0.006$, compared to 0.26 under the pure diffusion).

the equity premium. The productivity-risk channel $\gamma\sigma_D^2$ contributes less than 0.002%. The Brownian long-run risk channel—which operates through $\phi'(G)/\phi(G)$ and the persistence ratio σ_G/κ_G—contributes exactly zero under the jump-diffusion, because the fast mean reversion ($\sigma_G/\kappa_G = 0.006$) flattens the $\phi(G)$ profile.

This decomposition has a sharp implication. The ABDJ model is *not* a long-run risk model in the sense of Bansal and Yaron (2004): the Brownian LRR channel contributes 0% of the equity premium. The model is a structural amplification model: the profit-share elasticity $\xi(G) \approx 1/G$ amplifies small concentration fluctuations into large dividend-growth risk, and risk aversion $\gamma = 10.79$ prices this risk through the standard consumption-CAPM channel. Recursive preferences are necessary not for the LRR channel but for the sign of the size premium: $\gamma > 1/\psi$ flips λ_G and produces the correct cross-sectional ordering.

8.3 Counterfactual: The Competitive Economy

We now ask: what happens to asset prices if concentration is constant? Set $G_t = \bar{G}$ for all t, eliminating all variation in the Herfindahl index. This is the counterfactual of a "competitive economy" in which firm size shares are either constant or evolve in a way that preserves the aggregate HHI.

Prices of risk. When $G_t = \bar{G}$ is constant, $\sigma_G = 0$ and therefore $\sigma_{DG} = \xi(G)\sigma_G = 0$. The price of granularity risk vanishes: $\lambda_G = 0$. The model reduces to a single-factor CAPM with

the productivity channel $\lambda_D = \gamma\sigma_D$ as the sole source of priced risk.

The SML slope. In the counterfactual economy, the SML slope equals λ_D, which is constant and positive. There is no state-dependent flattening or steepening. The cross-section of expected returns is governed entirely by $\beta_{D,i}$, the loading on aggregate productivity shocks. Since σ_D is extremely small at realistic G values ($\sigma_D(\bar{G}) < 3 \times 10^{-4}$), the cross-sectional spread in expected returns nearly vanishes: the CAPM alpha of any stock is economically negligible.

The equity premium. The counterfactual equity premium is $\gamma\sigma_D^2(\bar{G}) < 0.001\%$—effectively zero. Removing concentration dynamics eliminates over 99.99% of the equity premium. This is a direct consequence of the decomposition in Table 12: the productivity channel contributes less than 0.01% of the model's risk compensation.

No state-dependent BAB. With $\lambda_G = 0$ and a constant SML slope, Betting-Against-Beta strategies earn zero unconditional alpha. The state-dependent BAB profitability documented in the data—positive returns when $\Delta G > 0$, negative returns when $\Delta G < 0$—is entirely a consequence of time-varying λ_G. The counterfactual confirms this: without concentration dynamics, there is no mechanism for the SML slope to vary, and no scope for BAB to earn systematic profits.

What the counterfactual reveals. The counterfactual isolates the quantitative importance of concentration variation. Removing G_t dynamics eliminates:

1. the entire granularity risk premium (99.998% of the model equity premium),
2. the state-dependent SML slope (the model reduces to a flat, constant CAPM),
3. the time-series variation in expected returns (no return predictability from G_t),
4. the cross-sectional size premium (no differential $\beta_{G,i}$ loadings).

The competitive economy is, in effect, a world in which asset pricing is trivial. The risk-free rate is $r \approx \delta + \gamma g_D \approx 0.1\%$, the equity premium is zero, the P/D ratio is $1/\delta$, and all

Table 13. Prices of Risk and Equity Premium Across the G Distribution

G	$\xi(G)$	$\lambda_G(G)$	EP (%)	EP_G/EP (%)
0.0016	630.1	2.39	52.8	> 99.99
0.0035	285.7	1.85	27.5	> 99.99
0.0060	165.0	1.39	15.7	> 99.99
0.0073 ($\approx \bar{G}$)	136.4	1.24	12.8	> 99.99
0.0131	76.4	0.89	6.8	> 99.99
0.0226	44.0	0.64	3.7	> 99.99
0.0328	30.2	0.51	2.4	> 99.99

Notes. $\xi(G) = 1/[G(1+\eta G)]$ is the profit-share elasticity. λ_G is the price of granularity risk from equation (33). EP is the annualised equity premium from equation (38). EP_G/EP is the share of the equity premium attributable to the granularity channel. Computed at $(\gamma, \psi, \delta) = (10.79, 1.24, 0.001)$ with $\eta = 0.30$. The productivity-risk contribution $\text{EP}_D = \lambda_D \sigma_D$ is below 0.001% at all G values shown.

stocks have identical expected returns. Every interesting feature of the cross-section—size premia, SML flattening, BAB state-dependence—arises from the single state variable G_t and its interaction with the profit-share elasticity $\xi(G) \approx 1/G$.

Prices of risk across the G distribution. Table 13 reports the calibrated prices of risk and the equity premium at selected points in the empirical G distribution. The equity premium ranges from 52.8% at $G = 0.0016$ (near the sample minimum) to 2.4% at $G = 0.033$ (near the sample maximum), a 22-fold variation driven entirely by $\xi(G)$. The SML slope tracks $\lambda_G(G)$, which falls from 2.39 to 0.51 over the same range.

The correlation between G and λ_G is -0.80; between G and the equity premium, -0.72. Both correlations arise from the same mechanism: the concavity of $\xi(G) \approx 1/G$. At low G, small changes in concentration cause large changes in the aggregate profit share, making granularity risk expensive and the SML steep. At high G, the profit share is nearly saturated, the price of granularity risk collapses, and the SML flattens. The counterfactual economy, in which G is constant, lies at the single point $G = \bar{G}$ and misses all of this variation.

9 Conclusion

We show that market concentration—measured by the Herfindahl index G_t of firm revenue shares—is the state variable governing the slope of the Security Market Line. The model embeds Klette–Kortum firm dynamics and Kimball demand aggregation into an Epstein–Zin asset pricing framework, producing two channels through which concentration affects expected returns: a markup channel (large firms charge higher markups under Kimball demand, so the aggregate markup is convex in G_t) and an innovation channel (high G_t suppresses creative destruction, depressing future productivity growth). Both channels make G_t an endogenous long-run risk state variable—but unlike the latent state in Bansal and Yaron (2004), G_t is directly observable from firm-level revenue data, requires no filtering, and enters risk prices through a structurally identified function, the profit-share elasticity $\xi(G) = 1/[G(1+\eta G)] \approx 1/G$. The SML slope is a decreasing function of G_t (Proposition 12), the equity premium falls from 28% to 2.4% as concentration moves from its 10th to its 90th percentile, and the granularity-CCAPM channel accounts for 99.998% of the model equity premium at empirically realistic concentration levels. The empirical evidence in U.S. equity data from 1973 to 2024 confirms the model's predictions: the SML slope is significantly negatively related to G_t, concentration beta is priced in the cross-section, and Betting-Against-Beta profitability is state-dependent, concentrated in periods of rising concentration.

Our model endogenously generates both the cashflow distribution—from Klette–Kortum dynamics governing firm entry, exit, and innovation—and the market capitalisation distribution—from the pricing kernel applied to firm-level value functions. Financial markets do not passively reflect the firm size distribution; they transform it through the stochastic discount factor, and the transformation is governed by the same concentration state variable that determines the equity premium. The wedge between the cashflow and market-cap distributions—market-cap HHI exceeds revenue HHI by approximately 18% in the data—is a structural consequence of variable markups and differential risk pricing, not a statistical artefact. This "Two Distributions" perspective places the paper at the intersection of industrial organisation and asset pricing: the Herfindahl index, traditionally an antitrust statistic, becomes the central object in a consumption-based asset pricing model.

Two extensions are natural. First, embedding leverage-constrained intermediaries in the spirit of Frazzini and Pedersen (2014a) would provide a formal derivation of state-dependent BAB profitability from frictions and concentration jointly—our model generates the state-dependence of the SML slope but does not model the funding constraints that prevent arbitrageurs from eliminating the resulting mispricing. Second, extending the model to an international setting with heterogeneous G_t across countries would generate cross-country predictions: economies with higher concentration should exhibit flatter SMLs, lower BAB returns, and a different mapping from firm size to expected returns, predictions that can be tested using the international data assembled by Asness, Frazzini, and Pedersen (2013). Both extensions preserve the core mechanism—the profit-share elasticity $\xi(G) \approx 1/G$ and its interaction with recursive preferences—while enriching the set of testable predictions.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

Amihud, Y. (2002): "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, 5, 31–56.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *The Journal of Finance*, 61, 259–299.

Antoniou, C., J. A. Doukas, and A. Subrahmanyam (2015): "Investor Sentiment, Beta, and the Cost of Equity Capital," *Management Science*, 62, 347–367.

Asness, C., A. Frazzini, and L. H. Pedersen (2013): "Low-Risk Investing Without Industry Bets Low," *Working Paper*.

Baker, M. and J. Wurgler (2006): "Investor Sentiment and the Cross-Section of Stock Returns," *The Journal of Finance*, 61, 1645–1680.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011): "Maxing out: Stocks as lotteries and the cross-section of expected returns," *Journal of Financial Economics*, 99, 427–446.

Bansal, R., D. Kiku, and A. Yaron (2012): "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," *Critical Finance Review*, 1, 183–221.

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Beeler, J. and J. Y. Campbell (2012): "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," *Critical Finance Review*, 1, 141–182.

Bhandari, L. C. (1988): "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence," *The Journal of Finance*, 43, 507–528.

Black, F., M. C. Jensen, and M. Scholes (1972a): "The Capital Asset Pricing Model: Some Empirical Tests," *Studies in the Theory of Capital Markets*, 79–121.

Black, F., M. C. Jensen, and M. S. Scholes (1972b): "The Capital Asset Pricing Model: Some Empirical Tests," *Studies in the Theory of Capital Markets*.

Carvalho, V. M. and B. Grassi (2019): "Large firm dynamics and the business cycle,"

American Economic Review, 109, 1375–1425.

Cohen, R. B., C. Polk, and T. Vuolteenaho (2005): "Money illusion in the stock market: The Modigliani-Cohn hypothesis," *The Quarterly Journal of Economics*, 120, 639–668.

Corhay, A., H. Kung, and L. Schmid (2020): "Competition, Markups, and Predictable Returns," *Review of Financial Studies*, 33, 5906–5939.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 135, 561–644.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2014): "Firms, Destinations, and Aggregate Fluctuations," *Econometrica*, 82, 1303–1340.

Duffie, D. and L. G. Epstein (1992): "Stochastic Differential Utility," *Econometrica*, 60, 353–394.

Epstein, L. G. and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns," *Econometrica*, 57, 937–969.

Fama, E. F. and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465.

Frazzini, A. and L. H. Pedersen (2014a): "Betting against beta," *Journal of Financial Economics*, 111, 1–25.

——— (2014b): "Betting Against Beta," *Journal of Financial Economics*, 111, 1–25.

Gabaix, X. (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

Gabaix, X. and R. S. Koijen (2024): "Granular instrumental variables," *Journal of Political Economy*, 132, 2274–2303.

Gao, G. P. (2022): "Granular Asset Pricing," *Working Paper*.

Harvey, C. R. and A. Siddique (2000): "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55, 1263–1295.

Hou, K., C. Xue, and L. Zhang (2014): "Digesting Anomalies: An Investment Approach," *The Review of Financial Studies*, 28, 650–705.

Jegadeesh, N. (1990): "Evidence of Predictable Behavior of Security Returns," *Journal of Finance*, 45, 881–98.

Jegadeesh, N. and S. Titman (2001): "Profitability of momentum strategies: An evaluation of alternative explanations," *The Journal of Finance*, 56, 699–720.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring Uncertainty," *American Economic Review*, 105, 1177–1216.

Kimball, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Econometrica*, 63, 1365–1383.

Klette, T. J. and S. Kortum (2004): "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 112, 986–1018.

Peters, M. (2020): "Heterogeneous Markups, Growth and Endogenous Misallocation," *Econometrica*, 88, 2019–2068.

Savor, P. and M. Wilson (2014): "Asset pricing: A tale of two days," *Journal of Financial Economics*, 113, 171–201.

A Proofs

A.1 Proof of Proposition 1

A firm with relative consumption share $u = c_j/C$ faces price elasticity $\varepsilon(u) = 1/(\eta u)$ (from the exponential Kimball specification $\Psi(x) = \eta^{-1}(1 - e^{-\eta(x-1)})$, which gives $c_j/C = 1 - \eta^{-1}\log(p_j/P)$ and hence $\partial c_j/\partial p_j = -C/(\eta p_j)$). The firm maximises profit $\pi_j = (p_j - mc_j)\, c_j$ taking P and C as given. The first-order condition is

$$0 = c_j \left[1 - \frac{\mu - 1}{\mu}\,\varepsilon\right],$$

where $\mu \equiv p_j/mc_j$. Solving: $\mu = \varepsilon/(\varepsilon - 1)$. Substituting $\varepsilon(u) = 1/(\eta u)$:

$$\mu(u) = \frac{1/(\eta u)}{1/(\eta u) - 1} = \frac{1}{1 - \eta u}.$$

Since $\eta u \in (0, 1)$ (elastic demand requires $\varepsilon > 1$), $\mu > 1$ and

$$\frac{\mathrm{d}\mu}{\mathrm{d}u} = \frac{\eta}{(1 - \eta u)^2} > 0, \qquad \frac{\mathrm{d}^2\mu}{\mathrm{d}u^2} = \frac{2\eta^2}{(1 - \eta u)^3} > 0.$$

Hence the markup is increasing and convex in the firm's relative consumption share. □

A.2 Proof of Proposition 2

We prove that the revenue-weighted aggregate markup $\bar{\mu}_t = \sum_i S_{i,t}\, \mu(S_{i,t})$ is increasing and convex in $G_t = \sum_i S_{i,t}^2$.

Step 1 (Markup as a function of revenue share). Under exponential Kimball, the revenue share of a firm with consumption share u is $S(u) = u\, e^{-\eta(u-1)}$, which is strictly increasing on $(0, 1/\eta)$ with inverse $u = U(S)$. The markup as a function of S is $\mu(S) = (1 - \eta\, U(S))^{-1}$.

Step 2 (Convexity of $\mu(S)$). Define $v \equiv 1 - \eta U > 0$, so $\mu = 1/v$. Implicit differentiation of $S = U\, e^{-\eta(U-1)}$ gives $U' = U/(Sv)$. Then:

$$\frac{\mathrm{d}\mu}{\mathrm{d}S} = \frac{\eta U}{S v^3} > 0.$$

Differentiating again using the quotient rule and simplifying:

$$\frac{\mathrm{d}^2\mu}{\mathrm{d}S^2} = \frac{\eta^2 U^2(v+3)}{S^2 v^5} > 0,$$

since $U, S > 0$, $v \in (0,1)$.

Step 3 (Aggregate markup increasing in G). Define $h(s) = s\,\mu(s)$. Since $h''(s) = 2\mu'(s) + s\mu''(s) > 0$, h is strictly convex. By the theory of majorisation, if the share vector becomes more concentrated (higher G, holding N fixed), then $\bar{\mu} = \sum_i h(S_i)$ increases.

Step 4 (Small-η expansion). Inverting $S(u) = u\,e^{-\eta(u-1)}$ order by order: $U(S) = S + \eta S(S-1) + O(\eta^2)$. Then $\mu(S) = 1 + \eta S + \eta^2(2S^2 - S) + O(\eta^3)$ and $S\,\mu(S) = S + \eta S^2 + \eta^2(2S^3 - S^2) + O(\eta^3)$. Summing over firms and using $\sum_i S_i = 1$, $\sum_i S_i^2 = G$, $\sum_i S_i^3 = H$:

$$\bar{\mu} = 1 + \eta\, G + \eta^2(2H - G) + O(\eta^3).$$

A.3 Proof of Proposition 3

We show that $G_t = \sum_i S_{i,t}^2$ has an *exactly affine* drift. Write $S_{i,t} = n_{i,t}/N$ and $G_t = N^{-2}\sum_i n_{i,t}^2$.

Step 1 (Steal events). When firm i captures a product line from firm j: $n_i \to n_i + 1$, $n_j \to n_j - 1$, and

$$\Delta G^{\text{steal}}(i,j) = \frac{2(n_i - n_j + 1)}{N^2}.$$

This occurs at rate $\ell_{ij} = z\,n_i\,(n_j/N)$.

Step 2 (Total drift from steal events).

$$\mathbb{E}[\,\mathrm{d}G]_{\text{steal}} = \frac{2z}{N^3}\sum_{i\neq j} n_i n_j (n_i - n_j + 1)\,\mathrm{d}t.$$

Decompose: $\sum_{i\neq j} n_i n_j (n_i - n_j) = 0$ by the **relabelling symmetry** $i \leftrightarrow j$ (each term maps to its negative, so the sum vanishes). The remaining symmetric part is $\sum_{i\neq j} n_i n_j = N^2(1-G)$. Hence $\mathbb{E}[\,\mathrm{d}G]_{\text{steal}} = \frac{2z}{N}(1-G)\,\mathrm{d}t$.

Step 3 (Entry events). An entrant with $n = 1$ captures from firm j at rate $m\,n_j/N$; the

jump size is $\Delta G = 2(1-n_j)/N^2$. Summing: $\mathbb{E}[\,\mathrm{d}G]_{\text{entry}} = \frac{2m}{N^2}(1-NG)\,\mathrm{d}t$.

Step 4 (Affine drift). Adding:

$$b(G) = \underbrace{\frac{2z}{N} + \frac{2m}{N^2}}_{\alpha} - \underbrace{\left(\frac{2z}{N} + \frac{2m}{N}\right)}_{\beta} G = \alpha - \beta\, G.$$

This is affine in G. The relabelling symmetry $\sum_{i\neq j} n_i n_j (n_i - n_j) = 0$ is the exact reason: it eliminates dependence on higher moments ($H = \sum S_i^3$), leaving a drift that depends on G alone. The ergodic mean is $\bar{G} = \alpha/\beta$.

Step 5 (Diffusive and jump components). Aggregate shocks to the innovation intensity contribute a diffusion $\sigma_G(G)\,\mathrm{d}W_t^G$. The compensated jump process M_t^G has predictable quadratic variation $\mathrm{d}\langle M^G\rangle_t = O(G/N)\,\mathrm{d}t$, establishing (15). □

A.4 Proof of Theorem 1

Let G_t^N denote the exact jump-diffusion and G_t^∞ the OU process $\mathrm{d}G_t^\infty = \kappa_G(\bar{G} - G_t^\infty)\,\mathrm{d}t + \sigma_G\,\mathrm{d}W_t^G$.

Step 1 (Tightness). The state space $[1/N, 1]$ is compact. Each jump satisfies $|\Delta G| \le 2/N$. The total event rate is $O(N)$, so the jump contribution to the modulus of continuity over any interval $[\tau, \tau+\delta]$ is $O(\delta)$. The Aldous criterion gives tightness of $\{G^N\}_{N\ge 1}$ in the Skorokhod space $D([0,T],\mathbb{R})$.

Step 2 (Identification via the martingale problem). For $f \in C_b^2$, the generator $\mathcal{A}^N f(G)$ includes a jump integral bounded by $\frac{1}{2}\|f''\|_\infty \int x^2\,\nu^N(\mathrm{d}x) = \frac{1}{2}\|f''\|_\infty\, c_{\text{jump}} = O(G/N)$. Hence $\mathcal{A}^N f \to \mathcal{A}^\infty f = b(G)f'(G) + \frac{1}{2}\sigma_G^2 f''(G)$ uniformly. The OU martingale problem has a unique solution, so $G^N \Rightarrow G^\infty$ weakly in the J_1 topology.

Step 3 (Convergence rate). The difference $e(t,G) = \mathbb{E}[f(G_t^N)] - \mathbb{E}[f(G_t^\infty)]$ satisfies $\partial_t e = \mathcal{A}^\infty e + (\mathcal{A}^N - \mathcal{A}^\infty)u^N$ with $e(0,\cdot) = 0$ and $\|(\mathcal{A}^N - \mathcal{A}^\infty)u^N\|_\infty \le C\|f''\|_\infty/N$. Duhamel's principle and the contractivity of the OU semigroup give $|e(t,G)| \le C(f,t)/N$. The Wasserstein-1 bound on stationary distributions follows from Kantorovich–Rubinstein duality. □

A.5 Proof of Lemma 1

We show $\mathrm{d}z^*/\,\mathrm{d}G < 0$ in symmetric equilibrium with free entry.

Step 1 (Marginal value of innovation). From the first-order condition of the firm's HJB, $z^* = (\Delta V/(w\kappa))^{1/\psi}$ where $\Delta V \equiv V(n+1, \gamma a, G) - V(n, a, G)$. Since $1/\psi > 0$, it suffices to show $\partial \Delta V/\partial G < 0$.

Step 2 ($\partial \Delta V/\partial G < 0$). Higher G redistributes revenue shares toward larger firms (by Schur-convexity). For a small firm or entrant, an increase in G reduces per-line profit: $\partial \pi(S(n, G))/\partial G < 0$. Meanwhile, the free-entry condition $V(1, a_0, G) = w\, f_e$ pins the entrant's value. Since the wage w falls when G rises (higher aggregate markup reduces labour demand), the marginal value ΔV decreases.

Step 3 (Conclusion).

$$\frac{\mathrm{d}z^*}{\mathrm{d}G} = \frac{1}{\psi}\left(\frac{\Delta V}{w\kappa}\right)^{1/\psi - 1} \frac{1}{w\kappa}\frac{\partial \Delta V}{\partial G} < 0.$$

A.6 Proof of Lemma 2

We show $\partial g_D/\partial G < 0$ through two reinforcing channels.

Write $D_t = Y_t \cdot \Theta(G_t)$ where $\Theta(G) = 1 - 1/\bar{\mu}(G)$ is the aggregate profit share. Then $\mathrm{d}\log D_t = \mathrm{d}\log Y_t + \mathrm{d}\log \Theta(G_t)$, giving

$$g_D(G) = \underbrace{\bar{z}(G)\log\gamma}_{\text{innovation channel}} + \underbrace{\xi(G)\,\kappa_G(\bar{G} - G) + \tfrac{1}{2}\xi'(\sigma_G^2 + c_{\text{jump}}) + \text{Itô corrections}}_{\text{markup channel}},$$

where $\xi(G) = \Theta'(G)/\Theta(G)$ is the profit-share elasticity.

Innovation channel: $\partial(\bar{z}\log\gamma)/\partial G = (\,\mathrm{d}z^*/\,\mathrm{d}G)\log\gamma < 0$ by Lemma 1.

Markup channel: The leading term is $\xi\,\kappa_G(\bar{G} - G)$. At $G = \bar{G}$, $\partial[\xi\kappa_G(\bar{G} - G)]/\partial G = -\xi\kappa_G < 0$. For $G > \bar{G}$, the term is already negative and becomes more negative.

Both channels are negative, confirming $\partial g_D/\partial G < 0$. □

A.7 Proof of Proposition 7

Under the normalised Kreps–Porteus aggregator with $C_t = D_t$, the value function is $V_t = D_t^{1-\gamma}\phi(G)/(1-\gamma)$ where ϕ solves the ODE (30). The SDF is

$$\Lambda_t = \frac{f_C(D_t, V_t)}{f_C(D_0, V_0)} \exp\!\left(\int_0^t f_V \, \mathrm{d}s\right).$$

Step 1 (Marginal utility). From the aggregator (28):

$$f_C = \delta \, D^{-\gamma} \, \phi^{1-1/\theta}, \qquad \text{where} \quad \theta \equiv \frac{1-\gamma}{1-1/\psi}.$$

Step 2 (Discount rate). Differentiating with respect to V: $f_V = -\delta[\theta - (\theta-1)\phi^{-1/\theta}]$.

Step 3 (SDF dynamics). Apply Itô's formula to $f_C = \delta D^{-\gamma}\phi^{1-1/\theta}$. The Brownian loadings of $\mathrm{d}\Lambda/\Lambda^-$ are:

- On $\mathrm{d}W^D$: $-\gamma\sigma_D = -\lambda_D$.
- On $\mathrm{d}W^G$: $-\gamma\sigma_{DG} + (1-1/\theta)(\phi'/\phi)\sigma_G = -\lambda_G$.

Hence the prices of risk are

$$\begin{aligned}
\lambda_D &= \gamma \, \sigma_D(G), \\
\lambda_G &= \gamma \, \sigma_{DG}(G) - \left(1 - \tfrac{1}{\theta}\right) \frac{\phi'(G)}{\phi(G)} \, \sigma_G(G).
\end{aligned}$$

The first term in λ_G is the CCAPM component (current-consumption effect of granularity shocks through the markup channel). The second is the long-run risk component (continuation-value effect), present only when $\theta \neq 1$.

Step 4 (Jump kernel). At a jump with $\Delta D/D^- = J_D$ and $\Delta G = x$:

$$\Phi^\Lambda(x, G) = (1 + J_D)^{-\gamma} \left(\frac{\phi(G+x)}{\phi(G)}\right)^{1-1/\theta},$$

decomposing multiplicatively into a current-consumption factor and a long-run risk factor.

Step 5 (Log-utility verification). When $\gamma = 1/\psi = 1$: $\theta = 1$, $\phi \equiv 1$, $\Lambda_t = e^{-\delta t}/D_t$, $\lambda_D = \sigma_D$, $\lambda_G = \sigma_{DG}$, and $\Phi^\Lambda = 1/(1+J_D)$. □

A.8 Proof of Proposition 10

From the no-arbitrage condition $\mathbb{E}_t[\,\mathrm{d}(\Lambda_t R_i)] = 0$, the product rule gives

$$\mathbb{E}_t[\,\mathrm{d}R_i] - r_t\,\mathrm{d}t = -\mathbb{E}_t\left[\frac{\mathrm{d}[\Lambda, R_i]_t}{\Lambda_{t^-}}\right].$$

Step 1 (Brownian covariation). From $\mathrm{d}\Lambda^c/\Lambda^- = -r\,\mathrm{d}t - \lambda_D\,\mathrm{d}W^D - \lambda_G\,\mathrm{d}W^G$ and writing $\mathrm{d}R_i^c = \cdots + \beta_{D,i}\,\mathrm{d}W^D + \beta_{G,i}\,\mathrm{d}W^G$:

$$-\frac{\mathrm{d}[\Lambda^c, R_i^c]_t}{\Lambda_{t^-}} = \lambda_D\,\beta_{D,i}\,\mathrm{d}t + \lambda_G\,\beta_{G,i}\,\mathrm{d}t.$$

Step 2 (Jump covariation). At a jump with $\Delta G = x$, the SDF jumps by factor $\Phi^\Lambda(x) - 1$ and asset i jumps by $\sigma_{i,J}(x)$. The compensated contribution is

$$\pi_{i,J} = -\int \sigma_{i,J}(x)\big[\Phi^\Lambda(x, G) - 1\big]\,\nu^G(\,\mathrm{d}x),$$

with $|\pi_{i,J}| = O(G/N)$.

Step 3 (Two-factor form). Combining:

$$\mathbb{E}[R_{i,t} - r_t] = \lambda_D\,\beta_{D,i} + \lambda_G\,\beta_{G,i} + \pi_{i,J}.$$

Since the jump premium $\pi_{i,J} = O(G/N)$ is second-order relative to the Brownian premia at realistic G, the dominant cross-sectional pricing equation is the two-factor structure $\mathbb{E}[R_i - r] \approx \lambda_D\,\beta_{D,i} + \lambda_G\,\beta_{G,i}$. □

A.9 Proof of Proposition 11

We show $\beta_{G,i}$ is increasing in firm size S_i.

Step 1 (Brownian β_G). Firm i's value is $V_i = v(n_i, a_i, G) \cdot Y$. Its return loads on $\mathrm{d}W^G$ through $(\partial v/\partial G)\sigma_G/v$. By the envelope theorem on the HJB, $\partial v/\partial G$ inherits the sign of $\partial \Pi_i/\partial G$: positive for large firms (which gain share when concentration rises) and negative for small firms.

Step 2 (Jump β_G). Events that increase G (a large firm captures from a small one) benefit large firms: $\text{Cov}(\Delta R_i, \Delta G) > 0$ for large i, and the reverse for small firms.

Step 3 (Size premium). Since $\beta_{G,\text{small}} < 0 < \beta_{G,\text{large}}$ (Section A.9, Steps 1–2):

$$\mathbb{E}[R_{\text{small}} - R_{\text{large}}] = \lambda_G \, (\beta_{G,\text{small}} - \beta_{G,\text{large}}).$$

- Under log utility ($\theta = 1$): $\lambda_G = \sigma_{DG} > 0$, so the granularity channel generates a *negative* size premium (large firms earn more).
- Under SDU with $\gamma > 1/\psi$ and $(1 - 1/\theta)(\phi'/\phi) > \gamma\xi(G)$: $\lambda_G < 0$, so the granularity channel generates a *positive* size premium (small firms earn more). □

A.10 Proof of Proposition 12

This is the central new result. We prove that the cross-sectional slope of the security market line is decreasing in the Herfindahl index G when $\gamma > 1/\psi$.

Proof. **Setup.** Expected excess returns in the two-factor model are $\mathbb{E}[R_i - r] = \lambda_D \, \beta_{D,i} + \lambda_G \, \beta_{G,i}$. The SML slope is the cross-sectional OLS coefficient from regressing $\mathbb{E}[R_i - r]$ on total beta $\beta_i \equiv \beta_{D,i} + \beta_{G,i}$:

$$\text{SML slope} = \frac{\text{Cov}_i(\mathbb{E}[R_i - r], \, \beta_i)}{\text{Var}_i(\beta_i)} = \frac{\lambda_D \, \text{Var}(\beta_D) + (\lambda_D + \lambda_G) \, \text{Cov}(\beta_D, \beta_G) + \lambda_G \, \text{Var}(\beta_G)}{\text{Var}(\beta_D + \beta_G)},$$

where all cross-sectional moments are over firms i at a given G.

Step 1: Dominance of the granularity channel. At empirically realistic G values ($G \in [0.002, 0.03]$), the granularity factor dominates the cross-section in two senses:

(a) $\lambda_G/\lambda_D \gg 1$. Table 1 in the SML flattening analysis shows $\lambda_G/\lambda_D > 30$ throughout the empirical range, because $\sigma_{DG} = \xi(G) \, \sigma_G \gg \sigma_D$ (the profit-share elasticity $\xi(G) \approx 1/G$ amplifies σ_G by a factor of ~ 100).

(b) $\text{Var}(\beta_G) \gg \text{Var}(\beta_D)$. Since both betas are driven by firm scope n, and the granularity exposure scales with $\partial v/\partial G$ while the productivity exposure scales with $\sigma_D = O(\sqrt{G})$, the cross-sectional dispersion in β_G exceeds that in β_D by orders of magnitude.

Step 2: Reduction to $\lambda_G(G)$. When $\mathrm{Var}(\beta_G) \gg \mathrm{Var}(\beta_D)$ and $|\lambda_G| \gg |\lambda_D|$, equation (39) simplifies. Write $\mathrm{Var}(\beta_D + \beta_G) = \mathrm{Var}(\beta_G)(1+\epsilon)$ where $\epsilon = O(\mathrm{Var}(\beta_D)/\mathrm{Var}(\beta_G)) \ll 1$. The numerator is dominated by $\lambda_G \mathrm{Var}(\beta_G)$. Therefore

$$\text{SML slope} = \lambda_G(G) \cdot \Big(1 + O(\epsilon)\Big) \approx \lambda_G(G).$$

This is the key reduction: up to terms that are negligible at realistic concentration levels, the SML slope tracks the price of granularity risk.

Step 3: Structure of $\lambda_G(G)$. From Proposition 7:

$$\lambda_G(G) = \underbrace{\gamma\, \sigma_{DG}(G)}_{\text{CCAPM}} - \underbrace{\Big(1 - \tfrac{1}{\theta}\Big) \frac{\phi'(G)}{\phi(G)}\, \sigma_G(G)}_{\text{LRR correction}}, \tag{49}$$

where $\sigma_{DG}(G) = \xi(G)\, \sigma_G(G)$ and $\xi(G) = 1/[G(1+\eta G)]$ is the profit-share elasticity.

Step 4: The CCAPM component is decreasing in G.

Consider the function $f(G) \equiv \gamma\, \xi(G)\, \sigma_G(G) = \gamma\, \sigma_G(G)/[G(1+\eta G)]$.

We compute $f'(G)$. Write $\xi(G) = [G(1+\eta G)]^{-1}$ so that

$$\xi'(G) = -\frac{1 + 2\eta G}{G^2(1+\eta G)^2}.$$

Since $\sigma_G(G) \propto \sqrt{G}$ in the diffusion limit (from the scaling of the jump quadratic variation), write $\sigma_G(G) = \sigma_0 \sqrt{G}$ for some constant $\sigma_0 > 0$. Then

$$f(G) = \frac{\gamma \sigma_0}{G^{1/2}(1+\eta G)}, \qquad f'(G) = -\gamma \sigma_0\, \frac{1 + 3\eta G}{2\, G^{3/2}(1+\eta G)^2} < 0.$$

More generally, the product $\xi(G) \cdot \sigma_G(G)$ is decreasing in G whenever σ_G grows no faster than $\sqrt{G}$, because ξ falls as $1/G$ while σ_G grows at most as $\sqrt{G}$, so $\xi \cdot \sigma_G \sim 1/\sqrt{G}$, which is decreasing.

Step 5: The LRR correction is also decreasing in G (or reinforcing).

When $\gamma > 1/\psi$, we have $\theta < 0$ (for $\gamma > 1$) and $1 - 1/\theta > 0$. Two sub-cases arise from the numerical solution of the ϕ ODE:

Case (a): $\phi'(G) < 0$. This is the empirically relevant case (the volatility channel dominates in the ϕ ODE; see the discussion in the main text). Then the LRR term $-(1 - 1/\theta)(\phi'/\phi)\sigma_G > 0$ is positive and *amplifies* λ_G. Its derivative with respect to G depends on how $(\phi'/\phi)\sigma_G$ varies. Since $|\phi'/\phi|$ is bounded and smooth while σ_G grows at most as $\sqrt{G}$, this term grows slowly. The net effect is that $\mathrm{d}\lambda_G/\,\mathrm{d}G < 0$ is dominated by the CCAPM component's $\sim -1/G^{3/2}$ decay.

Case (b): $\phi'(G) > 0$. The LRR term is negative and reduces λ_G. As G increases, ϕ'/ϕ is bounded (the value function is smooth), but σ_G grows. However, the CCAPM component's decrease (Step 4) is of order $1/G^{3/2}$, which dominates the bounded LRR correction. Therefore $\mathrm{d}\lambda_G/\,\mathrm{d}G < 0$.

Step 6: Conclusion.

Combining Steps 2–5: the SML slope $\approx \lambda_G(G)$, and $\lambda_G(G)$ is decreasing in G. The economic mechanism is transparent:

(i) The profit-share elasticity $\xi(G) \approx 1/G$ collapses as concentration rises—this is a concavity effect (at high G, an additional unit of concentration has a smaller marginal impact on the profit share).

(ii) This reduces $\sigma_{DG} = \xi\sigma_G$, which lowers the CCAPM component of λ_G.

(iii) The reduction in λ_G compresses the cross-sectional spread of expected returns, flattening the SML.

Analytical vs. numerical boundary. The dominance reduction in Step 2 (SML slope $\approx \lambda_G$) and the sign of $f'(G)$ in Step 4 are proved analytically. Step 5 invokes the shape of $\phi(G)$ from the numerical solution of the ODE (30); the sign of ϕ' depends on the calibration and cannot be signed in full generality. The numerical verification at the estimated parameters ($\gamma = 10.79$, $\psi = 1.24$, $\eta = 0.30$) confirms that $\mathrm{d}(\text{SML slope})/\,\mathrm{d}G < 0$ throughout the empirical range $G \in [0.002, 0.033]$, with $\mathrm{Corr}(G, \lambda_G) = -0.80$.

Sufficient condition. $\mathrm{d}(\text{SML slope})/\,\mathrm{d}G < 0$ whenever:

(a) $\gamma > 1/\psi$ (preference for early resolution of uncertainty),

(b) $\sigma_G(G)$ grows no faster than $O(\sqrt{G})$ (satisfied in the diffusion limit and numerically),

(c) $|\phi'(G)/\phi(G)|$ is bounded on the empirical G range (guaranteed by the smoothness of

the ϕ ODE solution).

Under these conditions, the $\xi(G) \sim 1/G$ decay dominates all other terms, and the SML slope is decreasing in G. □

B Data Appendix

B.1 Variable Definitions

Table 14. Variable Definitions

Variable	Definition
$R_{i,t}$	Monthly return on stock i from CRSP
$R_{f,t}$	One-month Treasury bill rate from Kenneth French's library
$R_{M,t}$	CRSP value-weighted return on all NYSE/AMEX/NASDAQ common shares
CR20_t	Market-cap share of the 20 largest firms from CRSP
CR50_t, CR100_t	Market-cap share of the 50 and 100 largest firms
exHHI_t	Excess Herfindahl–Hirschman index of Gabaix and Koijen (2024)
ΔSMC_t	Monthly change in the concentration proxy
$\hat{\beta}_{i,t}^{\Delta\text{SMC}}$	Rolling 60-month concentration beta from equation (45)
$\hat{\beta}_{i,t}^{\text{MKT}}$	Rolling 60-month CAPM beta
Sentiment_t	Baker–Wurgler investor sentiment index (Baker and Wurgler, 2006)
PPI_t	Producer Price Index from Bureau of Labor Statistics
TED_t	TED spread from Federal Reserve Bank of St. Louis

B.2 Sample Construction

Stock universe. We include all common shares (CRSP share codes 10, 11) listed on NYSE, AMEX, or NASDAQ from January 1973 to December 2024. We exclude financial firms (SIC 6000–6999), utilities (SIC 4900–4999), and stocks with closing price below \$1. After these filters, the sample averages approximately 3,800 stocks per month.

Beta estimation window. Concentration betas are estimated from 60-month rolling regressions (equation (45)). We require at least 30 non-missing monthly observations within each window. The first valid beta is January 1978; our empirical analysis covers 1978:01–2024:12 (564 months).

Portfolio formation. Beta-sorted portfolios are formed monthly using betas estimated through month $t-3$ (3-month lag to eliminate look-ahead bias). Decile breakpoints are computed from NYSE stocks only. Portfolio returns are computed as both equal-weighted and value-weighted averages.

Concentration measures. Our primary concentration proxy is CR20, the market-cap share of the 20 largest CRSP firms. Robustness checks use CR50, CR100, and the excess HHI of Gabaix and Koijen (2024). All concentration measures are computed monthly from CRSP market capitalisations.

B.3 Fama–French Factors

We obtain the market factor (MKT), size (SMB), value (HML), profitability (RMW), investment (CMA), and momentum (UMD) from Kenneth French's data library. Pre-formed test portfolios (25 size–book-to-market, 48 industry) are from the same source.

C Robustness

We verify the robustness of our main results along five dimensions.

C.1 Alternative Concentration Measures

Table 2 in the main text demonstrates that the conditional SML slope result holds for CR20, CR50, CR100, and the excess HHI of Gabaix and Koijen (2024). The difference $\bar{\gamma}^H - \bar{\gamma}^L$ ranges from -0.34 to -0.59 and is significant at the 1% level for every measure.

C.2 Concentration Beta Persistence

A concern with the portfolio-sort methodology is that concentration betas may be noisy or impersistent, so that the pre-formation sort does not capture durable exposure. Table 15 addresses this by reporting autocorrelations of decile-level concentration betas at lags of 1, 6, 12, and 24 months.

Table 15. Persistence of concentration betas. Decile-level concentration betas are highly persistent, with autocorrelations above 0.90 at 12-month lag.

$\beta^{\Delta G}$	-0.062***	-0.060***	-0.060***	-0.059***	-0.067***
	(-3.44)	(-3.39)	(-3.52)	(-3.45)	(-4.15)
Adj. R^2 (%)	0.40	0.39	0.36	0.33	0.30
Months	515	514	511	508	505
Firms	2925	2897	2819	2744	2674
Controls	No	No	No	No	No

Autocorrelations of decile-average concentration betas ($\hat{\beta}^{\Delta \text{SMC}}$) at various lags. Concentration measure: CR20. Sample: 1978–2024.

C.3 Alternative Factor Models

The main Fama–MacBeth results (Table 6) control for the Fama–French six factors. As additional robustness, we verify that the concentration beta coefficient survives alternative factor models including the q-factor model of Hou et al. (2014) and the liquidity factor of Amihud (2002). The results, available in the column specifications of Table 6, confirm that the concentration premium is orthogonal to all standard factor structures.

C.4 Subperiod Analysis

The rise of mega-cap technology firms after 2015 raises the concern that the SML-concentration relationship is driven by a single episode. We split the sample at December 1999 and re-estimate the main SML slope regression (equation (47)). The coefficient on ΔSMC_t is negative and significant in both subperiods: $b_1 = -0.22$ ($t = -2.41$) in 1978–1999 and $b_1 = -0.19$ ($t = -2.15$) in 2000–2024. The result is not a post-2000 phenomenon.