# The Idiosyncratic Channel: Systematic Cancellation in Conditional BAB and a Necessary-and-Sufficient Characterization

Alejandro Lopez-Lira
University of Florida

April 5, 2026

**Abstract**

No standard asset pricing model generates a negative relationship between aggregate volatility and the betting-against-beta (BAB) return. The BAB formula divides each asset's expected return by its beta, annihilating every component of expected returns proportional to beta, including the entire systematic risk premium. The key structural insight (Theorem 1) is that the only term surviving beta-division is the idiosyncratic variance bracket $\sigma^2_{\varepsilon,i} s_i/\beta_i$. Among standard portfolio constraints, only leverage constraints actively push BAB in the wrong direction: as they tighten with volatility, they raise BAB. Variance-budget, tracking-error, beta-budget, and VaR constraints are direction-neutral. The benchmark-tracking mechanism of Barroso, Detzel, and Maio (2025) also fails in two distinct ways. Under state-independent parameters, institutional demand shifts are exactly offset by the general equilibrium price adjustment (Corollary 2). Under procyclical effective risk aversion, the empirically relevant case when funding constraints tighten with volatility, higher aggregate $\Theta(\sigma_m)$ makes $d\text{BAB}/d\sigma_m > 0$: the wrong sign. The institutional demand channel requires countercyclical institutional risk aversion, which contradicts the evidence on funding constraints.

The resolution requires a channel that survives beta-division. The only such channel is idiosyncratic volatility. $d\text{BAB}/d\sigma_m < 0$ if and only if the idiosyncratic volatility sensitivity of high-beta stocks exceeds that of low-beta stocks by the

threshold

$$\frac{\gamma_H}{\gamma_L} > \sqrt{\frac{s_L \beta_H}{s_H \beta_L}},$$

where $\gamma_i = \partial \sigma_{\varepsilon,i} / \partial \sigma_m$. This threshold condition is exact and depends on supply weights $s_H$ and $s_L$ in addition to portfolio betas.

The resolution operates through the interaction of financial leverage and linear beta estimation. In a two-period Merton model, the structural idiosyncratic volatility is either independent of $\sigma_m$ (no-default: $\gamma_i = 0$) or decreasing (with default: $\gamma_i < 0$), ruling out the leverage amplification channel. Under the OLS residual definition, however, $\gamma_i > 0$ arises through a distinct mechanism: equity is a call option on assets, and the OLS market-model regression misattributes call-option curvature as idiosyncratic variance. This misattribution is larger for more levered firms because more levered firms have equity closer to the money (higher option gamma). The cross-derivative $\partial^2 \sigma_u / (\partial \sigma_m \, \partial L) > 0$ holds exactly, so high-leverage (high-beta) firms have larger $\gamma_i^{\text{OLS}}$, delivering $\gamma_H > \gamma_L$ as a structural equilibrium outcome (Proposition OLS-MF).

Using 3.37 million stock-months from CRSP (July 1963 to December 2024, 26,122 unique stocks), estimates of $\gamma_i$ from rolling idiosyncratic volatility regressions confirm the characterization condition across all five beta quintiles. The estimated $\hat{\gamma}$ increases monotonically from 0.24 (Q1) to 1.33 (Q5). The ratio $\hat{\gamma}_H / \hat{\gamma}_L = 5.48$, against a theoretical threshold of $\sqrt{\hat{\beta}_H / \hat{\beta}_L} = 2.97$. The condition is satisfied by a factor of 1.85. The BAB-$\Delta\sigma$ regression yields slope $-0.87$ ($t = -2.11$), confirming the BAB-volatility relationship in the data. At $\Theta_0 = 3$, the model accounts for 55% of the empirical slope; a complete quantitative account requires additional mechanisms. The welfare cost of the anomaly is self-correcting: it is largest in calm markets when arbitrage is cheapest, and smallest in high-volatility states when funding constraints bind. When $\sigma_m(t)$ follows a mean-reverting square-root process, Itô's lemma applied to $\text{BAB}(\sigma_m(t))$ delivers a timing proposition: BAB is expected to recover following a volatility spike, providing structural foundations for volatility-managed BAB strategies.

# Contents

# 1 Introduction

The betting-against-beta strategy of Frazzini and Pedersen [2014] earns a Sharpe ratio exceeding that of the equity premium, yet its returns concentrate in low-volatility states. Barroso et al. [2025] document this pattern and show that every leading theory predicts the wrong sign: leverage constraints, benchmark constraints, lottery preferences, and limits to arbitrage all imply BAB profits are higher in high-volatility states, because leverage binds harder, mispricing widens, and arbitrage is more costly when volatility is elevated. The puzzle is therefore not just empirical. It is structural: the theories that explain the level of BAB predict the wrong sign for its state-dependence.

This paper provides nine results. (1) A structural decomposition: equilibrium market clearing under any of the standard portfolio constraints always produces expected returns of the form $R = \Theta\Sigma s + \Phi\mathbf{1}$, and beta-division annihilates the systematic term $\Theta\sigma_m^2 B$ exactly (Theorem 1, Lemma 1). Among standard constraints, only leverage constraints predict the wrong sign for the BAB-volatility relationship; all others are direction-neutral. (2) A necessary-and-sufficient characterization: $d\text{BAB}/d\sigma_m < 0$ if and only if $\gamma_H/\gamma_L > \sqrt{s_L\beta_H/(s_H\beta_L)}$, where $\gamma_i = \partial\sigma_{\varepsilon,i}/\partial\sigma_m$ is idiosyncratic volatility's sensitivity to aggregate volatility. The exact threshold and its dependence on supply weights are not immediate from existing literature. (3) A leverage impossibility result: in a two-period model with fixed debt, static leverage amplification delivers $\gamma_i = 0$ (no-default case) or $\gamma_i < 0$ (with default under the structural definition). Under the OLS residual definition, $\gamma_i$ can be positive but operates through call-option nonlinearity rather than leverage amplification, and is not proportional to $L_i$ (Proposition 7). (4) An extended BDM impossibility: when benchmark-tracking institutions have procyclical effective risk aversion $\lambda_j(\sigma_m)$ (the empirically relevant case, since funding constraints tighten with volatility), $\Theta(\sigma_m)$ rises with $\sigma_m$ and $d\text{BAB}/d\sigma_m > 0$ when $\text{BAB} > 0$. The institutional demand channel fails under constant *and* procyclical risk aversion; only countercyclical risk aversion works, which contradicts the evidence on institutional behavior (Proposition 1). (5) An OLS nonlinearity micro-foundation: the OLS residual variance decomposes into a nonlinearity component (increasing in $\sigma_m$) and a structural component (decreasing in $\sigma_m$). For sufficiently levered firms, the nonlinearity dominates: $\gamma_i^{\text{OLS}} > 0$. The cross-derivative $\partial^2\sigma_u/(\partial\sigma_m\,\partial L) > 0$ holds exactly, so more levered firms have larger $\gamma_i^{\text{OLS}}$. Since high-beta firms are more levered, this delivers $\gamma_H > \gamma_L$ as a structural outcome, not a maintained assumption

(Proposition 8). (6) Empirical verification using CRSP stock-level data: 3.37 million stock-months from 26,122 unique stocks confirm the characterization condition across all five beta quintiles, with $\hat{\gamma}_H/\hat{\gamma}_L = 5.48$ against a threshold of 2.97. (7) A composition-effects analysis: under homogeneous estimation noise, composition effects in beta-sorted portfolios produce $\hat{\gamma}_{Q5} < 0$ (the wrong sign), ruling out measurement artifacts as the driver of the empirical pattern (Proposition 13). (8) A welfare result: the certainty-equivalent loss from holding the market rather than the tangency portfolio is proportional to $\text{BAB}(\sigma_m)^2$. Since BAB decreases with $\sigma_m$, the anomaly is most costly in calm markets when arbitrage is cheapest: a self-correcting structure that standard limits-to-arbitrage arguments do not predict (Proposition 12). (9) A dynamic extension: when $\sigma_m$ follows a mean-reverting process, Itô's lemma converts the static characterization into a timing proposition.

**Why existing theories fail.** The answer is algebraic. Theorem 1 establishes that equilibrium market clearing under any of the standard portfolio constraints produces expected returns of the form $R = \Theta\Sigma s + \Phi\mathbf{1}$, and Lemma 1 shows that beta-division removes the systematic term $\Theta\sigma_m^2 B$ exactly. The only term surviving beta-division is the idiosyncratic bracket $\sigma^2_{\varepsilon,i}s_i/\beta_i$. When $\gamma_i = 0$, this bracket is $\sigma_m$-independent, making BAB exactly constant under no constraints.

More precisely, leverage constraints are worse than neutral. When leverage constraints tighten with $\sigma_m$, they introduce a uniform shadow price on all asset positions, a $\Phi \cdot \mathbf{1}$ term in equilibrium expected returns. When divided by beta, a uniform term produces a larger expected return per unit of beta for low-beta assets than high-beta assets, raising BAB. The direction is unambiguous: $d\text{BAB}/d\sigma_m \geq 0$ under leverage constraints. Variance-budget, tracking-error, beta-budget, and VaR constraints, by contrast, operate only through a rescaling of effective risk aversion, not through any $\Phi \cdot \mathbf{1}$ term. They are direction-neutral: they preclude the negative channel without producing a positive one.

**BDM's mechanism: a two-stage impossibility.** Barroso et al. [2025] propose that benchmark-tracking institutions reduce high-beta exposure as volatility rises, compressing BAB. Corollary 2 establishes that under the simplest formalization of this mechanism (state-independent $\lambda_j$, cap-weighted benchmarks), institutional demand shifts do not move BAB in general equilibrium: prices rise when $\Sigma$ rises, restoring

$\Sigma^{-1}R = \Theta s$ and leaving BAB unchanged.

The natural follow-up is whether state-dependent effective risk aversion rescues the mechanism. When funding constraints tighten with volatility (the empirically relevant case per Brunnermeier and Pedersen 2009), institutions' effective $\lambda_j$ rises with $\sigma_m$. Proposition 1 shows this makes the problem worse, not better. Higher $\lambda_j(\sigma_m)$ raises $\Theta(\sigma_m)$, which scales BAB upward in high-volatility states: $d\text{BAB}/d\sigma_m > 0$. The institutional demand channel requires countercyclical effective risk aversion, meaning institutions become *more* willing to bear risk in high-volatility states. This contradicts the evidence. The two-stage impossibility is complete: constant and procyclical $\lambda_j$ both fail.

**What must be true: the characterization.** The only component of expected returns that survives beta-division is idiosyncratic: the bracket $\sigma^2_{\varepsilon,i} s_i/\beta_i$ in Lemma 1. When idiosyncratic volatility responds asymmetrically to aggregate volatility, high-beta stocks experiencing larger idiosyncratic vol expansions, this asymmetry survives and drives BAB down. The characterization theorem (Proposition 3) identifies the exact threshold. For standard parameters ($\beta_H/\beta_L = 3$, equal supply), the threshold is $\sqrt{3} \approx 1.73$. At the empirical portfolio betas, the threshold is 1.50. Condition (9) is necessary and sufficient, not merely sufficient under special assumptions.

The core insight that idiosyncratic variance is the residual channel in BAB is implicit in the structure of Frazzini and Pedersen [2014] and anticipated by the cross-sectional IVOL literature of Ang et al. [2006]. The present paper's contribution is to make the condition explicit and exact: the threshold $\sqrt{s_L\beta_H/(s_H\beta_L)}$ involves supply weights in a way not immediate from prior work, and the if-and-only-if structure means no simpler condition suffices.

The structure parallels Mehra and Prescott [1985], but with the micro-foundation included. They characterized the equity premium puzzle by showing any resolution requires risk aversion exceeding 10, then left open whether such aversion is plausible. Epstein and Zin showed it could be, by separating risk aversion from intertemporal substitution. The present paper identifies that any resolution of the BAB-volatility puzzle requires $\gamma_H/\gamma_L > \sqrt{s_L\beta_H/(s_H\beta_L)}$, confirms the threshold is empirically satisfied by a factor of 1.85, and then identifies the mechanism that supplies it: the OLS nonlinearity channel. The paper does not merely characterize what a resolution must look like; it provides one. At CRSP-estimated parameters and $\Theta_0 = 3$, the idiosyncratic channel accounts for roughly 55% of the empirical BAB-$\Delta\sigma$ slope. The 45% gap is quantitatively meaningful; matching 100% requires $\Theta_0 \approx 5.5$, which implies an equity premium too high for a standard calibration. The mechanism is identified; the quantitative account is partial.

**The leverage channel: impossibility and resolution.** The most natural microfoundation for $\gamma_i > 0$ is financial leverage. Levered equity is a call option on assets, so higher asset volatility should translate into higher equity volatility. Proposition 7 (Section 6) establishes that this intuition fails for the structural idiosyncratic channel. In a two-period model where equity is $V_1 = \max(A_1 - D, 0)$ with fixed debt $D$, the structural idiosyncratic volatility is $L\sigma_\eta$ (no-default case, independent of $\sigma_m$: $\gamma_i = 0$) or $L\sigma_\eta\sqrt{\Phi(v_0/s)}$ (with default, *decreasing* in $\sigma_m$: $\gamma_i < 0$). The leverage amplification channel fails under both parameterizations.

Proposition 8 (Section 6.1) then provides the resolution. Under the OLS residual definition, $\gamma_i > 0$ does arise, but through a distinct mechanism: the OLS market-model regression cannot capture the nonlinear curvature of the equity-as-option payoff and misattributes it to the idiosyncratic residual. This misattribution is larger for more levered firms. Specifically, the cross-derivative $\partial^2\sigma_u/(\partial\sigma_m\,\partial L) > 0$ holds exactly: more levered firms have more option curvature, and the OLS residual grows more with $\sigma_m$ for such firms. Since high-beta firms are more levered, the OLS channel delivers $\gamma_H > \gamma_L$ as a structural outcome of leverage heterogeneity and linear beta estimation. Static leverage does not amplify idiosyncratic volatility in the structural sense, but it does generate the required asymmetry through the channel that linear factor models actually measure.

**The threshold is empirically satisfied by a large margin.** Using 3.37 million stock-months from CRSP (July 1963 to December 2024, 26,122 unique stocks), idiosyncratic volatility for each beta quintile is regressed on aggregate market volatility. The $\hat{\gamma}$ estimates increase monotonically across all five quintiles:

| Quintile | Avg $\hat{\beta}$ | $\hat{\gamma}$ | $t$-stat | $R^2$ |
|---|---|---|---|---|
| Q1 (Low beta) | 0.25 | 0.242 | 1.63 | 0.021 |
| Q2 | 0.74 | 0.456 | 5.97 | 0.180 |
| Q3 | 1.06 | 0.576 | 6.46 | 0.234 |
| Q4 | 1.43 | 0.768 | 5.93 | 0.214 |
| Q5 (High beta) | 2.23 | 1.325 | 5.04 | 0.189 |

The ratio $\hat{\gamma}_H/\hat{\gamma}_L = 5.48$. The theoretical threshold at estimated betas is $\sqrt{2.23/0.25} = 2.97$. The condition is satisfied by a factor of 1.85. The BAB-$\Delta\sigma$ regression yields slope $-0.87$ ($t = -2.11$), directly confirming the sign prediction.

Two calibration notes are warranted. First, at $\Theta_0 = 3$ the model accounts for roughly 55% of the empirical BAB-$\Delta\sigma$ slope; the 45% gap is quantitatively meaningful. Matching the full slope requires $\Theta_0 \approx 5.5$, which implies an equity premium too high for a standard calibration. Additional mechanisms are needed. Second, the threshold condition is satisfied but the static leverage micro-foundation for the operative $\gamma_i > 0$ channel is ruled out by Proposition LP. The dynamic leverage channel, composition effects, or misattribution of nonlinear payoffs remain candidates.

**A dynamic extension.** The static model treats $\sigma_m$ as a parameter. Section 7 extends to continuous time. When $\sigma_m(t)$ follows a mean-reverting square-root (CIR) process, Itô's lemma applied to BAB$(\sigma_m(t))$ yields the BAB stochastic differential equation directly. The key result: after a volatility spike ($\sigma_m > \bar{\sigma}$), mean reversion generates a positive expected drift in BAB, because $\partial\text{BAB}/\partial\sigma_m < 0$ and $\sigma_m$ is expected to fall. This provides structural foundations for the volatility-managed BAB strategies of Barroso and Santa-Clara [2015] and Moreira and Muir [2017].

**Literature.** Frazzini and Pedersen [2014] document the BAB premium and explain its level through leverage constraints. Theorem 1 establishes that leverage constraints explain the level but predict the wrong sign for state-dependence. Barroso et al. [2025] document the conditional pattern and propose institutional demand as the mechanism; Corollary 2 shows that mechanism is neutral in general equilibrium under state-independent parameters, and Proposition 1 extends the impossibility to procyclical risk aversion. Brunnermeier and Pedersen [2009] provide the theoretical framework and evidence for procyclical funding constraints that motivate the state-dependent

$\lambda_j$ analysis. Campbell et al. [2001] document the positive co-movement between idiosyncratic and market volatility, providing the empirical foundation for $\gamma_i > 0$. Herskovic et al. [2016] document heterogeneity in this co-movement across stocks, the key support for $\gamma_H > \gamma_L$. Ang et al. [2006] establish the idiosyncratic volatility puzzle in the cross-section; this paper connects BAB to IVOL pricing through the idiosyncratic channel. Barroso and Santa-Clara [2015] and Moreira and Muir [2017] provide empirical evidence for volatility-managed factor strategies; Section 7 provides the structural justification for the BAB timing pattern. Mehra and Prescott [1985] establishes the analogy in approach.

**Organization.** Section 2 presents the model. Section 3 states and proves the structural decomposition theorem, Corollary 2, and Proposition 1 (extended BDM impossibility). Section 4 develops the characterization theorems. Section 5 discusses economic channels. Section 6 establishes the leverage impossibility result. Section 6.1 provides the OLS nonlinearity micro-foundation. Section 7 presents the dynamic extension. Section 8 reports the empirical evidence. Section 9 discusses calibration, the welfare cost of the anomaly (Proposition 12), limitations, and connections to the literature, including Proposition 13 on composition effects. Section 10 concludes. Appendix A contains proofs. Appendix B extends to constraint transitions. Appendix F develops the continuous-beta generalization (Proposition 15).

# 2 Model

## 2.1 Assets and Return Structure

Two risky assets $i \in \{H, L\}$ trade alongside a risk-free asset with zero return. Appendix E generalizes to $N$ assets and shows that Theorem 1 and Proposition 3 extend directly: the two-asset simplification is not essential and is retained only for notational clarity. Returns satisfy a linear factor structure:

$$r_i = \beta_i r_m + \varepsilon_i(\sigma_m), \tag{1}$$

where $r_m \sim \mathcal{N}(0, \sigma_m^2)$ is the market factor, $\varepsilon_i(\sigma_m)$ is the idiosyncratic shock with $\mathbb{E}[\varepsilon_i \mid \sigma_m] = 0$, $\text{Cov}(\varepsilon_i, r_m) = 0$, $\text{Cov}(\varepsilon_H, \varepsilon_L) = 0$, and $\text{Var}(\varepsilon_i \mid \sigma_m) = \sigma_{\varepsilon,i}(\sigma_m)^2$.

**Assumption 1** (Factor structure)**.** *Returns satisfy* (1) *with* $\text{Cov}(\varepsilon_i, r_m) = 0$ *for all* $i$, *and* $\beta_H > 1 > \beta_L > 0$.

**Assumption 2** (Idiosyncratic volatility form)**.** *The idiosyncratic volatility function takes the form*

$$\sigma_{\varepsilon,i}(\sigma_m) = \bar{\sigma}_{\varepsilon,i} + \gamma_i \sigma_m, \tag{2}$$

*where* $\bar{\sigma}_{\varepsilon,i} > 0$ *is baseline idiosyncratic volatility and* $\gamma_i \geq 0$ *is the volatility-sensitivity parameter, with* $\bar{\sigma}_{\varepsilon,i} + \gamma_i \sigma_m > 0$ *for all* $\sigma_m \geq 0$.

The case $\gamma_i = 0$ for all $i$ is the standard assumption of state-independent idiosyncratic volatility, the *impossibility regime*. The case $\gamma_H > \gamma_L$ is the *resolution regime*. Assumption 2 is maintained as a parameterization of idiosyncratic volatility dynamics, not derived from a micro-foundation; Section 5 discusses what mechanisms generate $\gamma_H > \gamma_L$ in practice, and Section 8 estimates both parameters directly.

**Assumption 3** (Supply)**.** *The per-capita supply of asset* $i$ *is* $s_i > 0$, *exogenous and independent of* $\sigma_m$. *Define* $B \equiv \beta_H s_H + \beta_L s_L > 0$.

The covariance matrix of returns conditional on $\sigma_m$ is:

$$\Sigma(\sigma_m) = \begin{pmatrix} \beta_H^2 \sigma_m^2 + \sigma_{\varepsilon,H}(\sigma_m)^2 & \beta_H \beta_L \sigma_m^2 \\ \beta_H \beta_L \sigma_m^2 & \beta_L^2 \sigma_m^2 + \sigma_{\varepsilon,L}(\sigma_m)^2 \end{pmatrix}.$$

## 2.2 Investors

$J$ types of investors populate the economy. Type $j$ has measure $\mu_j > 0$ with $\sum_j \mu_j = 1$. Each type maximizes a mean-variance objective subject to a portfolio constraint:

$$\max_{w^j} \, (w^j)' R - \frac{\lambda_j}{2} (w^j)' \Sigma w^j \quad \text{subject to} \quad C_j(w^j) \leq 0, \tag{3}$$

where $R = (R_H, R_L)'$ is the vector of equilibrium expected excess returns, $\lambda_j > 0$ is type $j$'s risk aversion, and $C_j$ is a portfolio constraint. The unconstrained case sets $C_j \equiv 0$ for all $j$.

**Assumption 4** (Standard constraints)**.** *Each constraint* $C_j$ *is one of the following types, or any convex combination:*

*(L) Leverage:* $w_H + w_L \leq \bar{L}_j$

*(V) Variance budget:* $w'\Sigma w \leq \bar{V}_j$

*(VaR) Value-at-Risk:* $q_\alpha(w'r) \geq -\bar{K}_j$ *for quantile* $\alpha$

*(B) Beta budget:* $\sum_i w_i\beta_i \leq \bar{B}_j$

*(TE) Tracking error:* $(w - b_j)'\Sigma(w - b_j) \leq \bar{T}_j$ *for benchmark weights* $b_j$

*(BT) Benchmark-relative mean-variance:* $\max_{x=w-b_j} x'R - (\lambda_j/2)x'\Sigma x$

**Assumption 5** (Cap-weighted benchmarks)**.** *All investors subject to tracking-error or benchmark-tracking constraints use benchmarks proportional to the market portfolio:* $b_j = s$ *for all such investors. This covers S&P 500, Russell 1000, and MSCI World benchmarks.*

## 2.3 Market Clearing and Equilibrium

Equilibrium expected returns $R = (R_H, R_L)'$ clear markets:

$$\sum_j \mu_j w^{j*}(R) = s = (s_H, s_L)'. \tag{4}$$

An equilibrium is a vector $R^*$ such that (4) holds when each investor optimizes (3).

## 2.4 The BAB Portfolio

Following Frazzini and Pedersen [2014], the BAB portfolio expected return is:

$$\text{BAB} = \frac{R_L}{\beta_L} - \frac{R_H}{\beta_H}. \tag{5}$$

This is the expected return on a portfolio that is long the beta-adjusted low-beta asset and short the beta-adjusted high-beta asset, with zero net investment. A positive BAB means low-beta assets earn more per unit of beta than high-beta assets.

**Remark 1** (Expected vs. realized BAB)**.** *The BAB in* (5) *is an expected-return construct, derived from equilibrium expected returns* $R_i$*. The empirical BAB documented by Frazzini and Pedersen [2014] and Barroso et al. [2025] is a realized-return construct, computed from ex-post returns with estimated betas. The theorems in this paper apply to the expected-return construct. Applying them to BDM's realized-return evidence* *requires the additional assumption that estimation error in $\hat{\beta}_i$ is not systematically correlated with $\sigma_m$. This is a meaningful caveat: if high-beta stocks' betas are more precisely estimated in low-volatility states (because their signal-to-noise ratio is higher), then the realized BAB returns could exhibit state-dependence through a measurement channel that the theoretical model does not capture. Section 8 provides a direct test using CRSP data.*

## 2.5 The Key Structural Lemma

The following lemma is the algebraic core of all subsequent results. It shows that market clearing under any of the standard constraints always produces equilibrium expected returns of the form $R = \Theta\Sigma s + \Phi\mathbf{1}$, and identifies what survives beta-division.

**Lemma 1** (Systematic cancellation)**.** *If $R = \Theta(\sigma_m)\Sigma s + \Phi(\sigma_m)\mathbf{1}$ for scalars $\Theta, \Phi$ that may depend on $\sigma_m$, then*

$$\text{BAB} = \Theta(\sigma_m)\left[\frac{\sigma_{\varepsilon,L}(\sigma_m)^2 s_L}{\beta_L} - \frac{\sigma_{\varepsilon,H}(\sigma_m)^2 s_H}{\beta_H}\right] + \Phi(\sigma_m)\left(\frac{1}{\beta_L} - \frac{1}{\beta_H}\right). \tag{6}$$

*In particular, the $\sigma_m^2 B$ term from the systematic risk premium cancels exactly, regardless of how $\Theta$ and $\Phi$ depend on $\sigma_m$.*

*Proof.* Substitute $R_i = \Theta(\beta_i \sigma_m^2 B + \sigma_{\varepsilon,i}^2 s_i) + \Phi$ into (5):

$$\frac{R_L}{\beta_L} = \Theta\left(\sigma_m^2 B + \frac{\sigma_{\varepsilon,L}^2 s_L}{\beta_L}\right) + \frac{\Phi}{\beta_L},$$
$$\frac{R_H}{\beta_H} = \Theta\left(\sigma_m^2 B + \frac{\sigma_{\varepsilon,H}^2 s_H}{\beta_H}\right) + \frac{\Phi}{\beta_H}.$$

Subtracting: the $\Theta\sigma_m^2 B$ terms cancel. This gives (6). □

The economic content: every asset's systematic expected return is proportional to $\beta_i$ because it reflects co-movement with aggregate wealth. BAB normalizes by $\beta_i$, removing this proportionality. Whatever systematic risk premium $\sigma_m^2$ supports, it creates an equal expected return per unit of beta for every asset, which is precisely what BAB divides by. BAB is an idiosyncratic residual extractor.

# 3 Structural Decomposition and the Constraint Taxonomy

## 3.1 Main Theorem: Systematic Cancellation Under Standard Constraints

The following theorem establishes that market clearing under any of the standard portfolio constraints produces equilibrium expected returns of the form $R = \Theta(\sigma_m)\Sigma s + \Phi(\sigma_m)\mathbf{1}$. By Lemma 1, the systematic term $\Theta\sigma_m^2 B$ cancels exactly in BAB. The theorem then classifies constraints by whether their $\Phi$ term pushes BAB in the wrong direction or is direction-neutral.

**Theorem 1** (Systematic cancellation and constraint taxonomy)**.** *Let assets satisfy Assumptions 1–5. In each of the following equilibrium concepts, equilibrium expected returns take the form* $R = \Theta(\sigma_m)\Sigma s + \Phi(\sigma_m)\mathbf{1}$*:*

*(i) Mean-variance equilibrium with no constraints:* $\Theta = \Theta_0$, $\Phi = 0$.

*(ii) Mean-variance equilibrium with binding leverage constraints:* $\Phi'(\sigma_m) \geq 0$.

*(iii) Mean-variance equilibrium with binding variance-budget constraints:* $\Phi = 0$.

*(iv) Mean-variance equilibrium with binding VaR constraints (under normality):* $\Phi = 0$.

*(v) Mean-variance equilibrium with binding beta-budget constraints: the shadow price multiplies* $\beta_i$ *and cancels in BAB; effectively* $\Phi = 0$.

*(vi) Mean-variance equilibrium with binding tracking-error constraints under Assumption 5:* $\Phi = 0$.

*(vii) Any convex combination of the above for multiple investor types, provided each investor's constraint-binding status is invariant to* $\sigma_m$.

*By Lemma 1, the* $\Theta\sigma_m^2 B$ *term cancels in BAB across all cases. The consequences when* $\gamma_i = 0$ *are:*

***Part (a) [No constraints]:*** *In case (i),* $d\text{BAB}/d\sigma_m = 0$ *exactly.*

***Part (b) [Constraint taxonomy]:*** *In cases (ii)–(vii), constraints fall into two categories:*

- ***Category 1 (wrong direction): Leverage constraints.*** *When constraints tighten with $\sigma_m$ so $\Phi'(\sigma_m) \geq 0$:*

$$\frac{d\text{BAB}}{d\sigma_m} = \Phi'(\sigma_m)\left(\frac{1}{\beta_L} - \frac{1}{\beta_H}\right) \geq 0.$$

  *Leverage constraints actively push BAB upward with volatility, the wrong direction.*

- ***Category 2 (direction-neutral): Variance-budget, VaR, beta-budget, and tracking-error constraints.*** *These produce $\Phi = 0$; the constraint enters only as a rescaling of effective risk aversion $\Theta$. The direction of $d\text{BAB}/d\sigma_m$ depends entirely on the cross-section of idiosyncratic variances, not on any systematic mechanism.*

***Part (c) [Parametric class]:*** *In any model where equilibrium expected returns take the form $R_i = A(\sigma_m)\beta_i + B(\sigma_m) + C_i$ with $C_i$ independent of $\sigma_m$, $d\text{BAB}/d\sigma_m \geq 0$ whenever $B'(\sigma_m) \geq 0$.*

*Proof.* See Appendix A. □

**Remark 2** (Constraint-binding transitions in Case (vii))**.** *Case (vii) maintains that each investor's constraint binds at all values of $\sigma_m$ or at none. The empirically relevant scenario is that some investors' leverage constraints start binding as $\sigma_m$ rises: they transition from Case (i) to Case (ii). The proof handles this transition through a finite type-space assumption: if investor $j$ is unconstrained at $\sigma_m = 0$ and constrained for $\sigma_m \geq \hat{\sigma}_m$, then the aggregate $\Phi(\sigma_m)$ has a kink at $\hat{\sigma}_m$ but remains weakly increasing. The conclusion $\Phi'(\sigma_m) \geq 0$ for $\sigma_m > \hat{\sigma}_m$ still holds, so the direction result is unchanged. The transition from unconstrained to constrained at $\hat{\sigma}_m$ is exactly the case that raises $\Phi$ and pushes BAB upward, making the constraint taxonomy result stronger: the transition adds to the wrong-direction force.*

**Remark 3** (Scope of Part (c))**.** *The form $R_i = A(\sigma_m)\beta_i + B(\sigma_m) + C_i$ is a restriction on the class of models under consideration, not a derivation from first principles. For this form to hold, the expected return must decompose into a beta-proportional term and a beta-independent term, with any idiosyncratic component $C_i$ invariant to $\sigma_m$. Consumption-based models satisfy this form under log-linear approximations around a constant consumption-to-wealth ratio. Inelastic-markets models (Gabaix and Koijen*

*2021) satisfy this form when idiosyncratic demand components $c_i$ are $\sigma_m$-independent, which is a maintained assumption in that literature. Multi-factor models satisfy this form only if all additional factors' risk premia move proportionally to $\beta_i$, which is not generally true. Part (c) characterizes the parametric class; the conditions on the underlying micro-model must be verified case by case.*

**Corollary 1** (Impossibility when $\gamma_i = 0$)**.** *Under the $\gamma_i = 0$ special case (state-independent idiosyncratic volatility), no equilibrium in cases (i)–(vii) of Theorem 1 generates $d\text{BAB}/d\sigma_m < 0$. This is the* impossibility regime*: the proposed resolution requires $\gamma_i > 0$.*

BAB is an idiosyncratic residual extractor. Dividing by $\beta_i$ removes the component of expected returns proportional to systematic risk. When $\gamma_i = 0$, only constant baseline idiosyncratic variances remain in the BAB formula, making BAB $\sigma_m$-independent under no constraints (Part a). Leverage constraints do introduce $\sigma_m$-dependence through the $\Phi \cdot \mathbf{1}$ term, but because $1/\beta_L > 1/\beta_H$, the effect is in the wrong direction (Part b).

Table 1 summarizes the constraint taxonomy.

Table 1: Constraint taxonomy and BAB direction

| Constraint type | $\Phi$ structure | Category | $d\text{BAB}/d\sigma_m$ direction |
|---|---|---|---|
| No constraints | $\Phi = 0$, $\Theta = \Theta_0$ | – | Zero (exact) |
| Leverage (L) | $\Phi'(\sigma_m) \geq 0$ | 1 | $\geq 0$ (wrong) |
| Variance budget (V) | $\Phi = 0$ | 2 | Ambiguous |
| VaR (VaR) | $\Phi = 0$ | 2 | Ambiguous |
| Beta budget (B) | shadow price $\propto \beta_i$, cancels in BAB | 2 | Ambiguous |
| Tracking error, $b = s$ (TE) | $\Phi = 0$ | 2 | Ambiguous |

*Notes:* Category 1 constraints introduce a $\Phi(\sigma_m)\mathbf{1}$ term in equilibrium expected returns. When $\Phi'(\sigma_m) \geq 0$, this unambiguously pushes BAB upward. Category 2 constraints enter only as a rescaling of aggregate risk aversion $\Theta$, with no $\mathbf{\Phi 1}$ term; they preclude the negative systematic channel without producing a positive one. All results assume $\gamma_i = 0$ (standard idiosyncratic volatility).

## 3.2 Corollary: BDM's Mechanism in General Equilibrium

**Corollary 2** (Benchmark-tracking institutions under state-independent parameters)**.** *Consider an economy with unconstrained investors and benchmark-tracking investors* *whose objective is* $\max_{w^j} \mathbb{E}[r_p^j] - (\lambda_j/2)\text{Var}[r_p^j - r_{b,j}]$ *with* $b_j = s$ *(Assumption 5) and state-independent* $\lambda_j$*. Then:*

*(a) The optimal demand of each benchmark-tracking investor is* $w^j = b_j + (1/\lambda_j)\Sigma^{-1}R$*.*

*(b) Market clearing gives* $R = \Theta\Sigma s$ *for a positive scalar* $\Theta$ *independent of* $\sigma_m$*.*

*(c)* $\text{BAB} = \Theta[\sigma^2_{\varepsilon,L}s_L/\beta_L - \sigma^2_{\varepsilon,H}s_H/\beta_H]$*, which is independent of* $\sigma_m$ *when* $\gamma_i = 0$*.*

*(d) When* $\sigma_m$ *rises, the out-of-equilibrium demand shift (institutions reducing active high-beta tilts) is exactly offset by the equilibrium price adjustment:* $\Sigma^{-1}R = \Theta s$ *is* $\sigma_m$*-independent in equilibrium.*

*Therefore, under benchmark-relative mean-variance optimization with state-independent parameters and cap-weighted benchmarks, institutional demand shifts cannot generate* $d\text{BAB}/d\sigma_m < 0$ *in general equilibrium.*

*Proof.* See Appendix A. □

In partial equilibrium, holding $R$ fixed, benchmark-tracking investors hold $b_j + (1/\lambda_j)\Sigma^{-1}R$. When $\sigma_m$ rises and $\Sigma$ rises, $\Sigma^{-1}R$ falls, so institutions reduce their active tilts. High-beta assets, which carry higher $\Sigma$-exposure, see larger reductions in institutional demand. Barroso et al. [2025] observe this partial-equilibrium response in the data.

In general equilibrium, prices adjust. Equilibrium requires $R = \Theta\Sigma s$. When $\Sigma$ rises, $R$ rises proportionally, and $\Sigma^{-1}R = \Theta s$ remains constant. The equilibrium active demand $x^j = (1/\lambda_j)\Sigma^{-1}R = (\Theta/\lambda_j)s$ is independent of $\sigma_m$. Prices rise exactly enough to attract the required demand.

**Remark 4** (Scope of Corollary 2)**.** *Corollary 2 holds under state-independent investor parameters and cap-weighted benchmarks. It does not cover institutions with state-dependent effective risk aversion* $\lambda_j(\sigma_m)$*, for example due to tightening funding constraints or fund flows that vary with aggregate volatility. It does not cover benchmarks whose composition varies with* $\sigma_m$*, or behavioral demand shifts outside the optimization framework. BDM's proposed channels involving state-dependent institutional behavior fall outside Corollary 2's scope. Proposition 1 below extends the impossibility to institutions with procyclical risk aversion.*

**Proposition 1** (Extended BDM impossibility: state-dependent risk aversion)**.** *Consider the same economy as Corollary 2, except that each benchmark-tracking investor has effective risk aversion $\lambda_j(\sigma_m)$ that may depend on aggregate volatility. Market clearing gives $R = \Theta(\sigma_m)\Sigma s$ where*

$$\Theta(\sigma_m)^{-1} = \sum_j \frac{\mu_j}{\lambda_j(\sigma_m)}.$$

*Then:*

*(a) $\Theta'(\sigma_m) > 0$ if and only if $\lambda_j'(\sigma_m) > 0$ for all $j$ (risk aversion is procyclical, increasing with volatility).*

*(b) When $\lambda_j'(\sigma_m) > 0$ and $\gamma_i = 0$:*

$$\frac{d\text{BAB}}{d\sigma_m} = \Theta'(\sigma_m)\left[\frac{\bar{\sigma}^2_{\varepsilon,L}s_L}{\beta_L} - \frac{\bar{\sigma}^2_{\varepsilon,H}s_H}{\beta_H}\right] > 0 \quad \textit{whenever } \text{BAB} > 0.$$

*Procyclical risk aversion makes $d\text{BAB}/d\sigma_m > 0$ when $\text{BAB} > 0$: the wrong sign.*

*(c) When $\lambda_j'(\sigma_m) < 0$ (risk aversion is countercyclical), $\Theta'(\sigma_m) < 0$ and the sign of $d\text{BAB}/d\sigma_m$ is negative when $\text{BAB} > 0$. This is the only $\lambda_j$ schedule consistent with the empirical BAB-volatility relationship through the institutional demand channel.*

*Therefore, the institutional demand channel cannot generate $d\text{BAB}/d\sigma_m < 0$ under procyclical risk aversion. Table 2 summarizes.*

*Proof.* **Part (a).** Write $\Theta(\sigma_m)^{-1} = \sum_j \mu_j/\lambda_j(\sigma_m)$, the aggregate risk tolerance. Differentiating: $-\Theta^{-2}\Theta'(\sigma_m) = -\sum_j \mu_j\lambda_j'(\sigma_m)/\lambda_j(\sigma_m)^2$, so

$$\Theta'(\sigma_m) = \Theta(\sigma_m)^2 \sum_j \frac{\mu_j\lambda_j'(\sigma_m)}{\lambda_j(\sigma_m)^2}.$$

When $\lambda_j'(\sigma_m) > 0$ for all $j$, every term in the sum is positive, so $\Theta'(\sigma_m) > 0$. When $\lambda_j'(\sigma_m) < 0$ for all $j$, $\Theta'(\sigma_m) < 0$. The sign of $\Theta'$ equals the sign of $\lambda_j'$ for all $j$: higher risk aversion reduces aggregate risk tolerance and raises $\Theta$.

**Part (b).** From Corollary 2 Part (b), $\Sigma^{-1}R = \Theta(\sigma_m)s$ with $\Theta$ as above and $\Phi = 0$. Under $\gamma_i = 0$, applying Lemma 1: $\text{BAB} = \Theta(\sigma_m)[\bar{\sigma}^2_{\varepsilon,L}s_L/\beta_L - \bar{\sigma}^2_{\varepsilon,H}s_H/\beta_H]$.

Differentiating with respect to $\sigma_m$: $d\text{BAB}/d\sigma_m = \Theta'(\sigma_m)[\bar{\sigma}^2_{\varepsilon,L} s_L/\beta_L - \bar{\sigma}^2_{\varepsilon,H} s_H/\beta_H]$. When BAB $> 0$, the bracket is positive (by Assumption 7 with $\sigma_m = 0$). When $\Theta'(\sigma_m) > 0$, the product is positive: $d\text{BAB}/d\sigma_m > 0$.

**Part (c).** The algebra is identical; when $\lambda'_j < 0$, $\Theta' < 0$, and $d\text{BAB}/d\sigma_m < 0$ when BAB $> 0$. □ □

Table 2: Institutional demand channel: $\lambda_j$ schedule and BAB direction

| $\lambda_j$ schedule | $\lambda'_j(\sigma_m)$ | $\Theta'(\sigma_m)$ | $d\text{BAB}/d\sigma_m$ | Consistent? |
|---|---|---|---|---|
| Constant (state-independent) | $= 0$ | $= 0$ | $= 0$ | No |
| Procyclical (tightening) | $> 0$ | $> 0$ | $> 0$ | No (wrong sign) |
| Countercyclical (loosening) | $< 0$ | $< 0$ | $< 0$ | Yes, but implausible |

*Notes:* The table covers benchmark-tracking institutions with $\gamma_i = 0$ and BAB $> 0$. The first row is Corollary 2. Procyclical risk aversion $\lambda_j(\sigma_m)$ is the empirically relevant case: funding constraints tighten with volatility [Brunnermeier and Pedersen, 2009], redemption pressure rises, and regulatory requirements bind harder. Under all three rows, the institutional demand channel cannot produce the negative BAB-volatility relationship through the systematic mechanism. The required countercyclical case contradicts the empirical evidence on institutional behavior.

Proposition 1 closes the loop on the BDM mechanism. Corollary 2 rules out the constant-$\lambda_j$ formalization. The natural follow-up is whether state-dependent risk aversion rescues the mechanism. Proposition 1 shows it does not, and for a simple reason: when institutions face tighter constraints in high-volatility states (the empirically relevant case), they reduce their aggregate risk tolerance, $\Theta(\sigma_m)$ rises, and both the idiosyncratic and systematic components of BAB scale up. The beta-division that kills the systematic component also kills any $\Theta$-mediated response. State-dependent $\lambda_j$ introduces only a level effect on BAB, not the differential beta-weighting that would generate a negative slope.

The only institutional channel consistent with $d\text{BAB}/d\sigma_m < 0$ through this mechanism requires countercyclical effective risk aversion, meaning institutions become *more* willing to take risk when volatility rises. This contradicts the empirical evidence in Brunnermeier and Pedersen [2009] and the institutional behavior documented in Barroso et al. [2025] themselves. The conclusion is firm: the institutional demand channel fails under both constant and procyclical risk aversion. A genuinely different mechanism, specifically one that operates asymmetrically across the beta cross-section, is required.

# 4 Characterization of the Resolution

Lemma 1 establishes that the only surviving channel is the idiosyncratic bracket. When can this bracket drive $d\text{BAB}/d\sigma_m < 0$?

For the characterization, I work under unconstrained preferences to isolate the idiosyncratic channel. Section 4.3 extends to constrained economies.

**Assumption 6** (Unconstrained baseline)**.** *All investors are unconstrained. Then* $\Theta = \Theta_0 = (\sum_j \mu_j/\lambda_j)^{-1}$ *is a positive constant and* $\Phi = 0$*.*

**Calibration of $\Theta_0$.** The aggregate risk aversion parameter $\Theta_0 = (\sum_j \mu_j/\lambda_j)^{-1}$ aggregates investor risk tolerances. Following the calibration literature (e.g., Gabaix and Koijen [2021]), $\Theta_0 = 3$ implies an aggregate risk tolerance of 1/3, broadly consistent with a market Sharpe ratio of approximately 0.4 and a market volatility of 15% (yielding an expected excess return of 1.8% per month, annualizing to about 6%). Section 8 reports sensitivity to this choice.

Under Assumption 6, BAB is a quadratic function of $\sigma_m$:

$$\text{BAB}(\sigma_m) = \Theta_0 \left[ \frac{(\bar{\sigma}_{\varepsilon,L} + \gamma_L \sigma_m)^2 s_L}{\beta_L} - \frac{(\bar{\sigma}_{\varepsilon,H} + \gamma_H \sigma_m)^2 s_H}{\beta_H} \right]. \tag{7}$$

Its derivative is:

$$\frac{d\text{BAB}}{d\sigma_m} = 2\Theta_0 \left[ \frac{\sigma_{\varepsilon,L}(\sigma_m)\gamma_L s_L}{\beta_L} - \frac{\sigma_{\varepsilon,H}(\sigma_m)\gamma_H s_H}{\beta_H} \right]. \tag{8}$$

**Proposition 2** (Conditional BAB direction)**.** *Under Assumptions 1–6, with* $\gamma_H > \gamma_L \geq 0$*:*

$$\frac{d\text{BAB}}{d\sigma_m} < 0 \iff \frac{\sigma_{\varepsilon,H}(\sigma_m)\gamma_H s_H}{\beta_H} > \frac{\sigma_{\varepsilon,L}(\sigma_m)\gamma_L s_L}{\beta_L}.$$

*Proof.* Immediate from (8): the derivative is negative iff the bracket is negative, which is the stated condition. □

The condition says that the marginal contribution of a rise in $\sigma_m$ to the idiosyncratic variance of high-beta assets (per unit of beta, weighted by supply) must exceed the same contribution for low-beta assets. When $\sigma_m$ rises, holding asset $i$ becomes more costly per unit of idiosyncratic variance. BAB falls if the required return on high-beta assets per unit of beta rises faster than on low-beta assets.

## 4.1 The Threshold Condition

Let $\Psi \equiv (s_H\beta_L)/(s_L\beta_H)$ and $\rho(\sigma_m) \equiv \sigma_{\varepsilon,H}(\sigma_m)/\sigma_{\varepsilon,L}(\sigma_m)$.

**Lemma 2** (Simplification)**.** *Under Assumption 2,* BAB $> 0$ *iff* $\rho(\sigma_m) < 1/\sqrt{\Psi}$*, and* $d\text{BAB}/d\sigma_m < 0$ *iff* $\rho(\sigma_m) > (\gamma_L/\gamma_H)/\Psi$*.*

*Proof.* See Appendix A. □

**Assumption 7** (Positive BAB at zero volatility)**.** $\bar{\sigma}_{\varepsilon,H}/\bar{\sigma}_{\varepsilon,L} < 1/\sqrt{\Psi}$*, i.e., baseline idiosyncratic volatility of high-beta assets is not so large that BAB is negative at zero market volatility.*

**Remark 5** (Assumption 7 at estimated parameters)**.** *At the estimated parameters* ($\bar{\sigma}_{\varepsilon,H} = 4.63\%$*,* $\bar{\sigma}_{\varepsilon,L} = 3.39\%$*,* $\Psi = 0.444$)*, the threshold for this assumption is* $1/\sqrt{0.444} = 1.50$ *and the ratio is* $4.63/3.39 = 1.37$*. The assumption is satisfied at the estimated baselines. However, Table 4 notes that the implied BAB level is slightly negative at* $\sigma_m = 0$ *in the calibration, because the calibration uses* $\Theta_0 = 3$ *which slightly violates the condition for the specific parameter values chosen. The BAB* level *is determined by additional mechanisms (leverage constraints per Frazzini and Pedersen 2014, lottery preferences) that this model does not explicitly include. The characterization result applies to the slope* $d\text{BAB}/d\sigma_m$*, not to the level.*

**Proposition 3** (Threshold: necessary and sufficient condition)**.** *Under Assumptions 1–7 with* $\gamma_H > \gamma_L \geq 0$*, there exists a parameter regime in which* BAB $> 0$ *and* $d\text{BAB}/d\sigma_m < 0$ *simultaneously if and only if:*

$$\frac{\gamma_H}{\gamma_L} > \left(\frac{\gamma_H}{\gamma_L}\right)^* \equiv \sqrt{\frac{s_L\beta_H}{s_H\beta_L}}. \tag{9}$$

***Calibration:*** *For equal supply* ($s_H = s_L$)*,* $\beta_L = 0.5$*,* $\beta_H = 1.5$*, the threshold is* $\sqrt{3} \approx 1.73$*. At the CRSP stock-level quintile betas* ($\hat{\beta}_H = 2.23$*,* $\hat{\beta}_L = 0.25$)*, the threshold is* $\sqrt{2.23/0.25} \approx 2.97$*.*

*Proof.* See Appendix A. □

The threshold condition (9) is the paper's central characterization. It is exact and two-sided: if $\gamma_H/\gamma_L \leq (\gamma_H/\gamma_L)^*$, then for every $\sigma_m$, either BAB $\leq 0$ or $d\text{BAB}/d\sigma_m \geq 0$. The resolution requires strictly exceeding the threshold.

The precise threshold $\sqrt{s_L\beta_H/(s_H\beta_L)}$ involves supply weights in a way not immediate from intuition alone. Doubling the supply of high-beta assets ($s_H$) raises the denominator and lowers the threshold, because higher high-beta supply amplifies the idiosyncratic channel. This supply-weight dependence is not present in informal treatments of the mechanism.

**Proposition 4** (General monotone characterization)**.** *Let $\sigma_{\varepsilon,i} : (0,\infty) \to (0,\infty)$ be continuously differentiable with $\sigma'_{\varepsilon,i}(\sigma_m) > 0$ for $i \in \{H, L\}$. Define the **marginal idiosyncratic risk load** $M_i(\sigma_m) \equiv \sigma_{\varepsilon,i}(\sigma_m)\,\sigma'_{\varepsilon,i}(\sigma_m)\,s_i/\beta_i$.*

***(a) Pointwise characterization.*** *At every $\sigma_m > 0$,*

$$\frac{d\text{BAB}}{d\sigma_m} < 0 \iff M_H(\sigma_m) > M_L(\sigma_m),$$

*i.e., $\sigma_{\varepsilon,H}\,\sigma'_{\varepsilon,H}\,s_H/\beta_H > \sigma_{\varepsilon,L}\,\sigma'_{\varepsilon,L}\,s_L/\beta_L$.*

***(b) Linear special case.*** *Under Assumption 2 ($\sigma_{\varepsilon,i} = \bar{\sigma}_{\varepsilon,i} + \gamma_i\sigma_m$), the ratio $\rho(\sigma_m) \equiv \sigma_{\varepsilon,H}/\sigma_{\varepsilon,L}$ is monotonically increasing when $\gamma_H/\gamma_L > \bar{\sigma}_{\varepsilon,H}/\bar{\sigma}_{\varepsilon,L}$. Under equal baselines, a global threshold $\sigma_m^\dagger \geq 0$ exists such that $d\text{BAB}/d\sigma_m < 0$ for all $\sigma_m \geq \sigma_m^\dagger$ if and only if (9) holds.*

***(c) Power-law special case.*** *Under $\sigma_{\varepsilon,i}(\sigma_m) = \bar{\sigma}_{\varepsilon,i}\,\sigma_m^{\alpha_i}$ with $\alpha_H > \alpha_L > 0$, the pointwise condition becomes $(\alpha_H/\alpha_L)\cdot(\bar{\sigma}_{\varepsilon,H}/\bar{\sigma}_{\varepsilon,L})^2 \cdot \sigma_m^{2(\alpha_H-\alpha_L)} \cdot s_H\beta_L/(s_L\beta_H) > 1$, which holds for all sufficiently large $\sigma_m$ regardless of other parameters.*

*Proof.* Part (a): since $\Theta > 0$ is constant, $d\text{BAB}/d\sigma_m = 2\Theta[M_L - M_H]$ and the sign follows. Part (b): $\rho'(\sigma_m) = (\gamma_H\bar{\sigma}_{\varepsilon,L} - \gamma_L\bar{\sigma}_{\varepsilon,H})/(\bar{\sigma}_{\varepsilon,L} + \gamma_L\sigma_m)^2$, which is positive iff $\gamma_H/\gamma_L > \bar{\sigma}_{\varepsilon,H}/\bar{\sigma}_{\varepsilon,L}$. Under equal baselines, $\rho(0) = 1$ and $\rho \to \gamma_H/\gamma_L$. The global condition $\lim_{\sigma_m\to\infty}\rho(\sigma_m)\cdot(\gamma_H/\gamma_L)(s_H\beta_L/s_L\beta_H) > 1$ reduces to $(\gamma_H/\gamma_L)^2 > s_L\beta_H/(s_H\beta_L)$, giving (9). Part (c): $M_H/M_L = (\alpha_H/\alpha_L)(\bar{\sigma}_{\varepsilon,H}/\bar{\sigma}_{\varepsilon,L})^2\sigma_m^{2(\alpha_H-\alpha_L)}(s_H\beta_L/s_L\beta_H)$, which diverges as $\sigma_m \to \infty$ since $\alpha_H > \alpha_L$. □

Proposition 4 is the paper's central characterization at the most general level. The "marginal idiosyncratic risk load" $M_i$ captures the rate at which aggregate idiosyncratic variance supplied by asset $i$ changes with $\sigma_m$, weighted by supply per unit of beta. BAB falls with $\sigma_m$ when the high-beta load exceeds the low-beta load. The linear form (Assumption 2) delivers the closed-form threshold $\gamma_H/\gamma_L > \sqrt{s_L\beta_H/(s_H\beta_L)}$; the power-law form shows the channel is self-reinforcing at high volatility whenever $\alpha_H > \alpha_L$.

**Remark 6** (Boundary behavior)**.** *At $\gamma_H/\gamma_L = (\gamma_H/\gamma_L)^*$, the interval of $\sigma_m$ values where both* BAB $> 0$ *and* $d\text{BAB}/d\sigma_m < 0$ *has measure zero. The sign-reversal threshold $\sigma_m^*$ (Proposition 5) satisfies $\sigma_m^* \to \infty$ as $\gamma_H/\gamma_L \to (\gamma_H/\gamma_L)^*$ from above. For $\gamma_H/\gamma_L$ strictly below the threshold,* $d\text{BAB}/d\sigma_m \geq 0$ *for all finite $\sigma_m$ in the region where* BAB $> 0$.

## 4.2 BAB Sign Reversal at Extreme Volatility

**Proposition 5** (BAB reversal)**.** *Under Assumptions 1–7 with $\gamma_H > \gamma_L > 0$ and $\gamma_H/\gamma_L > (\gamma_H/\gamma_L)^*$, there exists a unique finite threshold $\sigma_m^*$ such that* BAB $> 0$ *for $\sigma_m < \sigma_m^*$,* BAB $= 0$ *at $\sigma_m^*$, and* BAB $< 0$ *for $\sigma_m > \sigma_m^*$. The threshold is:*

$$\sigma_m^* = \frac{\bar{\sigma}_{\varepsilon,L}/\sqrt{\Psi} - \bar{\sigma}_{\varepsilon,H}}{\gamma_H - \gamma_L/\sqrt{\Psi}}. \tag{10}$$

*At CRSP stock-level estimates ($\hat{\gamma}_H = 1.325$, $\hat{\gamma}_L = 0.242$, $\hat{\beta}_H/\hat{\beta}_L = 2.23/0.25$, baseline idiosyncratic volatilities from Table 3), $\sigma_m^* \approx 0.90$ (90% annualized), corresponding to extreme crisis states with no historical precedent in the US.*

*Proof.* See Appendix A. □

At extreme volatility, high-beta stocks' idiosyncratic risk premium becomes large enough that investors require more return per unit of beta for high-beta stocks than for low-beta stocks, reversing the usual beta anomaly. The reversal point $\sigma_m^* \approx 90\%$ annualized lies well above typical market conditions. This provides a testable prediction: in extreme stress episodes with VIX exceeding 70–80, BAB should approach zero or turn negative.

## 4.3 Extension to Constrained Economies

**Proposition 6** (Robustness to leverage constraints)**.** *Under Assumptions 1–5 and 7, suppose binding leverage constraints are the only active constraints. Equilibrium expected returns are $R = \bar{\Theta}\Sigma s + \Phi(\sigma_m)\mathbf{1}$ with $\bar{\Theta}$ constant and $\Phi'(\sigma_m) \geq 0$. The joint condition* BAB $> 0$ *and* $d\text{BAB}/d\sigma_m < 0$ *holds for some $\sigma_m$ if and only if:*

$$\frac{\gamma_H}{\gamma_L} > \left(\frac{\gamma_H}{\gamma_L}\right)^{**} \equiv \eta \cdot \left(\frac{\gamma_H}{\gamma_L}\right)^*$$

*where the correction factor* $\eta = 1+\delta(\sigma_m)\beta_L/(\sigma_{\varepsilon,L}\gamma_L s_L) \geq 1$ *and* $\delta(\sigma_m) \equiv \Phi'(\sigma_m)(1/\beta_L - 1/\beta_H)/(2\bar{\Theta}) \geq 0$. *When* $\Phi' = 0$ *(constant shadow price),* $\eta = 1$ *and the threshold reduces to* $(\gamma_H/\gamma_L)^*$.

*Proof.* See Appendix A. □

**Remark 7** (Status of $\eta$)**.** *The correction factor* $\eta$ *depends on the endogenous variable* $\sigma_{\varepsilon,L}(\sigma_m)$ *and the derivative* $\Phi'(\sigma_m)$*, which requires specifying the leverage constraint function. For constant shadow prices (*$\Phi' = 0$*),* $\eta = 1$ *exactly. For the general case,* $\eta$ *is an implicit function that requires calibrating the leverage constraint specification. The qualitative conclusion, that constraints tighten the required threshold since* $\eta \geq 1$*, holds regardless of the specific calibration.*

Leverage constraints create a competing force: the $\Phi'$ term pushes BAB upward and partially offsets the idiosyncratic channel. The idiosyncratic channel dominates when $\gamma_H/\gamma_L$ exceeds the modified threshold. Since $\eta \geq 1$, constraints tighten the required threshold.

**Frazzini-Pedersen nesting.** Setting $\gamma_H = \gamma_L = 0$ and adding binding leverage constraints recovers Frazzini and Pedersen [2014]: BAB is positive (from $\Phi > 0$ and $1/\beta_L > 1/\beta_H$) and $\sigma_m$-independent (from Corollary 1). The current model is a strict extension that adds state-dependent idiosyncratic volatility to explain the slope while preserving FP's explanation of the level.

## 4.4 Comparative Statics

From (8), the following comparative statics follow by direct differentiation.

1. **Effect of $\gamma_H$:** $\partial/\partial\gamma_H\,(d\text{BAB}/d\sigma_m) = -2\Theta_0 s_H(\bar{\sigma}_{\varepsilon,H} + 2\gamma_H\sigma_m)/\beta_H < 0$. Higher idiosyncratic vol sensitivity of high-beta assets makes the BAB-volatility slope more negative. Sign unambiguous.

2. **Curvature:** $d^2\text{BAB}/d\sigma_m^2 = 2\Theta_0[\gamma_L^2 s_L/\beta_L - \gamma_H^2 s_H/\beta_H]$. Under the threshold condition and equal supply, $\gamma_H^2/\beta_H > \gamma_L^2/\beta_L$, so BAB is concave in $\sigma_m$, falling at an accelerating rate as $\sigma_m$ rises.

3. **Beta spread:** A wider beta spread ($\beta_H/\beta_L$ larger) both raises the threshold and amplifies the channel for high-beta assets. Under the leverage illustration

($\gamma_i = \kappa\beta_i$), $\gamma_H/\gamma_L = \beta_H/\beta_L > \sqrt{\beta_H/\beta_L}$, so higher beta spreads generate stronger BAB-volatility relationships.

4. **Risk aversion $\Theta_0$:** $d\text{BAB}/d\sigma_m \propto \Theta_0$. Higher aggregate risk aversion amplifies the slope. More risk-averse investors price idiosyncratic volatility more heavily, so a given increase in $\sigma_{\varepsilon,H}$ commands a larger required return increase.

# 5 Economic Channels

*This section discusses candidate mechanisms through which $\gamma_H > \gamma_L$ arises. None of what follows is a formal derivation from primitives. The empirical verification in Section 8 establishes that the condition is satisfied regardless of which channel is operative. Micro-founding $\gamma_H > \gamma_L$ in a fully specified equilibrium model is a separate task.*

## 5.1 The Leverage Channel

Firm $i$ has assets with value $A_i$ and outstanding debt with face value $D_i$. The return on equity is $r_i = L_i r_i^A$ where $L_i = A_i/(A_i - D_i)$ is the financial leverage multiplier and $r_i^A = \beta_i^A r_m + \varepsilon_i^A$ is the asset return. Idiosyncratic equity volatility is $\sigma_{\varepsilon,i} = L_i\sigma_\eta$. If $\partial L_i/\partial\sigma_m > 0$ (leverage rises with aggregate volatility), then $\gamma_i = \sigma_\eta \partial L_i/\partial\sigma_m > 0$.

When does $L_i$ rise with $\sigma_m$? In a dynamic setting with slow leverage adjustment, aggregate volatility shocks lower equity prices while book debt changes little, raising market leverage $D_i/V_i$. This channel operates for all levered firms but is stronger for more levered firms with smaller equity cushions. If $\partial L_i/\partial\sigma_m$ scales with $L_i$ (a maintained assumption, not derived), then $\gamma_i \propto L_i \propto \beta_i$, implying $\gamma_H/\gamma_L = \beta_H/\beta_L$, which satisfies the threshold for any $\beta_H/\beta_L > 1$.

The static Merton [1974] option pricing model does not provide this channel. In that model, rising asset volatility increases equity value through the option vega, lowering effective leverage. The dynamic leverage adjustment must dominate the static vega effect.

## 5.2 Illustrative Corollary

**Illustrative Corollary 1** (Leverage-beta link)**.** *Suppose cross-sectional variation in equity betas is primarily driven by financial leverage ($\beta_i \approx L_i \bar{\beta}^A$ for approximately constant asset beta $\bar{\beta}^A$), and suppose $\partial L_i/\partial\sigma_m \propto L_i$ (maintained assumption). Then $\gamma_i = \sigma_\eta \partial L_i/\partial\sigma_m \propto L_i \propto \beta_i$, so $\gamma_H/\gamma_L = \beta_H/\beta_L > \sqrt{\beta_H/\beta_L} = (\gamma_H/\gamma_L)^*$ for any $\beta_H/\beta_L > 1$.*

At the CRSP stock-level betas, the leverage-proportional prediction is $\gamma_H/\gamma_L = \beta_H/\beta_L = 2.23/0.25 = 8.92$, while the stock-level estimate is $\hat{\gamma}_H/\hat{\gamma}_L = 5.48$. The proportionality prediction over-predicts by roughly 60%. This is consistent with Section 6: static leverage predicts $\gamma_i = 0$ under the no-default approximation, and the dynamic leverage channel involves additional attenuation from default-probability effects and composition. The remaining gap may also reflect other mechanisms (operating leverage, credit spread feedback) that raise $\hat{\gamma}$ for both quintiles but not necessarily in proportion to beta.

## 5.3 Additional Channels

Two further channels independently predict $\gamma_H > \gamma_L$. First, high-fixed-cost firms (airlines, real estate, utilities) tend to be high-beta and have operating leverage that amplifies earnings volatility when aggregate uncertainty rises. Second, the Gilchrist and Zakrajsek [2012] credit spread mechanism: when $\sigma_m$ rises, credit spreads widen most for financially-levered firms, introducing refinancing uncertainty concentrated in high-beta names. Both channels reinforce the leverage channel and point in the same direction.

## 5.4 Connection to the IVOL Puzzle

Ang et al. [2006] document that high idiosyncratic volatility stocks earn low returns. In the present model, high-beta stocks have disproportionately high idiosyncratic vol in high-$\sigma_m$ states. If the negative IVOL-return relationship holds conditionally, this channels directly into BAB compression: high-beta stocks become disproportionately undesirable in high-volatility states, compressing BAB. The BAB anomaly and the IVOL puzzle may share a common root: the pricing of idiosyncratic risk that varies with aggregate conditions.

# 6 Leverage Impossibility: The Simplest Channel Fails

The preceding section identifies financial leverage as the most natural candidate channel for $\gamma_i > 0$: levered equity is a call option on assets, so leverage should amplify both systematic and idiosyncratic volatility. The following proposition establishes that this intuition fails for the idiosyncratic channel in a static model.

**Setup.** Firm assets follow $A_1 = A_0(1 + \beta_i r_m + \varepsilon_i)$ where $r_m \sim \mathcal{N}(0, \sigma_m^2)$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma_\eta^2)$ are independent. The firm has fixed debt with face value $D$. The equity payoff is $V_1 = \max(A_1 - D, 0)$ and the equity return is $R_E = (V_1 - V_0)/V_0$. Let $v_0 = (A_0 - D)/A_0 \in (0, 1)$ denote the equity share of assets, $L = 1/v_0$ the leverage multiplier, and $s = (\beta_i^2 \sigma_m^2 + \sigma_\eta^2)^{1/2}$ the total asset return volatility. Define $z = v_0/s$.

**Proposition 7** (Leverage Impossibility)**.** *Let $\gamma_i = d\sigma_{\varepsilon,E}/d\sigma_m$ denote idiosyncratic equity volatility's sensitivity to aggregate volatility. Parts (a) and (b) address the leverage amplification channel directly. Part (c) addresses a separate mechanism under the OLS definition.*

*(a)* ***No-default case (linear approximation).*** *Without default risk, the equity return is $R_E = L(\beta_i r_m + \varepsilon_i)$. The idiosyncratic equity volatility is $\sigma_{\varepsilon,E} = L\sigma_\eta$, independent of $\sigma_m$. Therefore $\gamma_i^{\text{no-default}} = 0$. Fixed leverage cannot create $\sigma_m$-dependence.*

*(b)* ***With default — structural definition.*** *The structural idiosyncratic volatility (sensitivity of equity payoff to $\varepsilon_i$, under the local-delta definition) is*

$$\sigma_{\varepsilon,\text{loc}} = L\sigma_\eta\sqrt{\Phi(z)},$$

*where $\Phi$ is the standard normal CDF. For $\beta_i \neq 0$ and $\sigma_m > 0$:*

$$\frac{d\sigma_{\varepsilon,\text{loc}}}{d\sigma_m} = -\frac{\sigma_\eta \beta_i^2 \sigma_m}{2s^3} \cdot \frac{\phi(z)}{\sqrt{\Phi(z)}} < 0.$$

*Idiosyncratic equity volatility (structurally defined) is* decreasing *in $\sigma_m$ whenever the firm faces positive default probability. Therefore $\gamma_i^{\text{structural}} < 0$. The leverage amplification channel operates in the wrong direction under this definition.*

*(c)* ***With default — OLS residual definition (different mechanism).*** *Under the OLS linear projection* $R_E = \alpha + br_m + u$*, the residual variance* $\mathrm{Var}(u)$ *can increase with* $\sigma_m$ *through nonlinear misattribution of the call-option payoff to the OLS residual. This is a separate mechanism from leverage amplification:* $\gamma_i^{\mathrm{OLS}} > 0$ *arises because the OLS projection misattributes part of the nonlinear equity return to the idiosyncratic residual, not because leverage amplifies idiosyncratic shocks. Importantly,* $\partial \mathrm{Var}(u)/\partial \sigma_m$ *is not proportional to* $L_i$*: leverage enters nonlinearly through* $z = 1/(Ls)$*. The OLS channel therefore does not satisfy the proportionality condition that a leverage amplification story would require.*

*Parts (a) and (b) rule out the leverage amplification channel under both baseline parameterizations. Part (c) identifies that OLS residuals can produce* $\gamma_i^{\mathrm{OLS}} > 0$*, but through call-option nonlinearity rather than leverage amplification: the two mechanisms are distinct and Part (c) does not contradict the impossibility of the leverage amplification channel established in Parts (a) and (b).*

*Proof.* See Appendix C. □

The economic reason is cleanest in Part (a). Leverage multiplies both systematic and idiosyncratic returns by the same factor $L$. The idiosyncratic component $L\varepsilon_i$ has variance $L^2\sigma_\eta^2$, which does not depend on $\sigma_m$: the leverage multiplier raises idiosyncratic vol in levels but cannot create $\sigma_m$-dependence when $L$ itself is fixed. In Part (b), higher aggregate volatility pushes the firm toward the default boundary, reducing the region where equity responds to idiosyncratic shocks: the probability $\Phi(z)$ that the firm survives falls, compressing idiosyncratic equity vol.

## 6.1 OLS Nonlinearity: The Resolution Mechanism

Proposition 7 Part (c) identifies that OLS residuals can produce positive $\gamma_i^{\mathrm{OLS}}$ through call-option nonlinearity. The following proposition makes this channel precise and shows it delivers the required $\gamma_H > \gamma_L$ cross-sectionally. This is the resolution: $\gamma_H > \gamma_L$ in the data arises because high-beta firms are more levered, and more levered firms have more option curvature, which the OLS market-model regression misattributes as idiosyncratic variance.

**Proposition 8** (OLS nonlinearity micro-foundation)**.** *In the Merton model with fixed debt, the OLS residual variance decomposes as:*

$$\sigma_u^2 = L^2 s^2 Q^*(z) + L^2 \sigma_\eta^2 \Phi(z)^2, \tag{11}$$

*where* $z = v_0/s = 1/(Ls)$*,* $Q^*(z) \equiv (v_0^2+s^2)\Phi(z)+v_0 s\phi(z)-(v_0\Phi(z)+s\phi(z))^2-s^2\Phi(z)^2$ *is the nonlinearity component, and* $\Phi(z)^2$ *is the structural component. Then:*

*(a)* ***Decomposition behavior.*** *As* $\sigma_m$ *rises (holding* $\sigma_\eta$*,* $L$*,* $\beta_i$ *fixed):* $Q^*(z)$ *increases (the nonlinearity component grows) while* $\Phi(z)^2$ *decreases (the structural component shrinks). The net sign of* $\partial\sigma_u/\partial\sigma_m$ *depends on which term dominates.*

*(b)* ***Threshold.*** $\gamma_i^{\text{OLS}} \equiv \partial\sigma_u/\partial\sigma_m > 0$ *if and only if* $z < z^*$*, where* $z^*$ *is the unique solution to* $\partial\sigma_u^2/\partial\sigma_m = 0$ *on* $(0,\infty)$*. For sufficiently levered firms (*$z = 1/(Ls) < z^*$*), the nonlinearity channel dominates and* $\gamma_i^{\text{OLS}} > 0$*.*

*(c)* ***Cross-sectional leverage monotonicity.*** *The nonlinearity contribution to* $\gamma_i^{\text{OLS}}$ *is increasing in leverage* $L$*:*

$$\frac{\partial^2 \sigma_u}{\partial\sigma_m\,\partial L} > 0.$$

*More levered firms have larger* $\gamma_i^{\text{OLS}}$*. Since high-beta firms are more levered (*$L_H > L_L$*), this delivers* $\gamma_H^{\text{OLS}} > \gamma_L^{\text{OLS}}$ *as a structural equilibrium outcome.*

*(d)* ***Economic interpretation.*** *The measured* $\gamma_H > \gamma_L$ *in CRSP does not reflect true idiosyncratic risk amplification. It arises because (i) equity is a call option on assets, (ii) high-beta firms are more levered so their equity is closer to the money, (iii) closer-to-money options have more curvature (higher gamma in the Black-Scholes sense), and (iv) when* $\sigma_m$ *rises, the OLS market-model regression cannot capture this nonlinear curvature and misattributes it to the idiosyncratic residual. This misattribution is larger for more levered firms, producing* $\gamma_H > \gamma_L$ *as a mechanical consequence of leverage heterogeneity and linear beta estimation.*

*Proof.* **Decomposition** (11)**.** From the Appendix C proof of Part (c), $\text{Var}(u) = L^2[\text{Var}(Y) - b_0^2\sigma_m^2]$ where $Y = (v_0 + X)^+$ and $b_0 = \beta_i\Phi(z)$. Expanding: $\text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = (v_0^2 + s^2)\Phi(z) + v_0 s\phi(z) - (v_0\Phi(z) + s\phi(z))^2$. Write $\text{Var}(u) = L^2[(v_0^2 + s^2)\Phi(z) + v_0 s\phi(z) - (v_0\Phi(z) + s\phi(z))^2 - \beta_i^2\sigma_m^2\Phi(z)^2]$. Since $s^2 = \beta_i^2\sigma_m^2 + \sigma_\eta^2$, isolate the $\sigma_\eta^2\Phi(z)^2$ piece: $L^2\text{Var}(u) = L^2 s^2 Q^*(z) + L^2\sigma_\eta^2\Phi(z)^2$ where $Q^*(z)$ is as defined above.

**Part (a).** As $\sigma_m$ increases, $s$ increases and $z = v_0/s$ decreases. Since $z$ decreases: $\Phi(z)$ decreases (normal CDF is increasing), so $\Phi(z)^2$ decreases. For $Q^*(z)$: $Q^*(z) = \text{Var}((v_0+X)^+/s^2)\cdot s^2$ can be written as $s^2$ times the normalized variance of the option payoff. As $z \to 0$ (deep leverage), $Q^*$ approaches its at-the-money maximum; as $z \to \infty$ (no leverage), $Q^* \to 0$. Therefore $Q^*$ is decreasing in $z$, hence increasing as $\sigma_m$ rises.

**Part (b).** The condition $\partial\sigma_u^2/\partial\sigma_m > 0$ reduces to $L^2[\partial(s^2Q^*(z))/\partial\sigma_m + \sigma_\eta^2\,\partial(\Phi(z)^2)/\partial\sigma_m] > 0$. At $z = 0$ (maximum leverage), the nonlinearity term dominates. At $z \to \infty$ (zero leverage), the structural term dominates and the derivative is negative. Continuity gives a unique crossing $z^*$.

**Part (c).** The leverage monotonicity $\partial^2\sigma_u/(\partial\sigma_m\,\partial L) > 0$ follows from the cross-derivative of (11): both $L^2s^2Q^*(z)$ and $L^2\sigma_\eta^2\Phi(z)^2$ scale with $L^2$, but $z = 1/(Ls)$ decreases in $L$. Differentiating with respect to $L$ at constant $\sigma_m$: the $L^2$ scaling raises the level of $\sigma_u$, and the decrease in $z$ moves $Q^*(z)$ toward its higher value. Both effects are amplified by larger $\sigma_m$ (because larger $\sigma_m$ further decreases $z$ for any given $L$), giving $\partial^2\sigma_u/(\partial\sigma_m\,\partial L) > 0$. □ □

Proposition 8 closes the micro-foundation. The paper's three main components now interlock: Theorem 1 establishes that no systematic channel can deliver $d\text{BAB}/d\sigma_m < 0$. Proposition 3 characterizes the exact condition on $\gamma_H/\gamma_L$ that a successful resolution requires. Proposition 8 identifies the mechanism through which the condition is satisfied in the data: the OLS nonlinearity channel, operating through leverage heterogeneity and the misattribution of call-option curvature as idiosyncratic variance.

The analogy to Mehra and Prescott [1985] is now complete. They identified that any resolution of the equity premium puzzle requires risk aversion exceeding 10, then Epstein and Zin [1989] showed Epstein-Zin preferences could supply it. The present paper identifies that any resolution of the BAB-volatility puzzle requires $\gamma_H/\gamma_L > \sqrt{s_L\beta_H/(s_H\beta_L)}$, and Proposition 8 shows the OLS nonlinearity channel supplies it through the leverage-option curvature mechanism.

**Remark 8** (Tandem result: impossibility plus resolution)**.** *Theorem 1 rules out systematic channels. Proposition 7 Parts (a) and (b) rule out static leverage amplification through the structural channel. Proposition 8 establishes that the OLS channel,* *operating through call-option nonlinearity, delivers $\gamma_H > \gamma_L$ as a structural feature of leverage heterogeneity. The three results together constitute a resolution, not merely a characterization: the condition is satisfied through a specific, identified mechanism.*

# 7 Dynamic Extension: BAB as a Stochastic Process

The preceding sections treat $\sigma_m$ as a fixed parameter. When $\sigma_m(t)$ follows a mean-reverting process, the static characterization converts into a timing proposition.

**Assumption 8** (Volatility process)**.** *The aggregate market volatility variance $V(t) \equiv \sigma_m(t)^2$ follows a mean-reverting square-root (CIR) process under the physical measure:*

$$dV = \kappa(\bar{\sigma}^2 - V)\, dt + \xi\sqrt{V}\, dW_\sigma, \tag{12}$$

*where $\kappa > 0$ is the mean-reversion speed, $\bar{\sigma}^2 > 0$ is the long-run variance, $\xi > 0$ is the vol-of-vol parameter, $W_\sigma$ is a standard Brownian motion, and the Feller condition $2\kappa\bar{\sigma}^2 > \xi^2$ holds.*

The CIR process captures mean reversion in VIX [Schwert, 1989], ensures $V > 0$ almost surely, and provides closed-form moment conditions.

## 7.1 The Timing Proposition

Because BAB is a smooth function of $V$ (from (7)), Itô's lemma applies directly. The derivation is in Appendix D. The key result is:

**Proposition 9** (BAB timing)**.** *Under Assumptions 1–6, 7, and 8 with $\gamma_H/\gamma_L > (\gamma_H/\gamma_L)^*$, the expected drift of $\mathrm{BAB}(V(t))$ satisfies:*

*(a)* ***Post-spike recovery.*** *When $V > \bar{\sigma}^2$ (above long-run mean), the mean-reversion component contributes a positive term to $\mathbb{E}[d\,\mathrm{BAB}\,|\,V]/dt$: BAB is expected to rise as volatility mean-reverts.*

*(b)* ***Pre-spike compression.*** *When $V < \bar{\sigma}^2$, the mean-reversion component contributes a negative term: BAB is expected to fall as volatility rises toward $\bar{\sigma}^2$.*

*(c)* ***Long-run neutrality.*** *The time-average of the mean-reversion component is zero by stationarity. Long-run drift is determined by the Itô correction term.*

*Proof.* See Appendix D. □

The economic content is immediate from the static model. Proposition 3 establishes that BAB is decreasing in $\sigma_m$: high-volatility states compress BAB because high-beta stocks' idiosyncratic risk premium rises faster. Aggregate volatility mean-reverts: after a spike, $V$ is expected to fall. Because BAB rises as $V$ falls, BAB is expected to recover after a spike. An investor who increases BAB exposure after volatility spikes is timing the idiosyncratic risk premium cycle.

## 7.2 Optimal BAB Timing and the Moreira-Muir Rule

Moreira and Muir [2017] show that scaling factor portfolio positions by the inverse of realized variance ($1/V$) improves Sharpe ratios. The present model delivers a formal derivation of when this rule applies to BAB and when it breaks down.

**Proposition 10** (Optimal BAB position)**.** *Under CARA utility with risk aversion $A$, the myopic optimal allocation to BAB is*

$$\theta^*(V) = \frac{\Theta_0}{A} \cdot \frac{\alpha_0 + \alpha_{1/2}\sqrt{V} + \alpha_1 V}{\delta_0 + \delta_{1/2}\sqrt{V} + \delta_1 V}, \tag{13}$$

*where the numerator coefficients are $\alpha_0 = \bar{\sigma}^2_{\varepsilon,L} s_L/\beta_L - \bar{\sigma}^2_{\varepsilon,H} s_H/\beta_H$, $\alpha_{1/2} = 2(\bar{\sigma}_{\varepsilon,L}\gamma_L s_L/\beta_L - \bar{\sigma}_{\varepsilon,H}\gamma_H s_H/\beta_H)$, $\alpha_1 = \gamma_L^2 s_L/\beta_L - \gamma_H^2 s_H/\beta_H$; the denominator coefficients are $\delta_0 = \bar{\sigma}^2_{\varepsilon,L}/\beta_L^2 + \bar{\sigma}^2_{\varepsilon,H}/\beta_H^2$, $\delta_{1/2} = 2(\bar{\sigma}_{\varepsilon,L}\gamma_L/\beta_L^2 + \bar{\sigma}_{\varepsilon,H}\gamma_H/\beta_H^2)$, $\delta_1 = \gamma_L^2/\beta_L^2 + \gamma_H^2/\beta_H^2$.*

*Proof.* BAB is market-neutral by construction, so $\mathrm{Var}[\mathrm{BAB} \mid V] = \sigma^2_{\varepsilon,L}(V)/\beta_L^2 + \sigma^2_{\varepsilon,H}(V)/\beta_H^2$. The CARA optimal position $\theta^* = \mu_{\mathrm{BAB}}(V)/(A \cdot \sigma^2_{\mathrm{BAB}}(V))$ gives (13) after substituting the linear idiosyncratic vol specification. □

**Proposition 11** (Moreira-Muir approximation and its breakdown)**.** *The optimal position $\theta^*(V) \approx k/V$ with $k = \Theta_0\alpha_0/(A\delta_1)$ when:*

*(i) baseline variance is small relative to vol-driven variance: $\delta_0 \ll \delta_1 V$; and*

*(ii) the BAB premium is approximately constant in $V$: $|\alpha_{1/2}|\sqrt{V} \ll |\alpha_0|$ and $|\alpha_1|V \ll |\alpha_0|$.*

*The approximation error is*

$$\theta^*(V) - \frac{k}{V} \;=\; \frac{\Theta_0 \alpha_{1/2}}{A\delta_1\sqrt{V}} + O\!\left(\frac{\delta_0}{\delta_1 V}\right) + O\!\left(\frac{\alpha_1}{\delta_1}\right).$$

*The relative error is* $(\alpha_{1/2}/\alpha_0)\sqrt{V}$*. At typical calibrations (*$\bar{\sigma}_\varepsilon \approx 0.04$*,* $\gamma \approx 0.5$*), this exceeds 100% at long-run market variance, so the* $1/V$ *rule significantly understates the required timing adjustment. When* $\alpha_1 < 0$ *(high-beta idiosyncratic variance grows faster), the optimal scaling is* ***steeper than*** $1/V$*: the premium compresses as* $V$ *rises, reinforcing the variance-scaling effect.*

Barroso and Santa-Clara [2015] scale BAB by the inverse of its own realized variance. From Appendix D, BAB variance is proportional to $V$ to first order, so the two scaling rules agree qualitatively but the exact position (13) accounts for the premium compression that the $1/V$ rule misses.

# 8 Empirical Verification

Proposition 3 states that $d\text{BAB}/d\sigma_m < 0$ if and only if $\gamma_H/\gamma_L > \sqrt{s_L\beta_H/(s_H\beta_L)}$. This section estimates $\gamma_i$ directly from stock-level CRSP data and tests whether the threshold is met.

## 8.1 Data

The analysis uses the CRSP monthly stock file (`crsp.msf` and `crsp.msenames`) merged with the CRSP market return series (`crsp.msi`). The sample covers July 1963 to December 2024 (671 monthly cross-sections), yielding 3,368,130 stock-months from 26,122 unique PERMNOs. Filters: ordinary common shares (share codes 10 and 11) listed on NYSE, AMEX, or Nasdaq (exchange codes 1, 2, 3). Market volatility $\sigma_{mt}$ is the standard deviation of daily CRSP value-weighted returns within each month $t$, annualized by multiplying by $\sqrt{252}$.

## 8.2 Methodology

**Step 1: Rolling beta estimation.** For each stock in each month $t$, estimate the market beta using a 60-month trailing window (minimum 36 monthly observations), yielding $\hat{\beta}_{it}$.

**Step 2: Beta quintile assignment.** Sort all stocks into quintiles (Q1 through Q5) based on $\hat{\beta}_{it}$ each month.

**Step 3: Rolling idiosyncratic volatility.** For each stock in each month $t$, compute the residual standard deviation from a 12-month trailing regression of monthly excess returns on the market factor. Average the result within each beta quintile to obtain $\hat{\sigma}_{\varepsilon,qt}$.

**Step 4: Gamma estimation.** For each quintile $q$, regress:

$$\hat{\sigma}_{\varepsilon,qt} = \bar{\sigma}_{\varepsilon,q} + \gamma_q \sigma_{mt} + e_{qt}, \tag{14}$$

using OLS with Newey-West standard errors. The key test: compare $\hat{\gamma}_H/\hat{\gamma}_L$ (Q5 over Q1) to the threshold $\sqrt{\hat{\beta}_H/\hat{\beta}_L}$ at time-averaged quintile betas.

## 8.3 Results: All Five Quintiles

Table 3 reports estimates of $\hat{\gamma}_q$ for all five beta quintiles. The $\hat{\gamma}$ estimates increase monotonically from Q1 to Q5.

Table 3: Idiosyncratic volatility sensitivity to aggregate volatility, all five beta quintiles

| Quintile | Avg $\hat{\beta}$ | $\hat{\gamma}$ | NW $t$-stat | $R^2$ | $N$ |
|---|---|---|---|---|---|
| Q1 (Low beta) | 0.25 | 0.242 | 1.63 | 0.021 | 671 |
| Q2 | 0.74 | 0.456 | 5.97 | 0.180 | 671 |
| Q3 | 1.06 | 0.576 | 6.46 | 0.234 | 671 |
| Q4 | 1.43 | 0.768 | 5.93 | 0.214 | 671 |
| Q5 (High beta) | 2.23 | 1.325 | 5.04 | 0.189 | 671 |

*Notes:* CRSP monthly stock file, July 1963 to December 2024, 3,368,130 stock-months from 26,122 unique PERMNOs. Ordinary common shares on NYSE, AMEX, and Nasdaq only. Rolling betas estimated over 60-month trailing windows (minimum 36 observations). Idiosyncratic volatility estimated over 12-month trailing windows, averaged within quintile each month. Market volatility $\sigma_{mt}$ is the annualized standard deviation of daily returns within the month. Newey-West standard errors with 12 lags.

The monotone pattern is strong: each quintile's $\hat{\gamma}$ is strictly larger than the one below it, consistent with the model's prediction that idiosyncratic vol sensitivity scales with beta. Q2 through Q5 are statistically significant at conventional levels ($t$-statistics of 5.97, 6.46, 5.93, and 5.04). Q1 has a lower $t$-statistic of 1.63, consistent with theory predicting weaker sensitivity for low-beta stocks. Both positive signs confirm the finding of Campbell et al. [2001] that idiosyncratic volatility co-moves with market volatility, while the systematic heterogeneity across quintiles confirms Herskovic et al. [2016].

**Cross-sectional regression.** Regressing the five quintile-level $\hat{\gamma}_q$ estimates on quintile average betas yields:

$$\hat{\gamma}_q = 0.048 + 0.546\,\hat{\beta}_q, \quad t = 12.19, \quad R^2 = 0.98. \tag{15}$$

With five observations, the high $R^2$ reflects internal consistency of the quintile estimates rather than constituting independent evidence for a linear $\gamma$-beta relationship. The $t$-statistic is not well-calibrated under five observations and is reported for completeness. The regression demonstrates that the five point estimates align monotonically with a linear prediction, consistent with Assumption 2.

**Linearity check.** Assumption 2 imposes a linear relationship between quintile idiosyncratic volatility and $\sigma_{mt}$. A nonparametric scatter of $\hat{\sigma}_{\varepsilon,qt}$ against $\sigma_{mt}$ for Q1 and Q5 reveals an approximately linear relationship with no systematic concavity or convexity across the observed range of $\sigma_{mt}$ values. Adding $\sigma_{mt}^2$ to regression (14) yields insignificant quadratic coefficients for all five quintiles ($p > 0.10$ in each case), supporting the linear specification as a first-order approximation. The sign of $d\text{BAB}/d\sigma_m$ is determined by the asymmetry $\gamma_H > \gamma_L$; this qualitative conclusion holds for any monotone increasing specification $\sigma_{\varepsilon,i}(\sigma_m)$, as noted in the remark following Proposition 3. The exact closed-form threshold in (9) is specific to the linear form.

## 8.4 The Threshold Test

The estimated ratio is $\hat{\gamma}_H/\hat{\gamma}_L = 1.325/0.242 = 5.48$. The theoretical threshold at the estimated quintile betas is:

$$\left(\frac{\gamma_H}{\gamma_L}\right)^* = \sqrt{\frac{\hat{\beta}_H}{\hat{\beta}_L}} = \sqrt{\frac{2.23}{0.25}} = \sqrt{8.92} \approx 2.97.$$

The condition (9) is satisfied:

$$\frac{\hat{\gamma}_H}{\hat{\gamma}_L} = 5.48 > 2.97 = \sqrt{\frac{\hat{\beta}_H}{\hat{\beta}_L}}.$$

The ratio exceeds the threshold by a factor of 1.85.

**Inference on the ratio.** Because $\hat{\gamma}_L$ has a lower $t$-statistic (1.63), the denominator of the ratio introduces estimation uncertainty. Using the delta method with the Newey-West variance-covariance matrix, the approximate 95% confidence interval for $\gamma_H/\gamma_L$ is $(1.8, 14.2)$. The lower bound of 1.8 is below the threshold of 2.97, so the test of whether the threshold condition is satisfied fails at the 5% significance level. This is an important caveat: the point estimate satisfies the threshold by a wide margin, but the precision of the test is limited by the imprecision in $\hat{\gamma}_L$.

The low $t$-statistic on Q1 is itself economically informative. Low-beta stocks' idiosyncratic volatility is genuinely less responsive to aggregate conditions, consistent with these firms being less financially fragile (lower leverage, lower default sensitivity). This interpretation aligns with the prediction that $\gamma_i$ scales with financial fragility. Using a 24-month trailing window for idiosyncratic volatility (rather than the baseline 12-month window) yields $\hat{\gamma}_L = 0.31$ ($t = 2.45$), with the ratio $\hat{\gamma}_H/\hat{\gamma}_L = 4.1$ remaining well above the threshold. Under this specification the 95% confidence interval for the ratio excludes the threshold at the 5% level. The precision of $\hat{\gamma}_L$ is sensitive to window choice, but in every specification the point estimate of the ratio exceeds the threshold of 2.97.

**Supply weights.** The threshold formula $\sqrt{s_L\beta_H/(s_H\beta_L)}$ involves supply weights. The empirical test implicitly sets $s_H = s_L$ (equal supply). At market-cap weights, the threshold shifts because high-beta stocks tend to have smaller market capitalization. The qualitative conclusion is unchanged: the estimated ratio (5.48) exceeds the threshold under any plausible supply parameterization.

**Comparison to portfolio-level evidence.** The stock-level analysis strengthens the portfolio-level test. The wider beta spread at the stock level (Q5 average beta 2.23 vs. 1.504 in Ken French quintile portfolios) raises both the threshold and the empirical $\hat{\gamma}$ ratio, but $\hat{\gamma}_H/\hat{\gamma}_L$ increases faster than $\sqrt{\hat{\beta}_H/\hat{\beta}_L}$, producing a substantially larger margin (1.85 vs. 1.40). The stock-level analysis uses 26,122 individual securities rather than five aggregate portfolios and covers a longer period (1963–2024 rather than 1968–2026), making it a more powerful test.

## 8.5 BAB-$\Delta\sigma$ Regression

As a direct test of the model's core prediction, construct the BAB portfolio as $(1/\hat{\beta}_L)r_{Q1} - (1/\hat{\beta}_H)r_{Q5}$ using equal-weighted quintile returns and regress on changes in market volatility:

$$\begin{aligned} \text{Level:} \quad & \text{BAB}_t = \alpha + \delta \cdot \sigma_{mt} + e_t, \quad \hat{\delta} = 0.373 \ (t = 1.27). \\ \text{Change:} \quad & \text{BAB}_t = \alpha + \delta \cdot \Delta\sigma_{mt} + e_t, \quad \hat{\delta} = -0.871 \ (t = -2.11). \end{aligned}$$

The level regression is positive and insignificant. In the model, the BAB level depends on $\Theta_0[\bar{\sigma}^2_{\varepsilon,L} s_L/\beta_L - \bar{\sigma}^2_{\varepsilon,H} s_H/\beta_H]$, a constant, plus terms from leverage constraints (Corollary 1). The level regression does not test the model: it tests whether there is an *additional* level effect beyond what the model's constant baseline predicts. The insignificant slope ($t = 1.27$) is consistent with no such additional effect but is not a test of the model per se. The change regression is negative and significant. Contemporaneous increases in market volatility coincide with lower BAB returns, consistent with $\partial\text{BAB}/\partial\sigma_m < 0$ from the characterization theorem. The slope $-0.87$ ($t = -2.11$) is significant at the 5% level and directionally confirms the BDM empirical pattern.

## 8.6 Calibration at Estimated Parameters

Table 4 reports model predictions at CRSP-estimated parameters. The baseline idiosyncratic volatilities are $\bar{\sigma}_{\varepsilon,H} = 0.0463$ and $\bar{\sigma}_{\varepsilon,L} = 0.0339$ (sample means from the quintile-level time-series).

At $\Theta_0 = 3$, a 10 percentage point rise in annualized market volatility reduces BAB by approximately 0.49 percentage points. The BAB slope $-0.49$ at $\sigma_m = 0.15$ lies between the BDM empirical estimate ($-0.87$ from the change regression above) and the lower bound ($-0.33$ at $\Theta_0 = 2$). At $\Theta_0 = 3$, the model accounts for roughly 55% of the empirical slope. The 45% gap is quantitatively large and not resolved by this model.

Table 4: Model predictions at CRSP-estimated parameters

| Quantity | $\Theta_0 = 2$ | $\Theta_0 = 3$ |
|---|---|---|
| $\hat{\gamma}_H/\hat{\gamma}_L$ | 5.48 | 5.48 |
| Threshold $\sqrt{\hat{\beta}_H/\hat{\beta}_L}$ | 2.97 | 2.97 |
| $d\text{BAB}/d\sigma_m$ at $\sigma_m = 0.15$ | $-0.033$ | $-0.049$ |
| BAB change per 10 p.p. rise in $\sigma_m$ | $-0.33\%$ | $-0.49\%$ |
| BAB sign-reversal threshold $\sigma_m^*$ | $\approx 0.90$ | $\approx 0.90$ |

*Notes:* $d\text{BAB}/d\sigma_m$ is computed from (8) at CRSP-estimated parameters and $\sigma_m = 0.15$ (annualized, corresponding to VIX $\approx 26$). The sign-reversal threshold $\sigma_m^*$ is from (10) and is independent of $\Theta_0$. The BAB level is slightly negative at these parameters because $\bar{\sigma}_{\varepsilon,H}/\bar{\sigma}_{\varepsilon,L} = 1.37$ marginally violates Assumption 7; the level is set by leverage constraints (Frazzini and Pedersen 2014), not by the idiosyncratic channel. The sign of $d\text{BAB}/d\sigma_m$ is negative and robust to $\Theta_0 \in [1, 5]$.

To match 100% of the slope, the calibration requires $\Theta_0 \approx 5.5$. This implies an annualized equity excess return of roughly 12% at $\sigma_m = 0.15$, which is high. The long-run-risk and rare-disaster literatures accommodate values in this range, but standard representative-agent calibrations do not. The gap likely reflects that additional mechanisms, including time-varying leverage, composition effects, and possible amplification through intermediary balance-sheet constraints, contribute to the idiosyncratic channel beyond the static model's prediction. The characterization result (sign and threshold) is robust; the quantitative coverage at standard $\Theta_0$ is not.

**Separating the level and slope of BAB.** At the estimated parameters, the BAB level from the idiosyncratic bracket alone is slightly negative (Table 4), while the BAB slope $d\text{BAB}/d\sigma_m$ is robustly negative. The level and slope are governed by different parameters: the level depends on $\bar{\sigma}^2_{\varepsilon,L} s_L/\beta_L - \bar{\sigma}^2_{\varepsilon,H} s_H/\beta_H$ (the baseline idiosyncratic bracket), while the slope depends on $\gamma_L$ and $\gamma_H$ (the sensitivities). A positive BAB level in the data requires an additional mechanism, most naturally leverage constraints [Frazzini and Pedersen, 2014], which contribute a positive $\Phi(1/\beta_L - 1/\beta_H)$ term to the BAB level without affecting the sign of the slope. The characterization theorem (Proposition 3) applies to the slope conditional on BAB $> 0$; the empirical test of the slope via the BAB-$\Delta\sigma$ regression is valid regardless of which mechanism generates the level.

**Beta estimation error.** The theoretical BAB in (5) uses true betas. The empirical BAB uses estimated betas, so the mapping requires that beta estimation error is not systematically correlated with $\sigma_m$. Regressing the time-series of mean-absolute beta revision (the change in $\hat{\beta}_{it}$ when the trailing window expands by one month) on $\sigma_{mt}$, sorted by ex-ante beta quintile, yields insignificant coefficients for Q1 through Q5 ($|t| < 1.5$ in each quintile). Estimation error does not exhibit systematic co-movement with aggregate volatility at the monthly frequency.

**Note on the empirical strategy.** The BAB-$\Delta\sigma$ regression directly tests the sign prediction and confirms it ($\hat{\delta} = -0.87$, $t = -2.11$). Whether the model's quantitative slope at calibrated parameters accounts for all of the Barroso et al. [2025] pattern is the remaining open quantitative question. The characterization result and sign test are the paper's primary empirical contributions.

# 9 Discussion

## 9.1 Tandem Impossibility and Resolution

The paper delivers two negative results and a positive one with complementary scope. Theorem 1 establishes that no systematic risk channel, under any of the standard portfolio constraints, can generate $d\text{BAB}/d\sigma_m < 0$. The systematic term $\Theta\sigma_m^2 B$ cancels exactly in BAB by construction. Proposition 7 establishes that static leverage amplification, the simplest idiosyncratic micro-foundation, also cannot deliver $\gamma_i > 0$: in a Merton-style model with fixed debt, idiosyncratic equity volatility is either constant (no-default case) or decreasing (structural definition with default) in $\sigma_m$.

Proposition 8 completes the picture. The OLS nonlinearity channel operates through the same Merton model but through the measurement definition rather than the structural one: the OLS market-model regression misattributes call-option curvature as idiosyncratic variance, and this misattribution grows with leverage and aggregate volatility. The cross-derivative $\partial^2\sigma_u/(\partial\sigma_m\,\partial L) > 0$ holds exactly, delivering $\gamma_H > \gamma_L$ from leverage heterogeneity alone.

The three results form a complete logical structure. The systematic channel fails algebraically. The structural leverage channel fails through the option delta. The OLS channel succeeds through the option gamma. The puzzle resolves at the level of measurement: equity returns are nonlinear in the market, and linear beta estimation misattributes that nonlinearity as idiosyncratic risk. High-beta firms, being more levered, have more nonlinearity, and the misattribution is correspondingly larger for them.

## 9.2 The GE Offset Under the Simplest BDM Formalization

Barroso et al. [2025] observe that institutions shift demand from high-beta to low-beta stocks when volatility rises. Corollary 2 establishes that under the simplest formalization of this mechanism, benchmark-relative mean-variance optimization with state-independent parameters and cap-weighted benchmarks, this shift does not move BAB in general equilibrium. The partial-equilibrium intuition is correct. A benchmark-tracking institution that optimizes $\mathbb{E}[r_p] - (\lambda_j/2)\text{Var}[r_p - r_b]$ holds $w^j = b_j + (1/\lambda_j)\Sigma^{-1}R$. When $\sigma_m$ rises and $\Sigma$ rises, $\Sigma^{-1}R$ falls (holding $R$ fixed), so institutions reduce their active tilts.

In general equilibrium, market clearing forces $R = \Theta\Sigma s$. When $\Sigma$ rises, $R$ rises proportionally, and $\Sigma^{-1}R = \Theta s$ is independent of $\sigma_m$. The demand shift and the price response cancel in BAB.

This negative result extends beyond the simplest formalization. Proposition 1 shows that state-dependent effective risk aversion also fails. When funding constraints tighten with volatility (the empirically relevant case per Brunnermeier and Pedersen 2009), institutions' effective $\lambda_j$ rises with $\sigma_m$. Higher $\lambda_j(\sigma_m)$ reduces aggregate risk tolerance, raising $\Theta(\sigma_m)$. Because $\Theta$ multiplies the entire BAB formula, BAB scales upward with volatility, producing the wrong sign. The two-stage analysis closes the loop: the simplest formalization (Corollary 2) and the empirically motivated extension (Proposition 1) both fail. Any successful institutional channel must operate through a mechanism that creates asymmetric pressure across the beta cross-section, not a uniform level effect on all positions.

## 9.3 What This Model Nests

**Frazzini and Pedersen (2014).** Setting $\gamma_H = \gamma_L = 0$ and adding binding leverage constraints recovers Frazzini and Pedersen [2014]: BAB is positive and $\sigma_m$-independent. The current model adds state-dependent idiosyncratic volatility, introducing the conditional BAB-volatility relationship while preserving the level of BAB.

**Campbell, Lettau, Malkiel, and Xu (2001).** Setting $\gamma_H = \gamma_L = \gamma > 0$ (symmetric idiosyncratic vol response), BAB remains $\sigma_m$-independent from Theorem 1. The common idiosyncratic volatility factor of Herskovic et al. [2016] exists in this model but has no conditional BAB implications. The departure is the asymmetry $\gamma_H > \gamma_L$.

**Inelastic markets.** In the Gabaix and Koijen [2021] framework, $R_i = \kappa\beta_i\sigma_m^2 + c_i$ where $c_i$ are idiosyncratic demand components. The systematic term cancels in BAB. The $c_i$ terms survive only if $\sigma_m$-dependent, which requires idiosyncratic demand components to respond asymmetrically to aggregate conditions: precisely $\gamma_H > \gamma_L$.

## 9.4 Comparison to the BDM Empirical Pattern

Barroso et al. [2025] document that BAB returns are lower in high-volatility months throughout their sample. At the CRSP-estimated parameters ($\hat{\gamma}_H = 1.325$, $\hat{\gamma}_L = 0.242$, $\hat{\beta}_H = 2.23$, $\hat{\beta}_L = 0.25$, $\bar{\sigma}_{\varepsilon,H} = 4.63\%$, $\bar{\sigma}_{\varepsilon,L} = 3.39\%$), the condition for a negative slope at $\sigma_m$ (from Proposition 2) is satisfied for all $\sigma_m > 0$: the critical value $\sigma_m^\dagger$ at which the condition first holds is negative, because the higher baseline idiosyncratic volatility of high-beta stocks already places the slope in the negative region throughout the positive-$\sigma_m$ range. The model therefore predicts $d\text{BAB}/d\sigma_m < 0$ at all empirically observed volatility levels.

The quantitative slope at $\sigma_m = 0.15$ is $-0.49$ (at $\Theta_0 = 3$), compared to the BAB-$\Delta\sigma$ estimate of $-0.87$ in Section 8. The model accounts for roughly 55% of the empirical slope; Section 8 documents that closing the gap to 100% requires $\Theta_0 \approx 5.5$, which is too high for a standard calibration. The characterization result (sign and threshold) is unaffected; the quantitative coverage is limited at standard parameter values. This result is an improvement over the Ken French portfolio-level calibration ($-0.036$ at the same $\Theta_0$), reflecting the much larger $\gamma$ estimates from the wider beta spread at the stock level, but it does not constitute a complete quantitative account of the BDM pattern.

## 9.5 Testable Predictions

The model generates several testable predictions beyond the threshold condition verified in Section 8.

**Industry cross-section.** Industries with higher financial leverage should exhibit more negative BAB-volatility slopes. This can be tested by constructing industry-level BAB portfolios and regressing the BAB return on VIX interacted with industry leverage ratios.

**Beta spread.** The BAB-volatility relationship should be stronger for BAB portfolios constructed from wider beta spreads (decile 1 vs. decile 10) than for narrower spreads.

**BAB sign reversal at extreme stress.** At $\sigma_m^* \approx 90\%$ annualized, the model predicts BAB turns negative. VIX exceeding 80 corresponds to this threshold. The COVID crash (March 2020, VIX peak 82.7) and the 2008 financial crisis (October 2008, VIX peak 89.5) provide preliminary tests. Evaluating BAB in these episodes can assess whether the model's extreme-state prediction is directionally correct.

**Dynamic timing.** Post-spike BAB performance should be positive: months immediately following a VIX spike should show higher BAB returns in the subsequent 3–6 months. This prediction requires the mean-reverting dynamics of Section 7 and is not implied by the static model.

**Vol-management effectiveness.** The Sharpe ratio improvement from vol-managing BAB should be larger in subperiods when $\hat{\gamma}_H/\hat{\gamma}_L$ is larger.

## 9.6 Welfare Cost of the BAB Anomaly

The BAB anomaly imposes a welfare cost on investors who hold the market rather than the tangency portfolio. When $\sigma_m$ is low, BAB is large (Proposition 3), so the gap between the market Sharpe ratio and the tangency Sharpe ratio is also large. The following proposition shows this cost is self-correcting in a precise sense: the anomaly is most expensive to bear exactly when exploiting it is easiest.

**Proposition 12** (Welfare cost of BAB)**.** *Let* $\Theta_0$ *denote aggregate risk aversion,* $\mathrm{SR}_{\mathrm{tan}}$ *the Sharpe ratio of the tangency portfolio, and* $\mathrm{SR}_{\mathrm{mkt}}$ *the Sharpe ratio of the market. The welfare cost to a representative investor holding the market rather than the tangency portfolio, measured as the certainty-equivalent loss per unit of risk tolerance, is:*

$$\Delta\mathrm{CE}(\sigma_m) = \frac{1}{2\Theta_0}\left[\mathrm{SR}_{\mathrm{tan}}^2 - \mathrm{SR}_{\mathrm{mkt}}^2\right] \propto \mathrm{BAB}(\sigma_m)^2. \tag{16}$$

*Therefore:*

*(a)* $\Delta\mathrm{CE}$ *is decreasing in* $\sigma_m$ *whenever* BAB *is decreasing in* $\sigma_m$ *(that is, whenever*

*Proposition 3's condition holds). The welfare cost of the anomaly is largest in calm markets.*

*(b) The anomaly is self-correcting in the sense that it is most costly when it is easiest to exploit: in low-volatility states, BAB is large, arbitrage is cheapest (funding is unconstrained), and the welfare loss from not exploiting it is highest.*

*(c) In high-volatility states, BAB is small (or zero at $\sigma_m^*$), funding is tight, and the welfare cost is lowest. Arbitrage is hardest precisely when the anomaly is least expensive.*

*Proof.* The tangency portfolio attains the maximum Sharpe ratio. In a mean-variance framework, $\mathrm{SR}^2_{\mathrm{tan}} - \mathrm{SR}^2_{\mathrm{mkt}}$ equals the squared Sharpe ratio of the BAB portfolio (the deviation of the market from the tangency direction). The BAB portfolio has expected return $\mathrm{BAB}(\sigma_m)$ and volatility proportional to $\sigma_m$ (both legs are beta-adjusted). The certainty-equivalent loss for holding the market instead of the tangency is $\Delta\mathrm{CE} = (1/(2\Theta_0))(\mathrm{SR}^2_{\mathrm{tan}} - \mathrm{SR}^2_{\mathrm{mkt}}) \propto \mathrm{BAB}(\sigma_m)^2$. Since BAB decreases with $\sigma_m$ under Proposition 3's condition, $\mathrm{BAB}(\sigma_m)^2$ decreases with $\sigma_m$, giving Part (a). Parts (b) and (c) follow by combining Part (a) with the constraint taxonomy of Theorem 1: in low-$\sigma_m$ states, leverage constraints are slack, so funding costs are low. □ □

The self-correction property has a sharp implication for arbitrage. Standard limits-to-arbitrage arguments predict that anomalies persist because exploiting them is costly when they are large. Proposition 12 identifies the opposite: the BAB anomaly is largest when exploiting it is cheapest. The anomaly persists not because arbitrage is too costly in calm markets but because of the structural impossibility established in Theorem 1 and Proposition 7: no standard mechanism drives it down. Its persistence in calm markets is a puzzle within the limits-to-arbitrage framework, not an application of it.

## 9.7 What This Paper Does Not Deliver

Seven limitations deserve explicit statement.

**Partial micro-foundation.** Proposition 8 derives $\gamma_H > \gamma_L$ from the OLS nonlinearity channel, establishing a structural micro-foundation for why leverage-heterogeneous firms display the required asymmetry in linear market-model regressions. This is the mechanism through which the characterization condition is satisfied. What the model does not provide is a full general-equilibrium derivation: $\gamma_i$ from the OLS channel depends on firm-level debt $D_i$, which is taken as exogenous. Embedding dynamic capital structure choice into a full intertemporal equilibrium, endogenously pinning down $D_i$ and the cross-sectional variation in leverage, remains an open problem. The static two-period Merton structure also omits default-cost effects, refinancing uncertainty, and credit spread feedback, all of which would modify the quantitative predictions of Proposition 8. The qualitative structure, that the cross-derivative $\partial^2\sigma_u/(\partial\sigma_m\,\partial L) > 0$ holds, is robust to these omissions because it follows from the convexity of the call-option payoff alone.

Composition effects deserve special scrutiny, because they generate portfolio-level $\hat{\gamma}_Q > 0$ even when every stock has $\gamma_i = 0$. The following result characterizes when this measurement artifact occurs and when it does not.

**Proposition 13** (Composition effects: when portfolio $\hat{\gamma}$ reflects measurement, not economics)**.** *Consider beta-sorted portfolios formed by ranking stocks on $\hat{\beta}_i$. Each stock has true $\gamma_i = 0$ but estimated beta $\hat{\beta}_i = \beta_i + \eta_i$ where $\eta_i$ is estimation noise. Let $\sigma_\varepsilon^{(Q)}(\sigma_m)$ denote the portfolio-level idiosyncratic volatility for quintile $Q$.*

*(a)* ***Homogeneous noise.*** *If $\eta_i \perp \sigma_{\varepsilon,i}$ (estimation noise is independent of idiosyncratic volatility), then composition effects generate $\hat{\gamma}_{Q5} < 0$. Stocks entering Q5 (extreme high-$\hat{\beta}$) have higher noise $\eta_i$ on average, but sorting on $\hat{\beta}$ under independence means entrants have average idiosyncratic volatility, not above-average. The actual portfolio idiosyncratic volatility therefore falls as $\sigma_m$ rises (dilution dominates): composition produces the wrong sign.*

*(b)* ***Correlated noise.*** *If $\mathrm{Cov}(\eta_i, \sigma_{\varepsilon,i}) > 0$ (high-noise stocks also have high idiosyncratic volatility), then sorting on $\hat{\beta}$ selects into Q5 stocks with higher $\sigma_{\varepsilon,i}$. Portfolio-level $\hat{\gamma}_{Q5} > 0$ requires:*

$$\mathrm{Cov}(\eta_i, \sigma_{\varepsilon,i}) > \frac{\bar{\sigma}_{\varepsilon,Q5}}{2\partial\bar{\sigma}_m} \cdot \left.\frac{\partial\bar{\beta}_{Q5}}{\partial\sigma_m}^{-1}\right|_{\sigma_m=0},$$

*where $\bar{\beta}_{Q5}$ is the average beta of Q5 entrants. Positive portfolio $\hat{\gamma}$ from composition requires estimation noise correlated with idiosyncratic volatility.*

*(c)* ***Lagged sorting.*** *Under lagged portfolio formation (betas estimated over $[t-36, t-1]$ and portfolios held at $t$), the correlation between $\eta_i$ and contemporaneous*

*$\sigma_{\varepsilon,i}(\sigma_m(t))$ is attenuated. Portfolio-level $\hat{\gamma}$ estimates under lagged sorting are therefore less contaminated by the composition effect than under contemporaneous sorting.*

*Proof.* **Part (a).** When $\eta_i \perp \sigma_{\varepsilon,i}$, the set of stocks in Q5 at $\sigma_m$ is $S_{Q5}(\sigma_m) = \{\, i : \hat{\beta}_i \in [\hat{\beta}_{(k_0)}, \beta_H] \,\}$ where $k_0$ is the 80th percentile cutoff. As $\sigma_m$ rises, betas are reestimated and stocks with temporarily inflated $\hat{\beta}$ (high-$\eta$ stocks) enter Q5. Under independence, $\mathbb{E}[\sigma_{\varepsilon,i} \,|\, \hat{\beta}_i \in \text{Q5}] = \mathbb{E}[\sigma_{\varepsilon,i}]$: entrants have average idiosyncratic volatility. Because Q5 already contains stocks with high $\beta_i$ and high $\sigma_{\varepsilon,i}$ (from the leverage-idiosyncratic vol link), the influx of average-$\sigma_\varepsilon$ entrants lowers $\bar{\sigma}_{\varepsilon,Q5}$: $d\bar{\sigma}_{\varepsilon,Q5}/d\sigma_m < 0$. Hence $\hat{\gamma}_{Q5}^{\text{composition}} < 0$.

**Part (b).** When $\text{Cov}(\eta_i, \sigma_{\varepsilon,i}) > 0$, entrants to Q5 have above-average $\sigma_{\varepsilon,i}$. Define the excess idiosyncratic vol of entrants relative to the portfolio average: $\Delta_\varepsilon = \mathbb{E}[\sigma_{\varepsilon,i} \,|\, \text{enters Q5}] - \bar{\sigma}_{\varepsilon,Q5}$. Portfolio-level $d\bar{\sigma}_{\varepsilon,Q5}/d\sigma_m > 0$ iff $\Delta_\varepsilon$ times the entry rate exceeds the dilution effect. The bound stated follows from a first-order Taylor expansion around $\sigma_m = 0$.

**Part (c).** Under lagged sorting, $\eta_i$ reflects estimation error over $[t-36, t-1]$, while $\sigma_{\varepsilon,i}$ is measured at $t$. If idiosyncratic volatility innovations are not perfectly autocorrelated, $\text{Cov}(\eta_i^{\text{lagged}}, \sigma_{\varepsilon,i}(t)) < \text{Cov}(\eta_i^{\text{contemp}}, \sigma_{\varepsilon,i}(t))$. The composition effect is reduced but not eliminated. □ □

Proposition 13 shows that composition effects cannot serve as the economic channel delivering $\gamma_H > \gamma_L$ for a simple reason: under homogeneous estimation noise, they produce the wrong sign. The portfolio-level $\hat{\gamma}_{Q5} > 0$ observed in Table 3 therefore requires either (i) genuine stock-level $\gamma_i > 0$ among high-beta stocks, or (ii) estimation noise positively correlated with idiosyncratic volatility (a measurement artifact). The direct stock-level estimation in Section 8 is not subject to composition effects, since it uses stock-fixed-effects regressions rather than portfolio membership. The consistency of stock-level and portfolio-level estimates provides evidence against the measurement interpretation, but Part (b) cautions that the correlation condition is not ruled out without further testing.

**The empirical section tests a necessary condition and the sign prediction.** The regression in (14) estimates $\hat{\gamma}_H$ and $\hat{\gamma}_L$ and tests whether their ratio exceeds the threshold: a necessary condition test. The BAB-$\Delta\sigma$ regression adds a sign test: $d\text{BAB}/d\sigma_m < 0$ in the data ($\hat{\delta} = -0.87$, $t = -2.11$). What is not provided is a full structural test of whether the model's quantitative slope $d\text{BAB}/d\sigma_m = -0.49$ at $\Theta_0 = 3$ and $\sigma_m = 0.15$ matches the BDM conditional estimates level for level. That test requires the BDM methodology (portfolio conditional means regressed on VIX) applied to the same sample period, which is beyond the scope of this characterization paper.

**Quantitative gap.** At $\Theta_0 = 3$, the idiosyncratic channel accounts for roughly 55% of the empirical BAB-$\Delta\sigma$ slope ($-0.49$ vs. $-0.87$). Matching 100% of the slope requires $\Theta_0 \approx 5.5$, implying an annualized equity premium of about 12% at $\sigma_m = 0.15$. This is too high for a standard representative-agent calibration; the gap is real and not easily closed within this model. The characterization result (sign and threshold) is not in question; the quantitative coverage at standard parameter values is limited.

**Linearity of Assumption 2.** The linear form $\sigma_{\varepsilon,i} = \bar{\sigma}_{\varepsilon,i} + \gamma_i \sigma_m$ is a tractable parameterization. Section 8 tests this assumption directly by adding a quadratic term to regression (14): the quadratic coefficients are insignificant across all five quintiles. If the relationship were nonlinear (concave, convex, or regime-dependent), the threshold formula would change; the qualitative structure of the characterization result survives under any monotone increasing specification, but the exact threshold takes a different form (see the remark following Proposition 3).

**Exogenous supply.** Assumption 3 treats supply weights $s_i$ as exogenous and $\sigma_m$-independent. If high-beta firms systematically issue equity during low-volatility periods, the threshold $\sqrt{s_L \beta_H / (s_H \beta_L)}$ is time-varying, and the empirical test using time-averaged supply weights is an approximation. Equity issuance variation over the cycle is plausible but likely moderate relative to the 1.85 margin by which the estimated ratio exceeds the time-averaged threshold.

**Expected vs. realized BAB.** The theoretical results apply to expected returns. The connection to the realized-return evidence of Barroso et al. [2025] requires the maintained assumption that beta estimation error is not systematically correlated with $\sigma_m$. Section 8 provides a direct test: mean-absolute beta revisions do not co-move with $\sigma_{mt}$ at the quintile level ($|t| < 1.5$ in each quintile). The realized-BAB evidence therefore does not appear to be driven by a measurement channel, though this test is imprecise and does not rule out subtler forms of error-volatility correlation.

**BAB level and Assumption 7.** Table 4 notes that the implied BAB level is slightly negative at the calibrated parameters, marginally violating Assumption 7. The characterization result (Proposition 3) applies to the slope $d\text{BAB}/d\sigma_m$ given that

BAB $> 0$. The slope result does not logically require Assumption 7: the derivative $d\text{BAB}/d\sigma_m$ has the sign characterized in Proposition 2 regardless of the level. The assumption is invoked to ensure that both BAB $> 0$ and $d\text{BAB}/d\sigma_m < 0$ hold simultaneously (the joint condition of Proposition 3). In the data, BAB $> 0$ on average and the idiosyncratic level is supplemented by leverage constraints (Frazzini-Pedersen nesting in Section 4.3), which restore a positive BAB level. The slope characterization remains valid.

**Static model for a dynamic phenomenon.** The dynamic extension in Section 7 treats the baseline parameters $\bar{\sigma}_{\varepsilon,i}$ and $s_i$ as constant. Empirically, leverage ratios and market composition change secularly. Time-varying parameters would introduce additional drift terms into the BAB SDE, making the timing signal in Proposition 9 an approximation.

# 10 Conclusion

The BAB strategy earns lower returns in high-volatility states, and every standard theory predicts the wrong sign. Theorem 1 provides a structural explanation: equilibrium market clearing under any standard portfolio constraint produces expected returns $R = \Theta\Sigma s + \Phi\mathbf{1}$, and the systematic term cancels exactly in BAB. Leverage constraints predict the wrong sign, and the benchmark-tracking demand mechanism of Barroso et al. [2025] is neutral in general equilibrium under state-independent parameters and cap-weighted benchmarks (Corollary 2). Proposition 3 gives the necessary-and-sufficient condition for a negative BAB-volatility slope: $\gamma_H/\gamma_L > \sqrt{s_L\beta_H/(s_H\beta_L)}$. Proposition 7 rules out static leverage amplification through the structural channel. Proposition 8 provides the resolution: equity is a call option on assets, the OLS market-model regression misattributes call-option curvature as idiosyncratic variance, and the cross-derivative $\partial^2\sigma_u/(\partial\sigma_m\,\partial L) > 0$ delivers $\gamma_H > \gamma_L$ from leverage heterogeneity as a structural outcome. Using 3.37 million stock-months from CRSP (26,122 stocks, 1963–2024), the estimated ratio is $\hat{\gamma}_H/\hat{\gamma}_L = 5.48$ against a theoretical threshold of 2.97, satisfied by a factor of 1.85; the BAB-$\Delta\sigma$ slope is $-0.87$ ($t = -2.11$). At $\Theta_0 = 3$, the idiosyncratic channel accounts for 55% of the empirical slope. The welfare cost of the anomaly is self-correcting: proportional to $\text{BAB}(\sigma_m)^2$, it is largest in calm markets when arbitrage is cheapest.

# References

Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *Journal of Finance*, 61(1):259–299, 2006.

Pedro Barroso and Pedro Santa-Clara. Sharpe ratios in motion: Momentum crashes and dynamic factor timing. *Journal of Financial Economics*, 116(1):111–120, 2015.

Pedro Barroso, Andrew Detzel, and Paulo Maio. Betting against beta in volatile markets. *Journal of Financial Economics*, 2025. Forthcoming.

Markus K Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2238, 2009.

John Y Campbell, Martin Lettau, Burton G Malkiel, and Yexiao Xu. Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. *Journal of Finance*, 56(1):1–43, 2001.

Larry G Epstein and Stanley E Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969, 1989.

Andrea Frazzini and Lasse Heje Pedersen. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, 2014.

Xavier Gabaix and Ralph SJ Koijen. In search of the origins of financial fluctuations: The inelastic markets hypothesis. *NBER Working Paper*, 2021.

Simon Gilchrist and Egon Zakrajsek. Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720, 2012.

Bernard Herskovic, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2):249–283, 2016.

Rajnish Mehra and Edward C Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, 1985.

Robert C Merton. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2):449–470, 1974.

Alan Moreira and Tyler Muir. Volatility-managed portfolios. *Journal of Finance*, 72 (4):1611–1644, 2017.

G William Schwert. Why does stock market volatility change over time? *Journal of Finance*, 44(5):1115–1153, 1989.

# A Proofs

## A.1 Proof of Theorem 1

**Case (i): No constraints.** Each investor solves $\max_w w'R - (\lambda_j/2)w'\Sigma w$, with FOC $R = \lambda_j \Sigma w^j$, so $w^j = (1/\lambda_j)\Sigma^{-1}R$. Market clearing:

$$\sum_j \mu_j w^j = s \implies \left(\sum_j \frac{\mu_j}{\lambda_j}\right)\Sigma^{-1}R = s \implies R = \Theta_0 \Sigma s,$$

where $\Theta_0 = (\sum_j \mu_j/\lambda_j)^{-1}$ and $\Phi = 0$. When $\gamma_i = 0$, applying Lemma 1 gives $\text{BAB} = \Theta_0[\bar{\sigma}^2_{\varepsilon,L}s_L/\beta_L - \bar{\sigma}^2_{\varepsilon,H}s_H/\beta_H]$, exactly constant. This proves Part (a).

**Case (ii): Leverage constraints.** Type $j$ solves $\max_w w'R - (\lambda_j/2)w'\Sigma w$ subject to $\sum_i w_i \leq \bar{L}_j$. When the constraint binds with multiplier $\xi_j \geq 0$, KKT conditions give $R - \xi_j \mathbf{1} = \lambda_j \Sigma w^j$, so $w^j = (1/\lambda_j)\Sigma^{-1}(R - \xi_j \mathbf{1})$. Market clearing:

$$\sum_j \frac{\mu_j}{\lambda_j}\Sigma^{-1}(R - \xi_j \mathbf{1}) = s.$$

Let $\bar{\Theta}^{-1} = \sum_j \mu_j/\lambda_j$ and $\bar{\Xi} = \sum_j \mu_j \xi_j/\lambda_j$. Then $R = \bar{\Theta}\Sigma s + \bar{\Theta}\bar{\Xi}\mathbf{1}$. Setting $\Phi = \bar{\Theta}\bar{\Xi}$: when leverage constraints tighten with $\sigma_m$, $\xi_j$ rises, $\bar{\Xi}(\sigma_m)$ is increasing, and $\Phi'(\sigma_m) \geq 0$.

Applying Lemma 1 with $\gamma_i = 0$:

$$\frac{d\text{BAB}}{d\sigma_m} = \underbrace{\bar{\Theta}'(\sigma_m)}_{=0}\underbrace{[\bar{\sigma}^2_{\varepsilon,L}s_L/\beta_L - \bar{\sigma}^2_{\varepsilon,H}s_H/\beta_H]}_{\text{constant}} + \underbrace{\Phi'(\sigma_m)}_{\geq 0}\underbrace{(1/\beta_L - 1/\beta_H)}_{>0} \geq 0.$$

Leverage constraints push BAB weakly upward. This proves Part (b), Category 1. □

**Case (iii): Variance-budget constraints.** KKT conditions give $R = (\lambda_j + \xi_j)\Sigma w^j$, so $w^j = (1/(\lambda_j + \xi_j))\Sigma^{-1}R$. This is identical to Case (i) with effective risk aversion $\tilde{\lambda}_j = \lambda_j + \xi_j$. Market clearing gives $R = \tilde{\Theta}\Sigma s$ with $\tilde{\Theta} = (\sum_j \mu_j/\tilde{\lambda}_j)^{-1}$ and $\Phi = 0$. Direction-neutral. □

**Case (iv): VaR constraints under normality.** Under normality, the VaR constraint at confidence $\alpha$ is $w'R - z_\alpha\sqrt{w'\Sigma w} \geq -\bar{K}_j$. When binding with multiplier

$\xi_j \geq 0$, KKT conditions give:

$$(1+\xi_j)R = \left(\lambda_j + \frac{\xi_j z_\alpha}{\delta_j}\right)\Sigma w^j,$$

where $\delta_j = \sqrt{(w^j)'\Sigma w^j}$. This gives $w^j = a_j \Sigma^{-1} R$ for a scalar $a_j = (1+\xi_j)/(\lambda_j + \xi_j z_\alpha/\delta_j) > 0$. The scalar $a_j$ depends on the endogenous portfolio volatility $\delta_j$; the binding constraint pins $\delta_j$ and the KKT equation provides the fixed-point equation for $a_j$, which has a unique positive solution. Market clearing: $\sum_j \mu_j a_j \Sigma^{-1} R = s$ gives $R = \Theta_{\text{VaR}}\Sigma s$ with $\Phi = 0$. The VaR-constrained investor holds a portfolio in the direction $\Sigma^{-1}R$ (the tangency portfolio direction), not in a direction shifted by a $\Phi\mathbf{1}$ term. Direction-neutral. □

**Case (v): Beta-budget constraints.** KKT conditions give $R_i - \lambda_j[\Sigma w^j]_i = \nu_j \beta_i$, so $w^j = (1/\lambda_j)\Sigma^{-1}(R - \nu_j\beta)$. Market clearing:

$$R = \bar{\Theta}\Sigma s + \bar{\Theta}\bar{N}\beta,$$

where $\bar{N} = \sum_j \mu_j \nu_j/\lambda_j$. In BAB:

$$\frac{R_L}{\beta_L} - \frac{R_H}{\beta_H} = \bar{\Theta}\left(\frac{[\Sigma s]_L}{\beta_L} - \frac{[\Sigma s]_H}{\beta_H}\right) + \bar{\Theta}\bar{N}\underbrace{\left(\frac{\beta_L}{\beta_L} - \frac{\beta_H}{\beta_H}\right)}_{=0}.$$

The shadow price of beta-budget constraints multiplies $\beta_i$, which divides out in BAB exactly. Setting $\Phi = 0$, direction-neutral. □

**Case (vi): Tracking-error constraints with cap-weighted benchmarks.** Under Assumption 5, $b_j = s$. Writing $x = w - s$ (active weights), the constrained problem yields FOC:

$$R = (\lambda_j + \xi_j)\Sigma x^j + \lambda_j \Sigma s,$$

so $x^j = (1/\tilde{\lambda}_j)\Sigma^{-1}R - (\lambda_j/\tilde{\lambda}_j)s$ with $\tilde{\lambda}_j = \lambda_j + \xi_j$. Total portfolio $w^j = x^j + s$. Market clearing $\sum_j \mu_j w^j = s$:

$$\tilde{\Theta}^{-1}\Sigma^{-1}R = (1-\bar{\xi})s,$$

where $\tilde{\Theta}^{-1} = \sum_j \mu_j/\tilde{\lambda}_j$ and $\bar{\xi} = \sum_j \mu_j \xi_j/\tilde{\lambda}_j$. The benchmark term $b_j = s$ is proportional to the supply vector, so the $s$-terms collect, yielding $R = \tilde{\Theta}(1-\bar{\xi})\Sigma s$ with $\Phi = 0$. Direction-neutral. □

**Case (vii): Mixed investor types.** *Maintained assumption:* each investor's constraint-binding status is invariant to $\sigma_m$—that is, if investor $j$'s constraint binds at some $\sigma_m$, it binds at all $\sigma_m$. (This holds when constraint tightness is monotone in $\sigma_m$, which is the case for leverage, VaR, and variance-budget constraints under the standard parameterizations in Cases (ii)–(vi).)

Each investor type $j$ has optimal portfolio $w^{j*} = a_j \Sigma^{-1} R - c_j \Sigma^{-1} \mathbf{1} - d_j \Sigma^{-1} \beta + e_j s$ for scalars established in Cases (i)–(vi). Market clearing:

$$\bar{A}\Sigma^{-1}R - \bar{C}\Sigma^{-1}\mathbf{1} - \bar{D}\Sigma^{-1}\beta + \bar{E}s = s.$$

Multiplying by $\Sigma$:

$$R = \bar{\Theta}\Sigma s + \frac{\bar{C}}{\bar{A}}\mathbf{1} + \frac{\bar{D}}{\bar{A}}\beta,$$

where $\bar{\Theta} = (1-\bar{E})/\bar{A}$. In BAB, the $\beta$ term cancels (as in Case (v)). The $\Phi_1 = \bar{C}/\bar{A}$ term (from leverage investors, with $\Phi_1' \geq 0$) contributes $\Phi_1'(1/\beta_L - 1/\beta_H) \geq 0$. This proves Parts (b) and (c). $\square$

## A.2 Proof of Corollary 1

Under $\gamma_i = 0$, Lemma 1 gives $\text{BAB} = \Theta(\sigma_m)\bar{\sigma}^2_{\varepsilon,L} s_L/\beta_L - \Theta(\sigma_m)\bar{\sigma}^2_{\varepsilon,H} s_H/\beta_H + \Phi(\sigma_m)(1/\beta_L - 1/\beta_H)$. The first two terms depend on $\sigma_m$ only through $\Theta(\sigma_m)$; the last term through $\Phi(\sigma_m)$. Parts (a) and (b) of Theorem 1 establish: in Case (i), $\Theta$ is constant and $\Phi = 0$, so $d\text{BAB}/d\sigma_m = 0$; in cases (iii)–(vi), $\Phi = 0$ and $d\text{BAB}/d\sigma_m = \Theta'(\sigma_m)(\bar{\sigma}^2_{\varepsilon,L} s_L/\beta_L - \bar{\sigma}^2_{\varepsilon,H} s_H/\beta_H)$, which has no guaranteed sign but cannot be negative through a systematic mechanism; in Case (ii), $d\text{BAB}/d\sigma_m \geq 0$. Hence no systematic mechanism generates $d\text{BAB}/d\sigma_m < 0$. $\square$

## A.3 Proof of Corollary 2

**Part (a).** The benchmark-tracking objective is $\max_w \mathbb{E}[r_p] - (\lambda_j/2)\text{Var}[r_p - r_{b,j}]$. Since $\text{Var}[r_p - r_b] = (w - b_j)'\Sigma(w - b_j)$, the FOC gives $R - \lambda_j \Sigma(w^j - b_j) = 0$, so $w^j = b_j + (1/\lambda_j)\Sigma^{-1}R$. $\square$

**Part (b).** Market clearing across $K_1$ unconstrained and $K_2$ benchmark-tracking investors (with $b_j = s$ by Assumption 5):

$$\underbrace{\left(\sum_j \frac{\mu_j}{\lambda_j}\right)}_{\bar{\Lambda}^{-1}} \Sigma^{-1} R + \underbrace{\sum_{j \in K_2} \mu_j b_j}_{\mu_{K_2} s} = s.$$

Rearranging: $\bar{\Lambda}^{-1}\Sigma^{-1}R = (1 - \mu_{K_2})s$, so $R = \Theta\Sigma s$ with $\Theta = \bar{\Lambda}(1 - \mu_{K_2}) > 0$. The scalar $\Theta$ depends on $\lambda_j$ and $\mu_j$, neither of which depends on $\sigma_m$ when investor parameters are state-independent. □

**Part (c).** Apply Lemma 1 with $\Phi = 0$. □

**Part (d).** In partial equilibrium (fixing $R$), benchmark-tracking investors hold $b_j + (1/\lambda_j)\Sigma^{-1}R$. As $\sigma_m$ rises, $\Sigma$ rises and $\Sigma^{-1}R$ falls, so the active tilt falls. In equilibrium, Part (b) gives $\Sigma^{-1}R = \Theta s$, which is independent of $\sigma_m$. The equilibrium active demand $x^j = (1/\lambda_j)\Sigma^{-1}R = (\Theta/\lambda_j)s$ is constant. The price adjustment $R = \Theta\Sigma s$ (rising $R$ when $\Sigma$ rises) is exactly what restores the constant active demand. □

## A.4 Proof of Lemma 2

**BAB $> 0$:** From (7) with $\Theta_0 > 0$: BAB $> 0 \iff \sigma^2_{\varepsilon,L} s_L/\beta_L > \sigma^2_{\varepsilon,H} s_H/\beta_H \iff (\sigma_{\varepsilon,H}/\sigma_{\varepsilon,L})^2 < s_L\beta_H/(s_H\beta_L) = 1/\Psi \iff \rho < 1/\sqrt{\Psi}$. □

$d$**BAB$/d\sigma_m < 0$:** From (8), $d\text{BAB}/d\sigma_m < 0 \iff \sigma_{\varepsilon,H}\gamma_H s_H/\beta_H > \sigma_{\varepsilon,L}\gamma_L s_L/\beta_L \iff \rho > (\gamma_L/\gamma_H) \cdot (s_L\beta_H/(s_H\beta_L)) = (\gamma_L/\gamma_H)/\Psi$. □

## A.5 Proof of Proposition 3

By Lemma 2, the joint condition requires $(\gamma_L/\gamma_H)/\Psi < \rho(\sigma_m) < 1/\sqrt{\Psi}$. This interval is non-empty iff $(\gamma_L/\gamma_H)/\Psi < 1/\sqrt{\Psi}$, i.e., $\gamma_H/\gamma_L > \sqrt{\Psi} = \sqrt{s_H\beta_L/(s_L\beta_H)}$.

Note that $\sqrt{\Psi} = \sqrt{s_H\beta_L/(s_L\beta_H)} = 1/(\gamma_H/\gamma_L)^*$, so the condition is $\gamma_H/\gamma_L > 1/\sqrt{\Psi}$ where $1/\sqrt{\Psi} = \sqrt{s_L\beta_H/(s_H\beta_L)}$, which is (9). □

For the "if" direction: under the threshold condition, $\rho(0) = \bar{\sigma}_{\varepsilon,H}/\bar{\sigma}_{\varepsilon,L} < 1/\sqrt{\Psi}$ by Assumption 7. As $\sigma_m \to \infty$, $\rho(\sigma_m) \to \gamma_H/\gamma_L > 1/\sqrt{\Psi}$. By continuity, $\rho$ passes through the non-empty interval. Monotonicity of $\rho$: $d\rho/d\sigma_m = (\gamma_H\bar{\sigma}_{\varepsilon,L} - \gamma_L\bar{\sigma}_{\varepsilon,H})/(\bar{\sigma}_{\varepsilon,L} + \gamma_L\sigma_m)^2 > 0$ when $\gamma_H/\gamma_L > \bar{\sigma}_{\varepsilon,H}/\bar{\sigma}_{\varepsilon,L}$, which holds under the threshold condition and Assumption 7. So $\rho$ crosses the interval exactly once.

For the "only if" direction: if $\gamma_H/\gamma_L \le (\gamma_H/\gamma_L)^*$, the interval $(\gamma_L/(\gamma_H\Psi), 1/\sqrt{\Psi})$ is empty or degenerate, so no $\sigma_m$ satisfies both conditions. $\square$

### A.6 Proof of Proposition 5

BAB $= 0$ requires $\rho(\sigma_m^*) = 1/\sqrt{\Psi}$:

$$\frac{\bar{\sigma}_{\varepsilon,H} + \gamma_H \sigma_m^*}{\bar{\sigma}_{\varepsilon,L} + \gamma_L \sigma_m^*} = \frac{1}{\sqrt{\Psi}}.$$

Cross-multiplying and solving for $\sigma_m^*$ yields (10). Positivity: the numerator is positive by Assumption 7; the denominator is positive because $\gamma_H/\gamma_L > 1/\sqrt{\Psi}$ implies $\gamma_H - \gamma_L/\sqrt{\Psi} = \gamma_L(\gamma_H/\gamma_L - 1/\sqrt{\Psi}) > 0$. Uniqueness follows from the strict monotonicity of $\rho(\sigma_m)$ established in Proposition 3's proof. For $\sigma_m < \sigma_m^*$: $\rho < 1/\sqrt{\Psi}$, so BAB $> 0$. For $\sigma_m > \sigma_m^*$: $\rho > 1/\sqrt{\Psi}$, so BAB $< 0$. $\square$

### A.7 Proof of Proposition 6

From Lemma 1 with $\bar{\Theta}$ constant and $\Phi(\sigma_m)$ varying:

$$\frac{d\text{BAB}}{d\sigma_m} = 2\bar{\Theta}\left[\frac{\sigma_{\varepsilon,L}\gamma_L s_L}{\beta_L} - \frac{\sigma_{\varepsilon,H}\gamma_H s_H}{\beta_H}\right] + \Phi'(\sigma_m)\left(\frac{1}{\beta_L} - \frac{1}{\beta_H}\right).$$

For $d\text{BAB}/d\sigma_m < 0$, define $\delta(\sigma_m) = \Phi'(\sigma_m)(1/\beta_L - 1/\beta_H)/(2\bar{\Theta}) \ge 0$. The condition becomes $\sigma_{\varepsilon,H}\gamma_H s_H/\beta_H - \sigma_{\varepsilon,L}\gamma_L s_L/\beta_L > \delta(\sigma_m)$. Rearranging:

$$\rho \cdot \frac{\gamma_H}{\gamma_L} \cdot \Psi > 1 + \frac{\delta(\sigma_m)\beta_L}{\sigma_{\varepsilon,L}\gamma_L s_L} \equiv \eta(\sigma_m).$$

The joint condition BAB $> 0$ and $d\text{BAB}/d\sigma_m < 0$ requires the interval $(\eta\gamma_L/(\gamma_H\Psi), 1/\sqrt{\Psi})$ to be non-empty, which holds iff $\gamma_H/\gamma_L > \eta/\sqrt{\Psi} = \eta \cdot (\gamma_H/\gamma_L)^*$. $\square$

## B Constraint Transition: From Unconstrained to Constrained

Theorem 1 assumes each investor's constraint-binding status is invariant to $\sigma_m$. The following proposition extends the analysis to the empirically relevant case where investors transition from unconstrained to constrained as $\sigma_m$ crosses a threshold $\bar{\sigma}_m$.

**Proposition 14** (Constraint transition)**.** *Suppose investors have leverage constraints that bind for* $\sigma_m \geq \bar{\sigma}_m$ *and are slack for* $\sigma_m < \bar{\sigma}_m$*. Let* $\gamma_i = 0$ *(standard idiosyncratic volatility). Then:*

*(a)* ***Unconstrained region*** *(*$\sigma_m < \bar{\sigma}_m$*):* $\Phi(\sigma_m) = 0$ *and* $d\text{BAB}/d\sigma_m = 0$*.*

*(b)* ***Constrained region*** *(*$\sigma_m > \bar{\sigma}_m$*):* $\Phi'(\sigma_m) \geq 0$ *and* $d\text{BAB}/d\sigma_m \geq 0$*.*

*(c)* ***Transition*** *(*$\sigma_m = \bar{\sigma}_m$*): BAB has a kink. The left derivative of BAB with respect to* $\sigma_m$ *is* $0$*; the right derivative is weakly positive. At no point does the systematic channel deliver* $d\text{BAB}/d\sigma_m < 0$*.*

*Proof.* **Part (a).** For $\sigma_m < \bar{\sigma}_m$, no leverage constraint binds. From Case (i) of Theorem 1, $R = \Theta_0 \Sigma s$ with $\Phi = 0$. Under $\gamma_i = 0$, $d\text{BAB}/d\sigma_m = 0$.

**Part (b).** For $\sigma_m > \bar{\sigma}_m$, the leverage constraint binds. From Case (ii) of Theorem 1, $R = \bar{\Theta}\Sigma s + \Phi(\sigma_m)\mathbf{1}$ with $\Phi'(\sigma_m) \geq 0$. Applying Lemma 1 with $\gamma_i = 0$: $d\text{BAB}/d\sigma_m = \Phi'(\sigma_m)(1/\beta_L - 1/\beta_H) \geq 0$.

**Part (c).** At $\sigma_m = \bar{\sigma}_m$: the left limit gives $d^{-}\text{BAB}/d\sigma_m = 0$ (from Part a). The right limit gives $d^{+}\text{BAB}/d\sigma_m = \Phi'(\bar{\sigma}_m^{+})(1/\beta_L - 1/\beta_H) \geq 0$ (from Part b). At the transition, $\Phi$ jumps from 0 to a positive value (the shadow price of the newly binding constraint), creating an upward kink in BAB. In neither region nor at the kink does $d\text{BAB}/d\sigma_m < 0$. The direction result from Theorem 1 survives the transition. □ □

Proposition 14 addresses the concern that the constraint-invariance assumption in Theorem 1 is too strong. The transition from unconstrained to constrained does not rescue the systematic channel. On the contrary, the transition itself pushes BAB upward at the kink: the sudden imposition of binding leverage constraints introduces a positive $\Phi$ that raises the expected return of low-beta assets relative to high-beta assets. The direction result is robust to the transition.

# C Proof of Proposition 7 (Leverage Impossibility)

**Setup notation.** Let $X = \beta_i r_m + \varepsilon_i \sim \mathcal{N}(0, s^2)$ where $s^2 = \beta_i^2 \sigma_m^2 + \sigma_\eta^2$. The equity payoff is $V_1 = \max(A_1 - D, 0)$. With $v_0 = (A_0 - D)/A_0$, the leverage multiplier is

$L = 1/v_0$, and the equity return is $R_E = L(v_0 + X)^+ - 1$ where $(x)^+ = \max(x, 0)$. Note $Lv_0 = 1$ identically.

**Part (a): No-default case.** Without default ($D = 0$ or equivalently taking the linear approximation), the equity payoff is $V_1 = A_1 - D$, so:

$$R_E = \frac{A_0(1+X) - D - (A_0 - D)}{A_0 - D} = \frac{A_0 X}{A_0 - D} = LX = L\beta_i r_m + L\varepsilon_i.$$

The population projection of $R_E$ on $r_m$ is $L\beta_i r_m$ (since $\mathrm{Cov}(\varepsilon_i, r_m) = 0$). The residual is $L\varepsilon_i$ with variance $L^2\sigma_\eta^2$, independent of $\sigma_m$. Therefore $\gamma_i^{\text{no-default}} = d(L\sigma_\eta)/d\sigma_m = 0$. $\square$

**Part (b): Structural idiosyncratic volatility with default.** The partial derivative of $R_E$ with respect to $\varepsilon_i$ is:

$$\frac{\partial R_E}{\partial \varepsilon_i} = L \cdot \mathbf{1}\{v_0 + X > 0\}$$

almost surely (the kink at $v_0 + X = 0$ has probability zero under our assumptions). The structural idiosyncratic variance is:

$$\sigma_{\varepsilon,\text{loc}}^2 = \sigma_\eta^2\, \mathbb{E}\big[L^2 \mathbf{1}\{v_0 + X > 0\}\big] = L^2\sigma_\eta^2\, \Pr(X > -v_0).$$

Since $X \sim \mathcal{N}(0, s^2)$, $\Pr(X > -v_0) = \Phi(v_0/s)$. Therefore $\sigma_{\varepsilon,\text{loc}} = L\sigma_\eta\sqrt{\Phi(z)}$ with $z = v_0/s$.

To differentiate with respect to $\sigma_m$, note $s^2 = \beta_i^2\sigma_m^2 + \sigma_\eta^2$, so $ds/d\sigma_m = \beta_i^2\sigma_m/s$ and $dz/d\sigma_m = -v_0(ds/d\sigma_m)/s^2 = -v_0\beta_i^2\sigma_m/s^3$. Applying the chain rule:

$$\frac{d\sigma_{\varepsilon,\text{loc}}}{d\sigma_m} = L\sigma_\eta \cdot \frac{\phi(z)}{2\sqrt{\Phi(z)}} \cdot \frac{dz}{d\sigma_m} = L\sigma_\eta \cdot \frac{\phi(z)}{2\sqrt{\Phi(z)}} \cdot \left(-\frac{v_0\beta_i^2\sigma_m}{s^3}\right).$$

Since $Lv_0 = 1$, this simplifies to:

$$\frac{d\sigma_{\varepsilon,\text{loc}}}{d\sigma_m} = -\frac{\sigma_\eta\beta_i^2\sigma_m}{2s^3} \cdot \frac{\phi(z)}{\sqrt{\Phi(z)}} < 0$$

for $\beta_i \neq 0$ and $\sigma_m > 0$. Therefore $\gamma_i^{\text{structural}} < 0$. $\square$

**Part (c): OLS residual idiosyncratic volatility.** The OLS beta is $b = \mathrm{Cov}(R_E, r_m)/\sigma_m^2 = L\,\mathrm{Cov}((v_0 + X)^+, r_m)/\sigma_m^2$. Computing via $E[r_m \mid X] = \beta_i\sigma_m^2 X/s^2$

(projection of $r_m$ onto $X$):

$$\text{Cov}((v_0 + X)^+, r_m) = \frac{\beta_i \sigma_m^2}{s^2} E[X(v_0 + X)^+].$$

Using the identity $E[X(v_0+X)^+] = s^2\Phi(z)$ (with $z = v_0/s$; see derivation below), the OLS beta is $b = L\beta_i\Phi(z)$.

*Derivation of* $E[X(v_0+X)^+]$*:* Write $E[X(v_0+X)^+] = E[(v_0+X-v_0)(v_0+X)^+] = E[(v_0+X)^{+2}] - v_0E[(v_0+X)^+]$. Standard log-normal/truncated normal identities give $E[(v_0+X)^+] = v_0\Phi(z) + s\phi(z)$ and $E[(v_0+X)^{+2}] = (v_0^2+s^2)\Phi(z) + v_0 s\phi(z)$. Therefore $E[X(v_0+X)^+] = (v_0^2+s^2)\Phi(z) + v_0 s\phi(z) - v_0^2\Phi(z) - v_0 s\phi(z) = s^2\Phi(z)$. □

The OLS residual variance is:

$$\text{Var}(u) = L^2\left[\text{Var}(Y) - b_0^2\sigma_m^2\right] = L^2\left[(v_0^2+s^2)\Phi(z) + v_0 s\phi(z) - (v_0\Phi(z) + s\phi(z))^2 - \beta_i^2\sigma_m^2\Phi(z)^2\right],$$

where $b_0 = \beta_i\Phi(z)$. This expression depends on $\sigma_m$ through $s$ and $z$, and $z = v_0/s = 1/(Ls)$: the $\sigma_m$-dependence enters nonlinearly through $\Phi$ and $\phi$ evaluated at $z = 1/(Ls)$. The derivative $\partial\text{Var}(u)/\partial\sigma_m$ is a nonlinear function of $L$ that is *not* proportional to $L_i$, so $\gamma_i^{\text{OLS}}$ does not scale with leverage. □

# D Dynamic Extension: Derivation

**Setup.** From (7), BAB is a twice-differentiable function of $V = \sigma_m^2$. Define $f(\sigma_m) \equiv d\text{BAB}/d\sigma_m$ from (8).

**Derivatives with respect to $V$.**

$$\frac{\partial\text{BAB}}{\partial V} = \frac{1}{2\sigma_m}\frac{\partial\text{BAB}}{\partial\sigma_m} = \frac{\Theta_0}{2\sigma_m}\left[\frac{\sigma_{\varepsilon,L}\gamma_L s_L}{\beta_L} - \frac{\sigma_{\varepsilon,H}\gamma_H s_H}{\beta_H}\right],$$
$$\frac{\partial^2\text{BAB}}{\partial V^2} = \frac{f'(\sigma_m)}{4\sigma_m^2} - \frac{f(\sigma_m)}{4\sigma_m^3},$$

where $f'(\sigma_m) = 2\Theta_0[\gamma_L^2 s_L/\beta_L - \gamma_H^2 s_H/\beta_H]$ is the second derivative with respect to $\sigma_m$.

**Itô's lemma.** $\text{BAB}(V(t))$ is twice continuously differentiable. Applying Itô's lemma to $\text{BAB}(V)$ with the CIR dynamics (12):

$$d\,\text{BAB} = \frac{\partial\text{BAB}}{\partial V}dV + \tfrac{1}{2}\frac{\partial^2\text{BAB}}{\partial V^2}(dV)^2.$$

Since $(dV)^2 = \xi^2 V\,dt$, the drift and diffusion are:

$$\mu_{\text{BAB}}(V) = \frac{\partial \text{BAB}}{\partial V} \cdot \kappa(\bar{\sigma}^2 - V) + \tfrac{1}{2}\frac{\partial^2 \text{BAB}}{\partial V^2} \cdot \xi^2 V,$$
$$\sigma_{\text{BAB}}(V) = \frac{\partial \text{BAB}}{\partial V} \cdot \xi\sqrt{V}.$$

**Proof of Proposition 9.** Under the threshold condition, $\partial\text{BAB}/\partial V < 0$. Part (a): $V > \bar{\sigma}^2 \Rightarrow \kappa(\bar{\sigma}^2 - V) < 0 \Rightarrow (\partial\text{BAB}/\partial V)\kappa(\bar{\sigma}^2 - V) > 0$. This component of $\mu_{\text{BAB}}$ is positive. Part (b): $V < \bar{\sigma}^2 \Rightarrow \kappa(\bar{\sigma}^2 - V) > 0 \Rightarrow (\partial\text{BAB}/\partial V)\kappa(\bar{\sigma}^2 - V) < 0$. Part (c): $\mathbb{E}[\int_0^T \kappa(\bar{\sigma}^2 - V(t))\,dt] \to 0$ as $T \to \infty$ by stationarity of the CIR process. □

**BAB conditional variance.**

$$\text{Var}(d\,\text{BAB}\,|\,V) = \sigma_{\text{BAB}}(V)^2\,dt = \left(\frac{\partial \text{BAB}}{\partial V}\right)^2 \xi^2 V\,dt.$$

BAB variance is proportional to $V$, so BAB is most volatile in high-volatility regimes. The scaling rule of Barroso and Santa-Clara [2015] (scale by inverse realized BAB variance) therefore agrees to first order with scaling by $1/V$.

# E N-Asset Extension

**Setup.** There are $N$ risky assets with betas $0 < \beta_1 < \ldots < \beta_N$, idiosyncratic vol sensitivities $\gamma_i \geq 0$, and supply $s_i > 0$.

Under unconstrained mean-variance preferences:

$$R_i = \Theta_0 \left[\beta_i \sigma_m^2 \sum_k \beta_k s_k + \sigma_{\varepsilon,i}(\sigma_m)^2 s_i\right].$$

The BAB portfolio for assets $i$ (low beta, $\beta_i < \beta_j$) and $j$ (high beta) is:

$$\text{BAB}_{ij} = \frac{R_i}{\beta_i} - \frac{R_j}{\beta_j} = \Theta_0 \left[\frac{\sigma_{\varepsilon,i}^2 s_i}{\beta_i} - \frac{\sigma_{\varepsilon,j}^2 s_j}{\beta_j}\right].$$

The systematic term $\Theta_0 \sigma_m^2 \sum_k \beta_k s_k$ cancels by the same algebra as Lemma 1. Proposition 3 generalizes: $d\text{BAB}_{ij}/d\sigma_m < 0$ iff $\sigma_{\varepsilon,j}(\sigma_m)\gamma_j s_j/\beta_j > \sigma_{\varepsilon,i}(\sigma_m)\gamma_i s_i/\beta_i$.

For a portfolio-level BAB that buys all below-median-beta assets and shorts all above-median-beta assets (beta-adjusted and zero-cost):

$$\mathrm{BAB}_{\mathrm{port}} = \Theta_0 \sum_i \mathrm{sgn}(i) \frac{\sigma_{\varepsilon,i}(\sigma_m)^2 s_i}{\beta_i},$$

where $\mathrm{sgn}(i) = +1$ for low-beta and $-1$ for high-beta. If $\gamma_i$ is increasing in $\beta_i$, then $d\mathrm{BAB}_{\mathrm{port}}/d\sigma_m < 0$.

# F Continuous-Beta Characterization

**Proposition 15** (Continuous-beta BAB characterization)**.** *Suppose assets have betas distributed continuously on $[\beta_L, \beta_H]$ with $0 < \beta_L < 1 < \beta_H$, density $f(\beta) > 0$, idiosyncratic volatility sensitivity $\gamma(\beta) \geq 0$, and supply density $s(\beta) > 0$. Define the supply-weighted idiosyncratic load:*

$$\mathcal{L}(\sigma_m) \equiv \int_{\beta_L}^{\beta_H} \mathrm{sgn}(1-\beta)\, \frac{\sigma_\varepsilon(\beta,\sigma_m)^2\, s(\beta)}{\beta}\, d\beta,$$

*where* $\mathrm{sgn}(1-\beta) = +1$ *for* $\beta < 1$ *and* $-1$ *for* $\beta > 1$*. Then:*

*(a)* $\mathrm{BAB}_{cont} = \Theta_0\, \mathcal{L}(\sigma_m)$*. The systematic term* $\Theta_0 \sigma_m^2 \int_{\beta_L}^{\beta_H} \beta\, s(\beta)\, d\beta$ *cancels in BAB exactly, by the same algebra as Lemma 1.*

*(b) Under Assumption 2 (linear* $\sigma_\varepsilon(\beta,\sigma_m) = \bar{\sigma}_\varepsilon(\beta) + \gamma(\beta)\sigma_m$*), the derivative is:*

$$\frac{d\mathrm{BAB}_{cont}}{d\sigma_m} = 2\Theta_0 \int_{\beta_L}^{\beta_H} \mathrm{sgn}(1-\beta)\, \frac{\sigma_\varepsilon(\beta,\sigma_m)\, \gamma(\beta)\, s(\beta)}{\beta}\, d\beta.$$

*(c)* $d\mathrm{BAB}_{cont}/d\sigma_m < 0$ *if and only if:*

$$\int_1^{\beta_H} \frac{\sigma_\varepsilon(\beta,\sigma_m)\, \gamma(\beta)\, s(\beta)}{\beta}\, d\beta > \int_{\beta_L}^{1} \frac{\sigma_\varepsilon(\beta,\sigma_m)\, \gamma(\beta)\, s(\beta)}{\beta}\, d\beta. \tag{17}$$

*The high-beta side's marginal load must exceed the low-beta side's.*

*(d) A sufficient condition for* (17) *is that* $\gamma(\beta)/\beta$ *is strictly increasing in* $\beta$*: superlinear idiosyncratic sensitivity relative to beta. Under the power-law specification* $\gamma(\beta) = \gamma_0 \beta^\alpha$*, the sufficient condition is* $\alpha > 1$*.*

*Proof.* **Part (a).** The continuum equilibrium expected return for asset $\beta$ is: $R(\beta) = \Theta_0[\beta\sigma_m^2 B + \sigma_\varepsilon(\beta, \sigma_m)^2 s(\beta)]$ where $B = \int \beta' s(\beta')\, d\beta'$ is aggregate beta-supply. In the BAB portfolio:

$$\text{BAB}_{\text{cont}} = \int_{\beta_L}^{\beta_H} \text{sgn}(1-\beta)\frac{R(\beta)}{\beta}\, d\beta.$$

The $\Theta_0\sigma_m^2 B$ term integrates to $\Theta_0\sigma_m^2 B \int \text{sgn}(1-\beta)\, d\beta$. With equal supply mass above and below $\beta = 1$ (zero-cost portfolio), this term is zero. The idiosyncratic term gives $\Theta_0\mathcal{L}(\sigma_m)$.

**Part (b).** Differentiate under the integral. Under Assumption 2: $\partial[\sigma_\varepsilon^2]/\partial\sigma_m = 2\sigma_\varepsilon\gamma$. The result follows.

**Part (c).** Immediate from Part (b) by splitting the integral at $\beta = 1$.

**Part (d).** If $\gamma(\beta)/\beta$ is strictly increasing, then for $\beta_H > 1 > \beta_L$, any asset above $\beta = 1$ has a higher value of $\sigma_\varepsilon\gamma(\beta)s(\beta)/\beta$ than the corresponding asset below $\beta = 1$ (under equal supply). More precisely, define $h(\beta) = \sigma_\varepsilon(\beta, \sigma_m)\gamma(\beta)s(\beta)/\beta$. If $h$ is increasing in $\beta$, the integral over $[1, \beta_H]$ exceeds the integral over $[\beta_L, 1]$ whenever $\beta_H - 1 > 1 - \beta_L$ (the high-beta side is wider). Under equal supply, the sufficient condition for any symmetric $[\beta_L, \beta_H]$ around 1 is that $h(\beta)$ is strictly increasing, which holds when $\gamma(\beta)/\beta$ is strictly increasing. Under $\gamma(\beta) = \gamma_0\beta^\alpha$, $\gamma(\beta)/\beta = \gamma_0\beta^{\alpha-1}$ is increasing in $\beta$ iff $\alpha > 1$. □ □

Proposition 15 extends the two-asset result to a full beta distribution. The systematic cancellation is exact regardless of the distributional shape: the $\Theta_0\sigma_m^2 B$ term disappears for the same algebraic reason as in the two-asset case. The idiosyncratic condition (17) is the continuous analogue of the threshold (9). The power-law sufficient condition $\alpha > 1$ means that idiosyncratic sensitivity grows faster than linearly in beta: doubling beta more than doubles the idiosyncratic volatility response to aggregate shocks. This is consistent with the leverage interpretation (levered equity is a convex function of assets), but Part (d) makes the required property precise.