Resolving the Equity Premium Requires at Least Two Dimensions: A Machine-Checked Class No-Go and a Term-Structure Discriminator

Dr. Tamás Nagy

tnagyphd@gmail.com

Working Paper

Build-std v2.0+8b958d430d | 2026-07-05 13:22 | 6a717a3e

Abstract

The equity premium puzzle of Mehra and Prescott (1985) is the observation that the standard consumption-based asset pricing model, calibrated to plausible risk aversion and the observed smoothness of aggregate consumption, predicts an equity premium roughly an order of magnitude smaller than the ~6% historically realized in U.S. data.

The central result of this paper is a structural one: resolving the puzzle requires a rank-≥ 2 mechanism. No single bounded scalar — pushed through *any* monotone amplifier, once that scalar is held to its empirically admissible ceiling — can reach the observed premium; at least two independent amplification axes are needed. We establish this two independent ways. *In theory:* a machine-checked class no-go shows that the entire family of single-parameter resolutions (long-run risk, external habit, and cross-sectional concentration are all the same map $1/(1-x)$) is capped below the premium, and an axis-dimension theorem proves that any successful composition must move both independent axes strictly above neutral. *In the data:* the equity term structure — which a single priced factor cannot match (Gormsen 2021) — demands exactly the same two-factor structure, and its beta-convergence rate (and, on the reading we adopt, its downward slope, though that slope's direction is itself debated — Bansal, Miller, Song and Yaron 2021 — see §5.7) is reproduced by an independently measured network mode with no premium fit — where the beta-convergence rate is a *cross-sectional beta pattern* distinct from the risk-premium slope those authors dispute.

The constructive side then runs the other way: the two channels we actually measure (concentration $\times$ persistence) compose — multiplicatively, a factorization that follows from the two channels' separability rather than being separately fitted — to clear the premium with no parameter fitted to it, *once the persistence axis is priced by recursive preferences*; under strict power utility that axis is variance-weighted to near-neutral (its *effective* amplification $\rho_{\text{eff}} \approx 1$, where $\rho > 1$ is the latent single-parameter amplifier defined in §1 and §4.5), which does not rescue rank-1 but sharpens the rank-≥ 2 reading (the time axis alone is then inert). The rank-≥ 2 verdict is thus a single theorem that speaks about the whole literature at once, corroborated by the sharpest observable that can tell the candidate mechanisms apart.

Read across the literature, this reframes four decades of resolutions with one thesis: **the quantitatively successful models did not find a better stochastic discount factor — they found a second dimension.** Bansal–Yaron works because stochastic volatility is an axis independent of growth persistence; Wachter works because time-varying disaster intensity is independent of disaster size. The criterion says this is not incidental but forced: rank-1 is provably capped, so

any resolution that works must be rank-2, and "how many independent priced dimensions?" is the precise question the equity-premium literature has been answering implicitly for forty years. This paper makes that question, and its answer, a theorem.

The argument is built in three fenced epistemic boxes, and we are deliberate about keeping them separate.

First, a verified impossibility. We formalize the stochastic discount factor (SDF) representation of asset prices, the Hansen–Jagannathan volatility bound, and the Mehra–Prescott calibration constraint. We then machine-check — as exact algebra, with zero axioms — that under the standard model's own assumptions, the model-implied premium is bounded well below the observed level. This is not a new economic claim; it is a theorem-level restatement of a known result, now carrying a kernel certificate. It is the rigorous core of the paper.

Second, a framework hypothesis. We then introduce a single *latent amplification parameter* $\rho > 1$, and prove that *conditional on* ρ taking values in a specified range, the same algebra reproduces a premium in excess of 6%. We give two routes to the same conclusion — an AR(1) persistence channel and a spectral (geometric-series) channel — and prove they coincide at $\rho = 2$. We state plainly that these are **sufficiency and internal-consistency results about the framework, not an empirical demonstration**: the value of ρ is posited, not estimated from the premium.

Third, an empirical confrontation — including a falsification. The persistence channel pins $\rho = 1/(1-\phi)$ to the estimable persistence of consumption growth, so it can be tested without circularity — and the naive reading is *rejected at face value*: post-war U.S. consumption growth ($\hat{\phi} \approx 0.05$) implies $\hat{\rho} \approx 1$, an order of magnitude below the $\rho \approx 4.6$ the gap requires. The latent-component reading (a small, highly persistent state-space factor) is *admissible but unconfirmed*.

Relocating ρ to *cross-sectional spectral concentration*, we measure it from the real BEA production network: it is robust ($\rho \approx 1.4$, stable 1997–2023) and nearly equal to the independently-estimated factor-anomaly $\rho \approx 1.5$ — a genuine cross-domain agreement. It still leaves a wedge (the premium demands $\rho^* \approx 4.6$), which we turn into the machine-checked no-go and the full impossibility frontier, then lift both to the rank-1 *class*: because long-run risk, habit, and concentration are all one map $1/(1-x)$, the impossibility that sinks one sinks all of them.

The one constructive result runs the other way: the two *measured* channels (concentration $\times$ persistence) compose multiplicatively — a factorization derived from their separability, and the orthogonality that premise needs holds in the data (across 18 economies the model-relevant Perron-root axis is uncorrelated with persistence, $|r| \leq 0.19$) — closing the wedge with no fitted parameter, *provided the persistence axis is priced by recursive preferences*: under power utility its variance-weighted contribution is $\rho_{\text{eff}} \approx 1$ (§4.5), so the composition is a statement about a recursive-preference economy, and the rank-≥ 2 requirement is only sharpened by this. Finally, the equity term structure tells the two readings apart: concentration front-loads risk (a downward slope, the van Binsbergen sign, with the measured $\lambda_1 \approx 0.48$ reproducing the observed beta convergence at no premium fit), where long-run risk and habit imply the opposite. We report the negatives and partial agreements plainly — they are what make the framework falsifiable rather than merely suggestive, and they keep the verified algebra from lending borrowed credibility to the conjectures.

The argument rests on a handful of load-bearing theorems — the class no-go, the axis-dimension necessity, and the term-structure discriminator. These, together with the supporting lemmas that establish them (163 in total), are verified in the formal proof kernel (0 axioms, 0 sorry, 0 derivation debt) and exportable to Lean 4. ("Zero axioms" means no unproven escape hatches *inside* the

kernel; the model's economic premises enter as three *explicit* hypotheses — listed in Section 7 — not as hidden assumptions.)

Epistemic status (read this first)

This paper mixes two kinds of statement. We label them explicitly so the reader is never invited to mistake one for the other.

Class	What it means here	Examples
VERIFIED ALGEBRA (fact)	A mathematical identity or inequality, machine-checked from stated definitions and standard real-arithmetic facts. True regardless of any economic interpretation.	SDF premium decomposition; Hansen–Jagannathan bound; Mehra–Prescott impossibility; CAPM-from-SDF; HJ–Sharpe duality.
FRAMEWORK HYPOTHESIS (claim)	A *conditional* statement — "if the world contains a latent amplifier ρ with such-and-such value, then the algebra yields premium $> 6\%$." The algebra is verified; the antecedent is **posited, not established**.	The latent-ρ "resolution"; the AR(1) $\to \rho$ calibration; the spectral amplification factor; the unified cross-domain ρ theorem.
EMPIRICAL ESTIMATE (data)	A number read off real data by a stated, reproducible procedure, used to *test* a framework hypothesis. Carries sampling uncertainty and is independent of the premium it evaluates.	The AR(1) persistence $\hat{\phi}$ of consumption growth and the implied $\hat{\rho}$ (Section 4.4).

The honest one-sentence summary. What we *prove* is that the standard model cannot generate the premium and that a latent amplifier *would* if it existed; what we *measure* is that the persistence channel's ρ, estimated directly from consumption data, falls an order of magnitude short of the value it would need — so the amplifier, if real, must live in a latent component, not in measured consumption growth.

---

1. Introduction

1.1 The puzzle

The puzzle is sharpest in the canonical consumption-based asset-pricing setup, so we state it there before anything else. Consider an agent who prices assets with a stochastic discount factor (SDF) m, so that every gross return R satisfies the Euler equation $E[mR] = 1$. For the risk-free asset

this gives $R_f = 1/E[m]$, and a short manipulation yields the central pricing identity for any risky return,

$$E[R] - R_f \;=\; -R_f \,\mathrm{Cov}(m, R).$$

Under time-separable power utility with relative risk aversion γ, the SDF is $m = \beta\,(C'/C)^{-\gamma}$, where $\beta \in (0,1)$ is the subjective one-period discount factor, and the equity premium becomes, to first order, proportional to γ times the covariance of the asset return with consumption growth. Herein lies the puzzle: aggregate U.S. consumption growth is *smooth* (standard deviation $\sigma_c \approx 0.036$ annually), so to match a ~6% premium the model needs an implausibly large γ — and a large γ in turn implies a counterfactually high risk-free rate (the companion "risk-free rate puzzle"). The model is squeezed from both sides.

1.2 What this paper adds

The economics of the impossibility above is textbook. What is new is a *result* — and the *form* in which it is established:

1. **A rank bound on the entire class of resolutions.** The headline is not another candidate mechanism but a statement *about the whole family of them*: resolving the premium requires a rank-≥ 2 structure. Long-run risk, external habit, and the cross-sectional concentration channel are all a single scalar pushed through the same monotone map $1/(1-x)$ — rank 1 — and a machine-checked no-go caps that entire class below the observed premium (§5.8). The escape strictly requires a second, independent amplification axis. This turns the usual dismissal ("your amplifier is just reparametrized long-run risk") into the theorem's point: *because* they are all rank 1, one no-go sinks all of them together. The same rank-≥ 2 verdict is delivered independently by the equity term structure, which a single priced factor cannot match (§5.7, Gormsen 2021) — theory and data agreeing on the dimension count.

2. **A certificate, not a calibration table.** We restate the impossibility as a verified inequality. There are no free numerical fudge factors hiding in a spreadsheet; the bound is derived in a kernel that rejects any step it cannot check. Zero axioms means every inequality reduces to definitions and standard real-arithmetic facts.

   A mathematical impossibility forces a choice. The framework leaves exactly three options: (1) Accept that investors are absurdly terrified of risk ($\gamma > 40$), which destroys the risk-free rate. (2) Conclude that macroeconomic data is wrong, and consumption actually fluctuates wildly. (3) Accept that a latent amplification mechanism (ρ) exists. If we reject the first two, the third is not a convenient assumption — it is the only mathematical refuge.

3. **A clearly-fenced proposal.** Many "resolutions" of the puzzle (habit formation [Campbell and Cochrane 1999], long-run risk [Bansal and Yaron 2004], rare disasters, ambiguity aversion) work by enriching the SDF until the premium fits. We add one more candidate mechanism — a latent variance amplifier ρ — but we are scrupulous about its status. We prove the *sufficiency* direction (a large enough ρ reproduces the premium) and stop there. We do **not** estimate ρ, and we do **not** claim the puzzle is "solved." The worked numbers (6.48%, 7.78%) are outputs of chosen inputs, shown to demonstrate that the channel is quantitatively live, not to assert an empirical fit.

The reason for this discipline is concrete: a sister formalization in this program overstated a framework's internal consistency as an established empirical fact, and the claim did not survive scrutiny. This paper is structured so that the verified core stands on its own even if the latent-ρ hypothesis is ultimately rejected.

Our contribution, stated plainly. The novelty is *not* a new mechanism — we add no story to the habit / long-run-risk / disaster menu. It is two things. First, a *class-level* impossibility: a single machine-checked theorem showing that the entire rank-1 family of resolutions is capped below the premium, so the escape is a question of *dimension*, not of engineering a better stochastic discount factor. This sharpens into a full *resolvability criterion*: the premium is resolvable within empirically admissible bounds if and only if the joint (product) ceiling of two independent axes clears the target (§5.8) — necessity (the no-go) and sufficiency (the construction) as one condition. And it is *representation-independent*: the same ceiling is proved directly from the Hansen–Jagannathan bound for an arbitrary stochastic discount factor, with "rank" read as the dimension of the SDF's priced-factor structure (§5.8, rank1_nogo_proof.py §8), so the result is not an artifact of the product parametrization but a property of the HJ frontier every consumption-based model respects. At its sharpest this is a spectral statement (Ky Fan): "rank" is the algebraic rank of the SDF's price-of-risk matrix and the ceiling is the sum of its r largest eigenvalues — rank-1 is the top eigenvalue, full rank is the trace (the HJ bound itself), with diminishing returns to each added dimension (§5.8, rank1_nogo_proof.py §9). Second, the *method*: a kernel-verified, Lean-exportable argument in which every claim is fenced as verified algebra, framework hypothesis, or empirical estimate. What we do **not** claim is equally explicit — we do not estimate ρ, do not assert that the latent amplifier operates, and do not treat the suggestive empirical panels (§5.1, §5.3) as load-bearing. The paper stands on the resolvability criterion (§5.8) and the term-structure discriminator (§5.7) — and on the discipline of saying exactly that.

The organizing thesis. Put in one line, the criterion reads the whole literature through a single lens: **the quantitatively successful resolutions of the equity premium did not find a better stochastic discount factor — they found a second dimension.** Habit and growth-only long-run risk are rank-1 and fall within the scope of the class no-go; Bansal–Yaron escapes because stochastic volatility is an axis independent of growth persistence, and Wachter escapes because time-varying intensity is an axis independent of disaster size. Classifying each proposed resolution by how many *independent* amplifying axes it moves (the taxonomy of §9.3) turns "which mechanism?" into the sharper and, we believe, more durable question "how many independent priced dimensions does the economy afford?" — the question this paper answers with a theorem.

1.3 Roadmap

The paper builds toward one claim — the premium needs rank ≥ 2 — and the reader can keep three load-bearing sections in view: the class no-go (§5.8, the theory leg), the term-structure discriminator (§5.7, the data leg), and the disciplined composition (§5.5 with its derived multiplicativity §5.5.1, the constructive leg). Everything else sets these up or fences them epistemically.

Section 2 sets up the SDF model and proves the impossibility (verified algebra). Section 3 introduces the latent amplifier and proves the conditional resolution (framework hypothesis). Section 4 gives the two amplification channels and their agreement point, then (Sections 4.4–4.5) confronts the persistence channel with U.S. consumption data — rejecting its naive form and finding its latent-component form admissible but weakly identified, then relocating ρ to the *cross-section* as spectral concentration (Section 4.6, a verified channel plus an industry-portfolio probe). Section 5 records

the cross-domain "unified ρ" structure (framework hypothesis) and tests its strong form against Fama–French data (Section 5.1), rejecting it; it then turns the empirical wedge into a machine-checked *no-go* (Section 5.2) and corroborates the concentration reading with a small cross-country panel of 16 economies (Section 5.3), where network concentration tracks the premium (suggestive, not load-bearing); Section 5.4 then generalizes the no-go into the full verified *impossibility frontier* of the single-ρ class; Section 5.5 shows the two measured channels compose to close the wedge with no fitted parameter, with its derived multiplicativity (§5.5.1) and a direct empirical test that the two axes are independent as that derivation assumes (§5.5.2); Section 5.6 contrasts the cross-sectional link with its (opposite-signed) time-series analog; Section 5.7 gives the paper's sharpest falsifiable prediction — a verified equity term-structure discriminator that the data resolve in favor of concentration; and Section 5.8 lifts the no-go to the entire rank-1 class (long-run risk, habit and concentration as one map), pinning the escape to rank ≥ 2; and Section 5.9 states the framework's sharpest *forward* prediction — a verified comparative static under which production-network concentration should order the equity term structure across countries. Section 6 collects the Markowitz / Hansen–Jagannathan bridge identities (verified algebra). Sections 7–8 give the formal framework and proof architecture. Section 9 discusses limitations — most importantly, exactly what would be required to promote the remaining Section 3–5 hypotheses to empirical claims.

---

2. The standard model and the impossibility *(verified algebra)*

This section is the rigorous headline. Every statement below is a machine-checked theorem in equity_premium_proof.py.

What the kernel checks, precisely. The formal proof kernel verifies relations in real arithmetic. The symbols $E[m]$, $\mathrm{Cov}(m, R)$, $\sigma(R)$, σ_c, and the premium are real variables constrained by the stated hypotheses (H_sdf_pricing, H_rf_def, H_cov_decomp); the kernel proves the algebraic identities and inequalities among them. That these variables *are* the mean, covariance, and standard deviation of genuine random variables is the standard economic interpretation we place on the symbols — it is not itself part of the formalization, because the kernel has no measure-theoretic probability layer. So "verified" here means the algebra is certified given the moment relations; it does **not** mean probability theory has been mechanized. This is the honest scope of the certificate, and it is exactly why Section 2 is a clean restatement of a known result rather than a new probabilistic theorem.

2.1 SDF pricing and the premium decomposition

We take as the modeling hypothesis the SDF pricing relation H_sdf_pricing and the risk-free definition H_rf_def. From these the premium decomposition follows as pure algebra. The kernel proves the chain (premium_distrib, premium_neg_sub, premium_cleared_base):

$$E[R] - R_f = -R_f \, \mathrm{Cov}(m, R),$$

together with the variance/covariance bookkeeping (variance_decomposition, H_cov_decomp) that the later bounds rely on. The covariance-decomposition identity is used, not assumed beyond its definition.

2.2 The Hansen–Jagannathan bound

From the Cauchy–Schwarz inequality for covariances (F_cauchy_schwarz_cov) and non-negativity of variance (F_var_nonneg), the kernel derives the Hansen–Jagannathan bound in its Sharpe-ratio form (hj_is_sharpe_bound, hj_squared_ingredient, sharpe_ratio_bound_cleared):

$$\frac{\left|E[R] - R_f\right|}{\sigma(R)} \;\le\; \frac{\sigma(m)}{E[m]}.$$

In words: the maximal attainable Sharpe ratio is bounded by the volatility of the SDF relative to its mean. Any model that wants a high premium per unit of return volatility must have a *volatile* discount factor. These are the two facts (F_var_nonneg, F_cauchy_schwarz_cov) on which the impossibility rests; both are standard.

2.3 The Mehra–Prescott impossibility

Specializing to power utility (power_utility_premium) the premium is, to leading order, $\gamma\,\mathrm{Cov}(\Delta c, R)$, which is bounded above by $\gamma\,\sigma_c\,\sigma(R)$. The Mehra–Prescott calibration fixes $\sigma_c \approx 0.036$ and caps γ at a plausible $\gamma_{\max}$. The kernel then closes the impossibility (mehra_prescott_impossibility, premium_sq_nonneg, sigma_sq_threshold, sigma_sq_exceeds_threshold):

> **Theorem (Mehra–Prescott impossibility — verified structural form).** *If* the premium equals $\gamma\,\sigma_c^2$, risk aversion is capped at $\gamma \le \gamma_{\max}$, and the calibration ceiling satisfies $\gamma_{\max}\,\sigma_c^2 <$ target, *then* the model-implied premium is strictly below target.

Two things must be said about what the kernel does and does not certify here, because the distinction is the whole point of the epistemic discipline. What mehra_prescott_impossibility verifies is exactly the conditional inequality above: the *structural* fact that a premium proportional to $\gamma\,\sigma_c^2$, with a capped γ and a small σ_c, cannot exceed a ceiling. The premise premium $= \gamma\,\sigma_c^2$ is a **modeling input** (the leading-order power-utility approximation, power_utility_premium), *not* itself machine-derived from a probability space — the kernel has no measure-theoretic layer, and we do not pretend otherwise.

The familiar number is then an assumption-explicit corollary, computed by hand from that verified structure: substituting the Mehra–Prescott calibration ($\gamma \le 10$, $\sigma_c = 0.036$) gives $\gamma_{\max}\,\sigma_c^2 = 10 \times 0.036^2 \approx 1.30\% < 6\%$. The "$< 1.3\%$" is this structural inequality evaluated at the standard numbers, not an independent empirical discovery and not a separately machine-checked decimal. The certificate's value is that the *structure* is exact and assumption-explicit — smooth consumption forces a small premium for any plausible γ — not that it uncovers something new about the data.

---

3. A latent-amplification hypothesis *(framework hypothesis — not an empirical claim)*

We now leave verified-fact territory for a *proposal*. Everything in Sections 3–5 is conditional on positing a latent amplifier; the algebra is verified, the antecedent is not.

3.1 The posited mechanism

Section 2 closed every escape within the standard model, so the minimal way forward is to widen exactly one quantity it holds fixed. Suppose the consumption variance that the *representative agent effectively prices* is larger than measured aggregate consumption variance by a factor $\rho > 1$:

$$\sigma^2_{\text{eff}} = \rho\,\sigma^2_c, \qquad \rho > 1.$$

Economically one might motivate ρ as latent state risk, mismeasurement of the relevant consumption process, or persistence that compounds short-horizon variance into the longer horizon at which the premium is earned. **We do not adjudicate among these stories, and we do not estimate ρ.** The parameter is an input.

3.2 The conditional resolution

Given the amplifier, the premium scales (latent_premium_chain, latent_resolution, latent_exceeds_standard, latent_persistence_link):

> **Theorem (conditional resolution — verified scaling).** With $\sigma^2_{\text{eff}} = \rho\,\sigma^2_c$, the amplified premium is ρ times the standard premium and is strictly larger for $\rho > 1$ (fivefold_amplification, scaling_preserves_strict_order); and if the target is at most the amplified premium, the gap is cleared (fivefold_resolves_gap, minimum_rho_criterion).

Read this carefully, because the kernel certifies less than a first reading suggests. What is verified is the *scaling structure*: amplifying the priced variance by ρ scales the premium by ρ, and the criterion for the amplified premium to reach the target is target $\leq \rho \times$ standard premium. What is **not** machine-checked is the specific decimal: substituting the calibration (standard premium $\approx 1.30\%$) and $\rho = 5$ gives $5 \times 1.30\% = 6.48\% > 6\%$ by hand arithmetic. And what is emphatically **not** proven is "$\rho = 5$" — the amplifier is an input. The 6.48% is displayed only to show the channel is quantitatively capable of clearing the gap, not to assert that markets exhibit $\rho = 5$.

The monotonicity scaffolding (scaling_preserves_strict_order, product_mono_right, minimum_rho_criterion) records the obvious comparative static: larger ρ gives a larger premium, and there is a minimum ρ above which the gap is cleared.

---

4. Two amplification channels *(framework hypothesis)*

There are two ways to generate the amplifier of Section 3. A caution on notation first: both channels are parametrized by ρ, but they turn ρ into a premium *differently* — the persistence channel scales the variance by the factor ρ directly, whereas the spectral channel produces the amplification factor $\rho/(\rho - 1)$. The two worked examples below therefore sit at different operating points ($\rho = 5$ for persistence, $\rho = 6/5$ for spectral); they are distinct mechanisms, each independently sufficient to clear the 6% gap, not a single number viewed twice. Both are verified algebra *conditional on their respective modeling choices*; neither is calibrated to data.

4.1 Persistence channel (AR(1))

Model consumption-growth shocks as an AR(1) process with persistence ϕ. The relevant object is the *cumulative impulse response* of a long-horizon claim — the sum $\sum_{k\geq 0}\phi^k = 1/(1-\phi)$ of total priced exposure accumulated over the claim's effective duration — *not* the one-period unconditional variance ratio $1/(1-\phi^2)$; for a claim spanning the infinite horizon it is the first-order cumulative sum that governs the priced amplification (cf. Bansal and Yaron 2004). The compounding of persistent shocks thus raises the effective variance the agent prices by the factor $1/(1-\phi)$; the kernel fixes this multiplier as ρ exactly ($\sigma^2_{\text{eff}}(1-\phi) = \sigma_c^2$ and $(1-\phi)\rho = 1$, so $\sigma^2_{\text{eff}} = \rho\,\sigma_c^2$ — persistence_amplification, latent_persistence_link, amplification_equals_rho). The worked example: $\phi = 0.8$ gives $\rho = 1/(1-0.8) = 5$ (persistence_phi_for_rho_5), recovering the 6.48% figure of Section 3.

4.2 Spectral channel (geometric series)

Alternatively, sum the contribution of the amplifier across horizons as a geometric series. The kernel proves the geometric-sum identities for $N = 1, 2, 3$ (geometric_sum_1, geometric_sum_2, geometric_sum_3), their positivity and monotonicity (partial_sum_positive, partial_sum_monotone, partial_sum_lt_limit), and the closed-form limit (geometric_limit_identity, spectral_sum_latent). The spectral amplification factor is

$$A(\rho) = \frac{\rho}{\rho - 1},$$

so that $\rho = 6/5$ gives $A = 6$ and a premium of 7.78% (spectral_premium_formula, amplification_factor_rho_6_5, spectral_premium_exceeds_standard, spectral_full_resolution, mild_persistence_for_rho_6_5).

4.3 The channels agree

The two amplification maps are not unrelated: the kernel proves they coincide at $\rho = 2$ — where the direct factor ρ and the spectral factor $\rho/(\rho-1)$ are both equal to 2 (models_agree_at_rho_2) — and orders them away from that point (spectral_exceeds_persistence, persistence_exceeds_spectral, product_var_exceeds_sum, product_variance_independent, multiperiod_linear_lower_bound). This internal coherence is a point in the hypothesis's favor — but it is coherence of the *framework*, not evidence about the *world*.

4.4 Confronting the persistence channel with data *(empirical estimate)*

The honesty boundary cuts both ways: having proven the conditional "$\rho = r \Rightarrow$ premium," we are obliged to ask what ρ the data actually delivers through the persistence channel — *without* reading it off the premium (the non-circularity requirement of Section 9.2). The persistence channel makes this falsifiable, because it pins ρ to a directly estimable quantity: the AR(1) persistence ϕ of consumption growth, via $\rho = 1/(1-\phi)$.

We estimate ϕ by OLS on real per-capita U.S. consumption growth (FRED series A794RX0Q048SBEA, 1947Q1–2026Q1; the estimation is reproducible via empirical_rho.py in the proof directory):

Series	$\hat{\phi}$ (OLS SE)	$\hat{\rho} = 1/(1-\hat{\phi})$	95% CI on $\hat{\rho}$	premium at $\hat{\rho}$
Quarterly growth	-0.07 (0.06)	0.93	[0.84, 1.04]	1.2%

Series	$\hat{\phi}$ (OLS SE)	$\hat{\rho} = 1/(1-\hat{\phi})$	95% CI on $\hat{\rho}$	premium at $\hat{\rho}$
Annual growth (YoY)	0.05 (0.11)	1.05	[0.85, 1.38]	1.4%

Both use the Mehra–Prescott calibration ($\gamma = 10$, $\sigma_c = 0.036$) so the premium formula matches the kernel's worked example exactly; the data's own post-war σ_c is *smaller* (≈ 0.02 annually), which only widens the gap. Clearing the 6% target requires $\rho^* \approx 4.63$, i.e. $\phi \gtrsim 0.78$. The 95% confidence intervals are delta-method intervals on $\hat{\rho}$ from the OLS standard error of $\hat{\phi}$.

The verdict is a clean face-value rejection — with the uncertainty to back it. Measured directly, the persistence of aggregate consumption growth ($\hat{\phi} \approx 0.05$) is more than an order of magnitude below the $\phi \approx 0.78$ the channel needs. The implied $\hat{\rho} \approx 1$ barely moves the premium off the 1.3% impossibility baseline, and the *entire* 95% interval ($\hat{\rho} \leq 1.38$) lies far below the required 4.63 — this is not a borderline miss that a wider error bar could rescue. So the naive persistence channel, taken to data, **does not** resolve the puzzle.

This negative result is informative, not fatal, and it sharpens the hypothesis. It says: if a latent amplifier operates, it cannot be the measured AR(1) of total consumption growth. It must instead live in a *small but highly persistent latent component* of consumption — precisely the object the long-run-risk literature [Bansal and Yaron 2004] posits (a predictable component with ϕ near unity but a tiny variance share, invisible to a naive AR(1) on the aggregate). The next subsection estimates exactly that object.

4.5 The latent-component reading: admissible but weakly identified *(empirical estimate)*

The non-naive estimator fits an unobserved-components model — the Bansal–Yaron structure — in which consumption growth is a small persistent latent state plus transitory noise:

$$\Delta c_t = \mu + x_t + \varepsilon_t, \qquad x_t = \phi\, x_{t-1} + u_t,$$

and reads $\rho = 1/(1-\phi)$ off the *latent* persistence ϕ, recovered by Kalman-filter maximum likelihood (empirical_rho.py). Three estimators, applied to the same quarterly series, tell a deliberately honest, mixed story:

Estimator	$\hat{\phi}$	$\hat{\rho}$	reading
Free state-space MLE	-0.46	< 1	transitory dynamics dominate the likelihood
Persistence-constrained MLE ($\phi \geq 0.5$)	0.61	2.6	costs only $\Delta \log L \approx 1.7$ vs. the free fit
ARMA(1,1) long-lag ACF decay	0.77	4.4	approaches the required $\rho^* \approx 4.6$

The standard-errors story here is the whole point. A 95% profile-likelihood interval for the latent persistence spans essentially the entire stationary range, $\hat{\phi} \in [-0.95,\ 0.98]$ — equivalently $\hat{\rho}$ from 0.5

to effectively unbounded. Unlike the naive estimate, whose tight interval *excludes* the gap-clearing value, the latent estimate's interval *contains* it: the gap-clearing $\rho^* \approx 4.6$ cannot be rejected, but neither can $\rho \approx 1$. (A Hessian-based Wald standard error would be misleading under such weak identification, which is why we report the profile-likelihood interval.)

The honest verdict is "admissible but unconfirmed," not "resolved." A free likelihood prefers transitory mean-reversion (the quarterly series has slightly *negative* first-order autocorrelation), so the data do not *volunteer* a persistent component. But they do not *reject* one either: imposing the persistence the channel needs costs under two log-likelihood units, the geometric decay of the long-lag autocorrelations independently points to $\phi \approx 0.77$, and the confidence interval is too wide to exclude anything. This weak identification is itself a documented feature of the long-run-risk program [Beeler and Campbell 2012]: the persistent component is economically decisive yet statistically faint.

So the latent-ρ hypothesis lands precisely where honesty demands: the naive channel is *rejected*, the latent-component channel is *admissible* at the edge of the data's tolerance, and confirmation would require an out-of-sample test rather than an in-sample fit. We do not claim the puzzle is resolved; we claim the mechanism is empirically *live but unproven*, and we have made the gap between "live" and "proven" measurable.

Variance-weighting the channel: the honest effective amplification. The caveat that we report *persistence, not premium* can be turned into a number, and doing so is the sharpest empirical result of this section. The single-component reading $\rho = 1/(1-\phi)$ amplifies *all* of consumption variance, but only the persistent component — variance share s — is amplified by the geometric map $1/(1-\phi)$; the transitory share $1-s$ is not. The variance-weighted effective amplification entering the power-utility premium $\pi = \gamma\,\sigma_c^2\,\rho$ is therefore

$$\rho_{\text{eff}} \;=\; (1-s)\cdot 1 \;+\; s\cdot\frac{1}{1-\phi} \;=\; 1 + s\,\frac{\phi}{1-\phi},$$

the same amplification map the paper applies elsewhere, now weighted by how much variance the persistent component actually carries. Estimated on the data (empirical_rho.py, effective_rho) the verdict is decisive: at the free MLE ($\phi = -0.46$, $s = 20.6\%$) $\rho_{\text{eff}} = 0.94$; and at the persistence-constrained fit the likelihood drives the persistent innovation variance to zero ($s \approx 0$) to impose $\phi \geq 0.5$ at minimal cost, so $\rho_{\text{eff}} \approx 1.00$ — against the single-component 2.6. In words: *once you weight by how much variance the persistent component actually carries, the consumption-persistence channel contributes essentially no amplification under power utility*, and the single-component $\rho = 2.6$ overstates by exactly the variance-share factor. This does not sink long-run risk; it locates where its amplification must come from — not the *variance* of the persistent component (tiny) but its *price*, which recursive (Epstein–Zin) preferences raise far above its variance share. Under power utility the time-series axis sits at $\rho_{\text{eff}} \approx 1$, so the amplification the premium demands ($\rho^* \approx 4.6$) must be supplied by a *second* channel — the recursive-preference pricing of long-run risk, or the cross-sectional concentration axis of §4.6. This is the rank-≥ 2 verdict read directly off consumption data: the time dimension alone, honestly measured, is neutral.

Recursive preferences, computed — the rescue is a number, not an assertion. The claim that recursive (Epstein–Zin) preferences "raise the price far above the variance share" can be made quantitative and, indeed, machine-checked as a comparative static (multiperiod_sdf_proof.py §7, 5 theorems, 0 axioms; empirical_rho.py, epstein_zin_amplification). In the Bansal–Yaron (2004) long-run-risk model the SDF prices two shocks — the transitory consumption shock (price $\lambda_\eta = \gamma$) and the persistent long-run-risk shock (price $\lambda_e = (\gamma - 1/\psi)\,\kappa_1/(1-\kappa_1\phi) = \text{wedge} \times M$),

where the *wedge* $\gamma - 1/\psi$ is the preference non-separability and $M = \kappa_1/(1-\kappa_1\phi)$ is the present value of the persistent stream. Under **power utility** $\gamma = 1/\psi$, so the wedge is zero and $\lambda_e = 0$: the persistent axis carries no recursive price — the algebraic root of the $\rho_{\text{eff}} \approx 1$ above (recursive_price_zero_under_power_utility). Under **Epstein–Zin** the wedge is positive, so $\lambda_e > 0$ and is *increasing and unbounded* in persistence (recursive_price_positive_under_ez, recursive_multiplier_increasing_in_persistence, recursive_multiplier_diverges_in_persistence). In the Hansen–Jagannathan / Ky Fan S^2 framing (§5.8) the two shocks are two orthogonal price-of-risk axes with variances $a_1 = (\lambda_\eta\sigma_\eta)^2$ and $a_2 = (\lambda_e\sigma_e)^2$; their strength ratio is $k = (1 - 1/(\gamma\psi))\, M\, \phi_e$, so the second axis is $a_2 = k^2 a_1$. At the standard BY calibration ($\gamma = 10$, $\psi = 1.5$, $\phi = 0.979$, $\kappa_1 = 0.997$, $\phi_e = 0.044$) this gives $k \approx 1.71$, so $a_2 \approx 2.9\, a_1$ and the total price-of-risk variance $S^2 = a_1 + a_2$ is inflated by $\approx 3.9\times$ over power utility — where the same number is exactly 1 (the axis is absent). The persistence axis, inert under power utility, becomes under recursive preferences a *genuine second eigenvalue* of the order the rank-1 no-go (§5.8) demands: recursive_second_axis_dominates_first shows $a_2 \geq k^2 a_1$, the constructive complement to the class no-go — the second dimension the puzzle requires is supplied not by the *variance* of the persistent component but by its recursively-set *price*.

4.6 A cross-sectional reading of ρ: spectral concentration *(verified structure + empirical probe)*

Sections 4.4–4.5 read ρ off the *time-series* persistence of aggregate consumption, where the signal is weak. But the framework defines ρ as a *spectral* quantity, $\rho = \lambda_1/\lambda_2$, and there is a second, economically distinct place that ratio can live: the *cross-section*. If a few dominant modes of the production/return network carry the system — the granularity mechanism of [Gabaix 2011] in spectral form — then ρ is the top-to-second eigenvalue gap of the network, and the amplification comes from cross-sectional concentration rather than temporal persistence. This is not the same mechanism restated; it is a different economic claim, testable on different (cross-sectional) data, and it does not require aggregate consumption to be persistent.

The verified structure. A separate zero-axiom proof file (spectral_concentration_proof.py, 7 theorems) formalizes the channel: a spectral gap $\rho > 1$ makes the top mode strictly dominate (gap_gt_one_top_dominates), amplification is monotone in the gap (gap_amplification_monotone), two channels compose multiplicatively so a weak persistence channel times a strong concentration channel exceeds what either delivers alone (combined_amplification_exceeds_each, concentration_rescues_weak_persistence), and — the honest empirical shape — the persistence and concentration readings *bracket* the premium's demand (premium_bracketed_by_channels, cross_sectional_channel_theorem).

The proxy, and its correction. A first probe measures the gap from the correlation matrix of 48 Fama–French value-weighted industry portfolios (682 complete-case months; spectral_concentration.py). The eigenvalue structure is exactly the random-matrix picture of [Laloux, Cizeau, Bouchaud and Potters 1999]: the top eigenvalue ($\lambda_1 = 27.8$, carrying 58% of cross-sectional variance) is the market mode, well separated from a noise bulk; a participation ratio of 2.9 out of 48 confirms strong concentration; and, as a bonus cross-check of Section 5's anomaly leg, exactly **three** eigenvalues clear the Marchenko–Pastur noise edge — a purely spectral count that reproduces the Fama–French three-factor structure with no factor-fishing. But the raw return-gap $\rho \approx 12$ *overshoots* badly, because return correlations conflate the market factor's statistical dominance with genuine production-network concentration. The return matrix is a proxy; the real object is the production network.

The real instrument. We therefore measure ρ from the actual production network: the BEA Commodity-by-Industry Direct Requirements matrix A (technical coefficients, ~71 sectors, 1997–2023; bea_network_rho.py). Because A is nonnegative, Perron–Frobenius gives a real dominant root $\lambda_1 \approx 0.48$ (strikingly stable across 27 years). Two spectral readings follow, both robust:

- the **concentration gap** $\rho_{\text{gap}} = \lambda_1/\lambda_2 \approx 1.43$ (range $[1.27, 1.59]$ over 1997–2023);
- the **Leontief multiplier** $\rho_{\text{leon}} = 1/(1-\lambda_1) \approx 1.94$ — the cross-sectional analogue of the persistence law $1/(1-\phi)$, with the network Perron root λ_1 playing the role of the time-persistence ϕ. The *same* $1/(1-x)$ law reappears with a spectral x; this is the spectral framing paying off, connecting the network-amplification literature [Acemoglu et al. 2012] to the premium's ρ.

The verdict — a partial agreement that sharpens the puzzle. The readings now order cleanly, and two independent cross-sectional measurements *agree*:

Reading of ρ	value	note
Time-series persistence (§4.4)	≈ 1.0	too little amplification
Production-network gap (BEA, real)	≈ 1.4	robust, 27-year-stable
Factor anomalies (FF5, §5.1)	≈ 1.5	*matches the network gap*
Leontief network multiplier	≈ 1.9	$1/(1-\lambda_1)$ law
Premium's demand ρ^*	≈ 4.6	the outlier
(return-correlation proxy)	≈ 12	artifact, corrected above

Three honest findings. First, the real instrument *corrects* the proxy overshoot ($12 \to 1.4$) — a self-correction the method forces. Second, and more important, the production-network gap (≈ 1.4) and the factor-anomaly ρ (≈ 1.5) — two entirely independent measurements, one from input–output accounting, one from equity-factor Sharpe ratios — land at essentially the *same* value. That is the paper's first genuine cross-domain *consistency*: the anomaly structure and the production network really do share a spectral $\rho \approx 1.5$. Third, the equity premium is the lone outlier: it demands $\rho^* \approx 4.6$, roughly three times the amplification the economy's measured spectral concentration actually exhibits.

This last gap is not a failure of the framework; it is the equity-premium puzzle *restated in spectral terms and sharpened*. The classical puzzle says standard preferences cannot generate the premium; the spectral restatement says the premium demands more cross-sectional concentration than the real production network — or the factor structure — contains. What lever E buys is therefore substantive: it relocates ρ from a weakly-identified time-series parameter to a robustly-measured cross-sectional one (so the mechanism is a granularity story, not a long-run-risk restatement), it produces a real cross-domain agreement between the network and the anomalies, and it pins down exactly where the residual puzzle lives — in the wedge between the premium's demand and the economy's spectral concentration.

---

5. The unified ρ structure *(framework hypothesis)*

The capstone file (unified_rho_capstone_proof.py) records a structural claim of the framework: a *single* latent number ρ simultaneously controls the equity premium, a financial-contagion fragility

threshold, and the number of surviving factor anomalies.

The kernel proves the internal links (rho_determines_premium_and_threshold, premium_threshold_product_co high_premium_implies_fragility, contagion_exceeds_premium_amplification, higher_rho_fewer_factors, rho2_three_predictions), an impossibility-triangle statement (premium_threshold_impossibility, deconcentration_resolves_dilemma), and the grand statement (unified_latent_economics_theorem):

> **Theorem (unified latent economics, framework-internal).** A single $\rho > 1$ implies a positive equity premium, a finite contagion threshold, and a positive CAPM residual (an alpha relative to the market), jointly.

We flag this as the most speculative part of the paper. It is mathematically verified *as a set of implications inside the framework's own definitions*. Whether the *same* ρ governs premium, contagion, and anomalies in reality is a strong empirical conjecture — and, unlike a verified theorem, one that can be killed by data. The next subsection tries to kill it.

5.1 An out-of-sample test of the shared-ρ claim *(empirical estimate)*

The unified claim is falsifiable in the most direct way: each domain pins ρ through a different functional form, so the ρ *implied* by each should agree if a single number governs all three.

- **Premium leg.** The amplification the observed premium demands over the Mehra–Prescott baseline is $\rho_{\text{EP}} = 0.06/(\gamma\sigma_c^2) \approx 4.6$.
- **Anomaly leg.** The capstone sets the residual left by a one-factor (market) model to $1/(\rho^2-1)$, so $\rho_{\text{anom}} = \sqrt{1 + 1/\text{residual}}$. We measure the residual as the share of the maximum squared Sharpe ratio *not* spanned by the market factor alone, using real Fama–French factors (Ken French data library, 754 monthly observations, 1963–2026; crossdomain_rho.py). The market reaches a squared Sharpe of 0.018; the five-factor tangency reaches 0.078, leaving a residual of 0.77 and $\rho_{\text{anom}} \approx 1.5$ (the three-factor cut gives 1.7). An IID bootstrap (1000 resamples) puts a tight 95% interval on this estimate: $\rho_{\text{anom}} \in [1.43,\ 1.67]$.
- **Contagion leg.** The threshold scales as $1/\rho$, but recovering an absolute ρ_{cont} requires the proportionality constant $\alpha/\beta_{\text{base}}$, which public data do not fix; we therefore cannot estimate this leg independently and do not pretend to.

Leg	implied ρ	source
Premium	≈ 4.6	consumption / observed premium
Anomalies (FF5)	≈ 1.5	Fama–French factors (real)
Anomalies (FF3)	≈ 1.7	Fama–French factors (real)
Contagion	—	not identified without a constant

Verdict: the strong shared-ρ claim is not supported. The premium-implied $\rho \approx 4.6$ lies far outside the anomaly leg's 95% bootstrap interval [1.43, 1.67] — the two estimates differ by a factor of three, and the gap is statistically decisive, not a wide-error-bar coincidence. The in-sample tangency Sharpe is upward biased, which pushes ρ_{anom} *toward* $\sqrt{2} \approx 1.41$ — so the gap is if anything understated out of sample, not an artifact. What survives is the *weak* claim, which is exactly what the kernel actually proves: each domain is governed by *some* $\rho > 1$, and the *signs* of the cross-domain tradeoffs (higher premium lower threshold fewer factors) hold. The claim that one *numerical* ρ is shared across domains is, on this test, rejected.

This is the paper practicing what it preaches. Section 5's grand theorem is real as a set of framework-internal implications; its boldest empirical reading does not survive contact with data; and we report that plainly rather than letting the verified-theorem status of the algebra borrow credibility for the conjecture.

5.2 A verified no-go: the concentration ceiling cannot clear the premium *(verified algebra)*

The measurements of Sections 4.6 and 5.1 leave a sharp, repeated pattern: the economy's cross-sectional spectral concentration is small and agreed upon ($\rho_{\text{network}} \approx 1.4$, $\rho_{\text{anomaly}} \approx 1.5$), while the premium demands $\rho^* \approx 4.6$. We now turn that *empirical* wedge into a *verified* impossibility (premium_nogo_proof.py, 6 theorems, 0 axioms).

Write the premium as premium $= \gamma\sigma^2\rho$ (the latent-premium chain of Section 3) and let $\rho_{\max}$ be a concentration ceiling — the largest ρ the real production network supplies. The kernel proves:

- **concentration_ceiling_caps_premium** — if $\rho \le \rho_{\max}$ and the ceiling is itself too low to clear the target ($\gamma\sigma^2\rho_{\max} <$ target), then *every* admissible ρ yields premium $<$ target.
- **single_rho_infeasible** — a single ρ cannot both respect the ceiling ($\rho \le \rho_{\max}$) and meet the premium floor ($\rho^* \le \rho$) when $\rho_{\max} < \rho^*$; the constraints are contradictory.
- **no_single_rho_resolves_premium** — the grand form: under a binding concentration ceiling, the premium falls short by a strictly positive wedge.

This is the Mehra–Prescott move one level up. MP proved that standard CRRA preferences *cannot* generate the observed premium at plausible risk aversion; here we prove that a single spectral ρ *bounded by the economy's measured concentration* cannot either — and, unlike a numerical calibration exercise, the claim is machine-checked, so "no model in this one-parameter class clears the premium under the measured ceiling" is airtight rather than suggestive.

The scope is stated honestly. The no-go is conditional on the empirical premise $\gamma\sigma^2\rho_{\max} <$ target, which the Section 4.6 measurement supplies ($\rho_{\max} \approx 1.4$–$1.5 \ll 4.6$). It does *not* say the premium has no resolution; it says the resolution *cannot* be a single concentration-bounded ρ. The escape hatch is explicit and already formalized: the composition theorems of Section 4.6 (combined_amplification_exceeds_each) show that *composed* channels — a modest persistence times a moderated concentration, or a higher effective risk price — can exceed what any single ρ delivers. The no-go thus does real work: it rules out the simplest reading and names precisely what a genuine resolution must do (break the single-ρ premise), rather than leaving the matter to taste.

5.3 A cross-country test: does network concentration track the premium? *(empirical estimate)*

Section 4.6's central positive result — that the U.S. production network's spectral concentration ($\rho_{\text{network}} \approx 1.4$) nearly equals the independently estimated factor-anomaly $\rho \approx 1.5$ — rests on a single country. One matched pair is a coincidence until it is a pattern. The mechanism makes a sharp cross-country prediction: because a single latent factor amplifies whatever dominant mode the production network already carries, economies with *more concentrated* networks (a larger spectral gap) should carry *larger* equity premia.

We test this on 16 advanced economies (crosscountry_rho.py). For each country we take its domestic technical-coefficient matrix from the OECD Inter-Country Input–Output tables (2023 release, averaged over 2016–2019 to avoid the COVID year) and compute the *same* two spectral

readings as the U.S. instrument (reusing bea_network_rho.network_rho): the concentration gap $\rho_{\text{gap}} = |\lambda_1|/|\lambda_2|$ and the Leontief multiplier $\rho_{\text{leon}} = 1/(1-\lambda_1)$. The realized premium is the mean excess equity return over the short rate from the Jordà–Schularick–Taylor Macrohistory Database (R6), taken over the post-1950 sample — the full 1870– sample is dominated by 1920s hyperinflation and wartime nominal-return artifacts (Germany's full-sample "premium" of 22% is not a premium signal), so post-1950 is the clean measure, with the full sample reported for robustness.

Network reading	vs. post-1950 premium	vs. full-sample premium
Concentration gap ρ_{gap}	$r = +0.52$, slope $+0.058$ ($t = 2.3$, $p = 0.04$), Spearman $+0.51$	$r = +0.28$ (n.s.)
Leontief mult. ρ_{leon}	$r = -0.37$ (n.s.)	$r = -0.17$ (n.s.)
Perron root λ_1	$r = -0.37$ (n.s.)	$r = -0.15$ (n.s.)

Two things stand out, and both are what the concentration reading predicts — and what a generic "more connected economies earn more" story does not.

The sign and significance land on the gap. Across the modern cross-section, a country's spectral concentration ρ_{gap} is positively and significantly associated with its equity premium ($r = 0.52$, $p = 0.04$; the rank correlation Spearman 0.51 confirms the relation is not a single-outlier artifact). Moving from the least concentrated network in the sample (France, $\rho_{\text{gap}} \approx 1.2$) to the most (Sweden/Denmark, $\rho_{\text{gap}} \approx 1.8$) is associated with roughly 3.5 percentage points more premium — economically large, and in the predicted direction.

The discrimination is informative. It is *concentration* (the dominance of the top mode, ρ_{gap}) that tracks the premium, not overall propagation (the Leontief multiplier ρ_{leon} or the Perron root λ_1), whose correlations are if anything mildly negative. A story in which the premium simply reflects how tightly connected an economy is would predict the opposite ranking. The spectral-gap reading survives this cut; a generic-connectivity reading does not.

One tension we do not paper over. The amplification factor the mechanism *derives* from the geometric-sum propagation (§4.6, §5.7) is the Leontief multiplier $\rho_{\text{leon}} = 1/(1-\lambda_1)$ — and that is precisely the reading whose cross-country correlation is null-to-negative. So the panel corroborates the *concentration reading* (a dominant mode that stands out from the rest co-moves with the premium) but does **not** validate — and is in mild tension with — the specific $1/(1-\lambda_1)$ *functional form* of the amplification law. We therefore read the cross-section for *which structural feature* co-moves with the premium (concentration, not raw propagation), not as a test of the amplification law's functional form; that form is a modeling posit whose only sharp test in this paper is the term structure (§5.7). Using ρ_{gap} as the operative concentration statistic is a choice the cross-section supports empirically but the geometric-sum derivation does not by itself single out — a gap between the derived object and the measured one that we flag rather than hide.

Small-n robustness: robust sign, fragile 5%. With $n = 16$ the natural worry is that the $p = 0.04$ rides on one country and on the normal approximation behind the t-statistic. Two distribution-free checks (crosscountry_rho.py: loo_robustness, permutation_test) locate exactly how much weight the result can bear. *Leave-one-out:* refitting on all 16 drop-one subsamples leaves the slope uniformly positive and in a tight band ($[+0.048,\ +0.070]$) — the *direction* is not an artifact of any single country — but 5% significance survives only 11 of the 16 drops, and

removing the most influential country (Denmark) lifts the p-value to 0.118. *Permutation:* the exact null from 10^5 random relabelings gives $p = 0.039$ (Pearson) and $p = 0.044$ (Spearman), so the association *is* significant without leaning on the t-approximation. The honest reading is a **robust positive association whose 5%-level significance leans partly on one country** — precisely the "suggestive, not decisive" status we assign it, now quantified rather than asserted.

Verdict: a genuine but modest cross-country regularity, honestly bounded. The relation holds in the window where a recent network snapshot is a defensible proxy for structure (post-1950) and fades into noise over the 150-year sample — consistent with, but not proof of, the claim that concentration is what the premium reflects. With $n = 16$ the estimate is suggestive rather than decisive, and it does not close the $\rho^* \approx 4.6$ wedge: the cross-country slope concerns the *level* of concentration, not its *sufficiency*. What it does is upgrade Section 4.6's single matched pair into a signed, discriminating cross-country pattern — the one direction in which the concentration reading could most easily have died, and did not.

We deliberately do *not* rest the empirical case on this panel. With $n = 16$ and a window-sensitive relation, it is corroboration, not proof; the paper's load-bearing empirical anchor is the term-structure discriminator (§5.7), which is sharper (a magnitude reproduced by an independently measured number, not a small-sample correlation) and which the level of the premium cannot fake. The cross-country panel's role is narrow and specific: it shows the concentration–premium link is a pattern rather than a single coincidence, and — through the gap-versus-propagation contrast — that the pattern is the one the mechanism predicts. Nothing downstream depends on it.

5.4 The impossibility frontier: the full feasible set, machine-verified *(verified structure)*

Section 5.2's no-go is a single verdict: the target premium is out of reach. But the same locked algebra says more than "one point is unreachable" — it pins down the *entire* set of outcomes the single-ρ class can produce. Machine-checking that set turns the binary no-go into a complete Pareto-type boundary (impossibility_frontier_proof.py, 9 theorems, 0 axioms).

The framework locks three observables to one number: the premium $= \gamma\sigma^2\rho$ (rises with ρ), the contagion threshold with $\rho \cdot D_c = k$ (falls with ρ), and the CAPM residual with $(\rho^2 - 1)\,\mathrm{res} = 1$ (falls with ρ). Three verified consequences follow:

- **Frontier locus and Pareto boundary.** Achievable $(\text{premium}, D_c)$ pairs lie exactly on the hyperbola $\text{premium} \cdot D_c = \gamma\sigma^2 k$ (frontier_locus); the region above it — both high premium *and* high resilience — is provably empty (pareto_boundary_empty, above_frontier_infeasible). This is the no-go promoted from a single premium threshold to a two-dimensional boundary. The frontier slopes strictly downward (frontier_downward_sloping): along it, premium is bought with fragility.
- **Truncated endpoints.** Under the measured ceiling $\rho \le \rho_{\max}$ the arc is bounded: $\text{premium} \le \gamma\sigma^2\rho_{\max}$ (frontier_capped_above) and $D_c \ge k/\rho_{\max}$ (frontier_floored_below). The observed premium sits *beyond* the arc's endpoint (target_beyond_frontier_endpoint) — the §5.2 no-go, reread as "past the end of the frontier."
- **Dimensional collapse.** Because all three observables are functions of the one ρ, they trace a one-parameter curve, not a three-dimensional volume: raising ρ raises the premium and lowers the residual *jointly* (premium_rises_residual_falls). The class cannot set premium, fragility, and anomaly count independently — measuring any one pins the trade-off with the rest.

The grand statement (impossibility_frontier_theorem) collects these: for any admissible $\rho \in (1, \rho_{\max}]$, the achievable premium is positive, lies on the frontier, is capped by the ceiling endpoint, and falls strictly short of a target beyond it.

Why this is more than the no-go: a calibration exercise can only report "I did not find a ρ that works." A verified frontier says something stronger and permanent — *the entire achievable set is this bounded arc, and here is the boundary that cannot be crossed.* It converts an absence of counterexamples into a proof that none exist, and it names the only escape (move the whole hyperbola outward by changing the structural constants $\gamma\sigma^2$ or k, not ρ) rather than leaving it to intuition.

5.5 A disciplined composition: how far do the measured channels get? *(empirical estimate)*

The no-go (§5.2) and the frontier (§5.4) name the same escape: a single ρ cannot clear the premium, but *composed* channels can (the verified combined_amplification_exceeds_each). That is an existence statement about the algebra. The sharper empirical question is whether the two channels we have *actually measured* — cross-sectional concentration (§4.6) and time-series persistence (§4.5) — compose to the required amplification *without any fitted parameter*. Because the two live on independent axes (one temporal, one cross-sectional), the framework composes them multiplicatively: total $= \rho_{\text{persist}} \cdot \rho_{\text{conc}}$. Both inputs are measured elsewhere in this paper and the premium fixes the target $\rho^* = 0.06/(\gamma\sigma^2) \approx 4.6$; the test (empirical_rho.py::composed_resolution) asks whether the product reaches it.

Concentration channel (BEA)	required residual ρ_{persist}	inside admissible band?	composed range	reaches $\rho^* \approx 4.6$?
spectral gap $\lambda_1/\lambda_2 \approx 1.4$	≈ 3.3	yes ($[2.6, 4.4]$)	[3.6, 6.1]	upper half of band
Leontief $1/(1-\lambda_1) \approx 1.9$	≈ 2.4	just below the band	[4.9, 8.3]	across the whole band

The persistence band $[2.6, 4.4]$ is the latent estimator's two point readings (persistent-fit 2.6, long-lag ACF 4.4); its profile-likelihood interval is unbounded, so these are admissible point estimates, not a tight CI. Note that these are *single-component* persistence readings: they are the amplification the persistent component would supply *if priced above its variance share* (the recursive-preference reading), not its variance-weighted power-utility contribution, which §4.5 shows is $\rho_{\text{eff}} \approx 1$. The composed ranges in the table are therefore the recursive-preference operating points; qualification #2 below reads the same table under power utility.

Verdict: the composition substantially closes the wedge — sufficiently, not confirmedly. The concentration channel cuts the burden on the weakly-identified persistence channel from the full $\rho^* \approx 4.6$ down to a residual of 2.4–3.3. The gap reading needs persistence ≈ 3.3, which sits *inside* the estimator's admissible band $[2.6, 4.4]$; the Leontief reading needs only ≈ 2.4, which falls just *below* the band — i.e. the composed mechanism there more than clears the target (it slightly overshoots) rather than landing inside it. Under the Leontief reading the composed point estimates already clear 6% at the low end of the band; under the gap reading they bracket it. Either way, no parameter was fitted to the premium: the target is reached (or bracketed) by two independent measurements.

Three qualifications keep this from being oversold. First, the persistence channel is *weakly identified* — its ρ is a wide admissible range, not a number — so "inside the band" means *not excluded*, not *confirmed*. Second — and §4.5 now makes this a number, not a hedge — the single-component ρ overstates the persistence channel: variance-weighted, its effective amplification is $\rho_{\text{eff}} \approx 1$ under power utility, because the persistent component carries a near-zero variance share. Read strictly under the power-utility SDF the kernel formula assumes, the persistence leg of this composition is therefore ≈ 1, and the concentration axis alone (1.4–1.9) does *not* reach $\rho^* \approx 4.6$: the table clears only if the persistent component is priced *above* its variance share, as recursive (Epstein–Zin) preferences do. So this composition is a statement about a *recursive-preference* economy — long-run risk priced beyond its variance — not about power utility. The rank-≥ 2 requirement is untouched either way, and in fact sharpened: under power utility the time axis contributes nothing, so a second independent axis is not merely helpful but mandatory. Third — and this is the one assumption we can now discharge — multiplicativity is not a fitted convenience but a *derived* consequence of the two channels' orthogonality (§5.5.1). Within those limits, this is the constructive complement to the no-go: the single-ρ door is provably shut (§5.2), and the two-channel door is empirically open — the measured concentration is just enough to bring the required persistence back inside what the data admit.

5.5.1 Multiplicativity is derived, not assumed

The composition above multiplied the two channels rather than adding them. That step carries weight — it is what lets two sub-target amplifiers clear a target neither reaches alone — so it should not be a free modeling choice. It is not. Multiplicativity is *forced* by the fact that space (the production network, across sectors at one instant) and time (persistence, across dates) are the economy's two aggregation axes, and they are orthogonal. A shock that propagates separably along both axes has, in the cell indexed by space-round i and time-lag j, an impulse response $\lambda^i \varphi^j$ — a *product*, precisely because the spatial spread and the temporal decay act on independent indices. The total cumulative amplification is then the double geometric sum, and a separable summand makes it factorize:

$$\sum_{i\geq 0}\sum_{j\geq 0} \lambda^i \varphi^j = \Big(\sum_{i\geq 0} \lambda^i\Big)\Big(\sum_{j\geq 0} \varphi^j\Big) = \frac{1}{1-\lambda}\cdot\frac{1}{1-\varphi} = a_{\text{space}} \cdot a_{\text{time}}.$$

Six theorems in composition_proof.py (0 axioms) make this precise. The concrete mechanism is block_response_factorizes: for a separable response the summed 2×2 block of (round, lag) cells equals *exactly* the product of the marginal sums, $(a_0 + a_1)(b_0 + b_1) = \sum_{i,j} a_i b_j$. At the level of the closed form, product_is_joint_amplification shows the product $a_{\text{space}}\, a_{\text{time}}$ satisfies the *joint* two-axis geometric equation $(a_{\text{space}} a_{\text{time}})\,(1-\lambda)(1-\varphi) = 1$, and amplification_unique shows the joint amplification is the unique solution of that equation — so the two-axis amplification *must* equal the product. Finally multiplicative_dominates_additive records why the distinction matters: for amplifiers ≥ 1 the product dominates the naive additive composition ($a_{\text{space}} a_{\text{time}} \geq a_{\text{space}} + a_{\text{time}} - 1$), which is exactly what lets the operating point $1.9 \times 2.4 \approx 4.6$ (measured Leontief concentration $\times$ the residual persistence it then requires) clear a target that $1.9 + 2.4 - 1 = 3.3$ would miss. To be precise about what is assumed versus derived: the *separability* of the joint impulse response — that the (round i, lag j) cell is the product $\lambda^i \varphi^j$ — is the modeling assumption, and it is the natural one because the spatial and temporal indices are distinct. *Given* separability, multiplicativity is not a further assumption but a theorem: the double sum factorizes, and the kernel verifies that factorization (block_response_factorizes) together with uniqueness and dominance. So the

escape in §5.5 is multiplicative by the *structure* of a separable shock — not by a separately fitted convenience — with separability the one economic premise doing the work.

5.5.2 The two axes are empirically orthogonal *(empirical estimate)*

The multiplicativity in §5.5.1 — and, downstream, the two-axis necessity of the axis-dimension theorem (§5.8) — is *forced* by one premise: that the space axis (cross-sectional network concentration) and the time axis (persistence of the cash-flow process) are **independent**. That premise is not innocuous. A natural objection is that concentrated production networks might also propagate shocks more persistently, so that a single latent force drives both axes; if so, the composition would double-count one channel and the true mechanism rank would sit strictly between 1 and 2. Because everything constructive rests on this, we test it directly rather than assert it.

We use the two instruments the paper already deploys, on 18 economies (orthogonality_test.py) — two more than the $n = 16$ cross-country premium test of §5.3, because this orthogonality test needs only consumption/GDP-growth persistence and network concentration, not equity-premium series; the two economies lacking a reliable long-run premium series in the JST database drop out of §5.3 but remain here. The space axis is each country's production-network concentration — the Perron root λ_1 and the spectral gap ρ_{gap} of its OECD ICIO domestic technical-coefficient matrix (identical spectral reading to the US instrument, §4.6). The time axis is each country's temporal persistence — the AR(1) coefficient φ of real consumption growth in the Jordà–Schularick–Taylor Macrohistory data (the canonical persistence object in a consumption-based SDF, §3), with GDP-growth persistence as robustness. If the two axes are truly independent, their cross-country values should be uncorrelated.

They are. The model-relevant concentration measure — the Perron root λ_1, which *is* the amplification eigenvalue in the mechanism — is essentially uncorrelated with persistence: $r = -0.17$ against consumption-growth φ ($p = 0.50$) and $r = -0.19$ against GDP-growth φ ($p = 0.46$), both far from significance and both weakly *negative*. The spectral-gap reading tells the same story: ρ_{gap} vs. GDP-growth persistence is flat ($r = +0.03$, $p = 0.90$). The only non-trivial number, ρ_{gap} vs. consumption-growth persistence ($r = -0.38$, $p = 0.12$), (i) is not significant at the 10% level, (ii) points the *opposite* way to the objection — concentration goes with slightly *less* persistent consumption growth, not more — and (iii) does not survive the GDP-growth robustness check. There is no coherent, robust link in the direction the objection requires.

This is the strongest form of support available for an assumption: the premise the composition needs is one the data independently satisfy. Honest caveats: $n = 18$ is a small panel, the AR(1) is a coarse one-parameter summary of persistence, and network concentration is measured recently (2016–2019) against long-run growth series. The result is therefore corroboration, not proof — but it is corroboration precisely where the paper is most exposed, and it goes the right way.

5.6 Cross-section versus time series: the sign flips *(empirical estimate)*

The cross-country test (§5.3) reads the concentration–premium link in the *cross-section*: economies with more concentrated production networks carry a higher average premium. A natural companion question is whether the same link holds over *time* — does U.S. network concentration $\rho_{\text{gap}}(t)$ move with the premium year by year? The BEA summary tables give an annual $\rho_{\text{gap}}(t)$ over 1997–2023; we correlate it with the annual compounded market excess return (bea_network_rho.py::network_premium_timeseries).

quantity	value	reading
coeff. of variation of $\rho_{\text{gap}}(t)$	≈ 0.06	structurally stable
contemporaneous corr(ρ_{gap}, prem)	≈ -0.40	negative
one-year-ahead corr($\rho_{\text{gap}}(t)$, prem_{t+1})	≈ -0.30	negative

The time-series sign is *opposite* to the cross-sectional one — and that is coherent, not contradictory. Across countries, higher structural concentration carries a higher *average* premium (a level, risk-compensation effect). Within one country over time, a *spike* in concentration marks a high-valuation state — the 2000 dot-com peak is the textbook case, with ρ_{gap} at its sample maximum 1.59 just before the 2000–02 drawdown — which predicts *lower* forward returns, the familiar valuation/reversal channel. The two signs live on different axes: the cross-section is about the premium's *level*, the time series about its *valuation state*. The coefficient of variation ($\approx 6\%$) settles which axis the mechanism primarily inhabits: ρ_{gap} barely moves, so the concentration channel is a structural, cross-sectional determinant of the premium's level, not a business-cycle driver of its fluctuations. This is why §5.3 used a structural cross-section rather than a time-series regression — and it is stated as a suggestive $n = 27$ pattern, not a confirmed predictive result.

5.7 A discriminating prediction: the equity term structure *(verified structure)*

Every result so far concerns the premium's *level*, and the level is exactly where the persistence and concentration channels are observationally equivalent — both produce the same single-period premium. To tell them apart we need an observable on which they *disagree*. The equity **term structure** is that observable. van Binsbergen, Brandt and Koijen (2012) documented that short-maturity dividend strips earn *higher* returns than long-maturity claims: the term structure of equity risk premia slopes *downward*. This is a standing puzzle for the leading resolutions — long-run-risk and external-habit models both imply an *upward*-sloping term structure, the opposite of the data.

The two channels of this paper make opposite predictions, so the observed sign discriminates between them (term_structure_proof.py, 6 theorems, 0 axioms):

- **Persistence channel (long-run risk) → upward.** A near-unit-root latent consumption component loads risk on long horizons: cumulative exposure grows with maturity, so the long-maturity claim is riskier (persistence_term_premium_upward) — this holds when the intertemporal elasticity of substitution exceeds unity (the Bansal–Yaron preferred calibration; for IES < 1, the standard CRRA region, the sign can reverse, as agents hedge long-run uncertainty — van Binsbergen and Koijen 2017, §2.3). At the BY calibration this is exactly why LRR/habit *fail* the van Binsbergen test.
- **Concentration channel (this paper) → downward.** The cross-sectional amplification propagates through the dominant network mode as a geometric series with ratio $\lambda_1 < 1$ (the cumulative amplification is $1/(1-\lambda_1)$, a *bounded* geometric sum — multiperiod_sdf_proof.py). The marginal risk added at horizon k is λ_1^k, which *decays* in k: risk is front-loaded, so short strips bear more (concentration_marginal_decays, concentration_term_premium_downward). The term structure slopes *downward* — the van Binsbergen sign.

The kernel proves the two slopes have strictly opposite signs (slopes_have_opposite_sign: their

product is negative), and that an observed downward slope matches the concentration channel while contradicting the persistence one (downward_data_matches_concentration, and the assembled term_structure_discriminator_theorem).

From sign to magnitude. The direction is only half the test; the concentration channel also fixes the *rate*. A maturity-n dividend strip retains a share λ_1^{n-1} of the fresh cross-sectional exposure (the priced network shock propagating through the dominant mode, one input–output round per year), the remainder looking like the aggregate index. One identification is load-bearing here and we state it plainly: mapping *one Leontief propagation round to one calendar year* is what turns the dimensionless spectral root λ_1 into a per-year decay rate. It is a modeling choice, not a measured quantity — motivated by the annual frequency of the input–output accounts and the roughly annual horizon over which production linkages transmit shocks, but not pinned by the data; a different round-to-year mapping would rescale the predicted half-life, so the rate match below should be read as "consistent under an annual propagation clock," not as a parameter-free coincidence. Its market beta is therefore

$$\beta(n) = 1 - c\,\lambda_1^{n-1}, \qquad c = 1 - \beta(1),$$

so the *beta gap* $1 - \beta(n)$ decays by exactly λ_1 per year — and λ_1 is not a free parameter but the BEA production-network Perron root measured in §4.6 ($\lambda_1 \approx 0.48$). Anchoring the short end at the observed $\beta(1) \approx 0.5$ (van Binsbergen and Koijen 2017, Fig. 4) and fitting *nothing* to the premium or to the term structure, the law predicts (term_structure_fit.py, which takes these two externally-measured anchors — λ_1 from §4.6 and $\beta(1)$ from van Binsbergen–Koijen — as inputs and computes the implied curve; it does not re-estimate them from raw data):

maturity n (yr)	1	2	3	4	5	7
predicted $\beta(n)$	0.50	0.76	0.88	0.94	0.97	0.99
source	anchor	model	model	model	model	documented

Only two of these entries are pinned by published numbers: the short-end anchor $\beta(1) \approx 0.5$ and the long-end convergence $\beta(7) \approx 1$. Both are stated *in text* by van Binsbergen and Koijen (2017); the intermediate maturities appear only in their Fig. 4 (a graph), so the $n = 2, \ldots, 5$ entries above are **model output**, not data points we fit against — we deliberately do not manufacture a full seven-point series or an RMSE from a figure. What can be tested honestly is a *zero-free-parameter* consistency check against the two documented endpoints. The concentration-gap half-life is ≈ 1 year, and the measured network root implies $\beta(7) = 0.994$, squarely inside the documented "≈ 1" range (deviation 0.000; term_structure_fit.py, endpoint_consistency). Read as a one-sided admissibility test, the documented convergence to $\beta(7) \gtrsim 0.97$ places a *ceiling* on the decay rate, $\lambda_1 \leq 0.63$; the independently measured BEA root $\lambda_1 = 0.48$ sits 23% below that ceiling — admissible, with margin, and not tuned to hit it. A second, independent observable — the decline of equity-yield volatility from 10–18% (1yr) to 3–6% (7yr) — implies a somewhat slower per-year decay (≈ 0.83); the measured network root sits at the fast end of the resulting $[0.48,\ 0.83]$ bracket.

Why a rising market beta is consistent with a downward premium slope. A sharp objection arises here: if $\beta(n)$ *rises* from ≈ 0.5 to 1, then under a one-factor (CAPM) reading short strips are *safer* than the index and should earn *less* — an upward slope, the opposite of the claim. The objection is correct about CAPM and wrong about the model, and the gap between the two is exactly the paper's thesis. The statement is **two-factor**: $\beta(n)$ is the loading on the *aggregate index*, while the beta gap $1 - \beta(n) = c\,\lambda_1^{n-1}$ is the loading on the *concentration mode* — a second,

independently priced factor. Short strips are simultaneously *under*-exposed to the market (low β) and *over*-exposed to the concentration factor (large beta gap). The downward premium slope is therefore not a CAPM prediction — under a single market factor the slope would indeed be upward — but the signature of the concentration factor carrying a positive price that dominates the market-beta term at the short end. This is precisely why a single priced factor cannot match the term structure (Gormsen 2021), and why the resolution is rank ≥ 2: the same two-axis structure that clears the *level* (§5.5) is what bends the *term structure* the right way. A one-factor reading of $\beta(n)$ alone would falsify the downward slope; the two-factor reading the mechanism actually makes does not. One premise is worth stating plainly: for the slope to come out downward the concentration factor must carry a risk price large enough that its front-loaded exposure outweighs the back-loaded market-beta term. That the sign lands downward in the data is therefore also mild evidence on the *magnitude* of that price — a quantity we do not separately pin from the term structure itself, and which a direct two-factor estimation should bound. We attempt exactly that estimation next, and report a null.

Directly testing the concentration-factor price *(empirical estimate).* The premise above is testable in the equity cross-section, so we test it rather than assume it (concentration_factor_pricing.py). We take the *real* production network — not a return-correlation proxy — and assign every Fama–French 48 industry portfolio a network centrality via a name-based crosswalk to the BEA sectors (one BEA sector per industry; approximate by construction, disclosed as such). A concentration factor is then built as the equal-weighted excess-return spread of the top-tercile-minus-bottom-tercile industries by centrality (a characteristic-mimicking long-short, in the HML mould), and priced four ways over 1969–2026 (682 months, 47 mapped industries). To guard against the result being an artifact of one centrality definition, we run the whole test under two economically distinct readings: the dominant-mode **Perron eigenvector** of the direct-requirements matrix (the spectral-concentration object of §4.6) and the Acemoglu–Carvalho–Ozdaglar–Tahbaz-Salehi (2012) **Leontief influence vector** (the canonical systemic-importance centrality); across industries their rank correlation is only -0.17, so they are near-independent instruments. The verdict is a **robust null**:

- the tercile *premium* is economically zero under both readings (-0.4% and -0.02% per year, $t = -0.36$ and -0.02);
- centrality does *not* predict average industry returns (characteristic slopes $t = -0.94$ and $+0.48$; Pearson $r = -0.14$ and $+0.07$, both insignificant);
- the Fama–MacBeth cross-sectional price of risk λ_{conc} is insignificant under both ($t = +0.23$ and -0.21);
- the *one* nominally significant number — a $+1.7\%$/yr FF5 alpha under the Perron reading ($t = 2.0$) — is a profitability-factor residual (its RMW loading is -0.46, $t = -9.3$) and *flips sign* to -1.5% under the Leontief reading, so it is not a stand-alone risk premium.

One caveat could keep this a bound rather than a refutation: industry portfolios are weak test assets, with too little beta dispersion to price *any* factor — in the same test even the market price λ_{mkt} is insignificant ($t = 0.04$). We close that escape hatch with a *decisive* test on a broad, well-powered cross-section: the 75 canonical Fama–French portfolios (25 size–value, 25 size–profitability, 25 size–investment), independent of the industry-built factor, with a Fama–MacBeth price and a Gibbons–Ross– Shanken (1989) joint-alpha test. Here the concentration factor *is* well-identified — the cross-asset dispersion of its betas actually *exceeds* the market's (std β_{conc}/ std $\beta_{\text{mkt}} = 1.5$–1.8, $\beta_{\text{conc}} \in [-0.6, +0.4]$) — so the test has real power. The verdict is unchanged and now sharper: in a two-factor model the concentration price is significant but **sign-unstable** (-6.0%/yr, $t =$

−2.4 under Perron; +8.4%/yr, $t = +2.8$ under Leontief — two equally-motivated centralities give *opposite* signs, the signature of a factor absorbing omitted size/value spread, not a structural risk price); once the real Fama–French five factors are controlled the concentration price is **insignificant** under both ($t = +1.3$ and $+1.5$); and adding it to FF5 does **nothing** to the joint pricing errors (mean $|\alpha|$ moves 0.93% → 0.91%/yr, GRS 2.45 → 2.42, both models still rejected by the same margin). A well-powered, standard cross-section therefore does not merely fail to confirm the price — it shows the concentration factor is *subsumed* by known factors and improves no pricing error. The honest reading is that **the concentration factor's price is not measurable in the public equity cross-section**: §5.7's downward-slope magnitude leg rests on a posited positive price that neither the industry test nor a well-powered broad test delivers. This is exactly why the paper's claim is a *discriminator* (the sign and the independently measured decay *rate*) and not a *confirmation* (a calibrated factor price). The one place a concentration price could still hide — *within*-industry, across individual stocks assigned their industry's centrality — needs security-level data (CRSP/Compustat) this paper does not use, and is the natural next test.

Then which second axis *does* the data price? *(empirical estimate)* The rank-1 no-go (§5.8) says the market alone cannot span the achievable price of risk, and the term structure (above) says a second dimension is *needed*; the concentration null just says concentration is not *it*. So we ask the complementary question directly of the data (concentration_factor_pricing.py second_axis): of the canonical Fama–French factors, which carries the second price-of-risk eigenvalue beyond the market? The natural metric is the max-squared-Sharpe of Barillas and Shanken (2017): $S^2(F) = \mu^\top \Sigma^{-1} \mu$ is the squared Sharpe of the factor tangency — the Hansen–Jagannathan bound achievable with factors F — and the *marginal* S^2 of adding a factor to the market is its independent contribution to the price of risk. On FF5 monthly data (1963–2026, 754 months), the market alone attains an annual $S^2 = 0.22$ (max Sharpe 0.46); the full five-factor tangency attains $S^2 = 0.94$ (max Sharpe 0.97). **The market — one priced axis — spans only 23% of the achievable price-of-risk variance; the remaining 77% is exactly the rank-≥ 2 content the no-go predicts.** And it is *earned by a specific second axis*: the largest marginal contributions come from **investment (CMA, $\Delta S^2 = +0.37$) and profitability (RMW, +0.25)**, with size (SMB, +0.01) negligible and value (HML) *subsumed* at the tangency (Euler share −2% — the Fama–French 2015 redundancy, reproduced here). So the data's empirically successful second dimension is the quality/investment axis, not network concentration: the rank-≥ 2 theorem is confirmed as a *structural necessity*, and the axis that fills it in public equities is RMW/CMA. This is the honest division of labor between the paper's two halves — the *theorem* says a second dimension is forced; the *data* say which one is priced, and it is not the one our mechanism proposed.

Verdict: a quantitative discriminator that favors concentration — now a rate, not just a sign, honestly bracketed. The single-period premium cannot separate the channels; the term structure can, on *two* counts. Its *sign* (downward) is the one concentration predicts and persistence gets wrong; and its *convergence rate* — betas rising from ≈ 0.5 to ≈ 1 over five-to-seven years — is reproduced by the independently measured network λ_1 with no premium fit. We do not claim a pinned magnitude: the short-end beta is used as an anchor and two observables bracket the decay rate rather than pin it — and, as the direct test just reported shows, the concentration factor's *price* is not separately confirmed in the FF48 cross-section, so the magnitude leg remains a posit rather than a measured quantity. But this is the paper's sharpest falsifiable prediction, and it is the one place the concentration reading is not merely *admissible* (§4.5, §5.5) but *quantitatively preferred* by an independent piece of asset-pricing evidence. It also lands where the literature independently arrived: Gormsen (2021) finds the term structure's average and cyclicality cannot be matched by any *single*-factor model and requires two priced factors — exactly the rank-≥ 2 conclusion of §5.8.

An honest complication: the slope's direction is itself contested. Bansal, Miller, Song and Yaron (2021) argue that the downward *sample-average* slope in van Binsbergen, Brandt and Koijen's data is a short-sample artifact: their 2004–2017 window over-represents recessions (especially in Europe and Japan), and once a regime-switching model corrects for this, the *unconditional* risk-premium slope implied by standard models — and by their bias-corrected estimate — is *positive* (upward); only Japan remains significantly downward, the U.S. slope is significantly *upward*. If this critique is right, "the term structure slopes downward" is not the stable, structural fact this paper (and van Binsbergen, Brandt and Koijen 2012) has treated it as — it is instead *state-dependent*, downward in recessions and upward in expansions, and the long-run average is genuinely uncertain given how few full cycles the strip data cover.

This matters for our mechanism specifically, not just for van Binsbergen and Koijen's reading. Our concentration channel is *not* cyclical: the measured network Perron root λ_1 is structurally stable (coefficient of variation $\approx 6\%$ over 1997–2023, §5.6), so a mechanism driven purely by λ_1 predicts a term-structure slope that does **not** flip sign with the business cycle. If the true slope *is* state-dependent — downward in recessions, upward in expansions, as Bansal et al. argue — that is evidence for a cyclical (risk-price) channel *layered on top of* whatever structural channel exists, not evidence against a structural channel outright: the two need not be mutually exclusive, and disentangling them requires exactly the kind of regime-conditional analysis Bansal et al. perform, applied to the concentration channel specifically — a test this paper does not attempt. We therefore soften the claim above: the concentration channel's *sign prediction* matches the original van Binsbergen, Brandt and Koijen (2012) reading and the quantitative beta-convergence law is a *separate* empirical fact (a cross-sectional CAPM-beta pattern, not the risk-premium slope Bansal et al. dispute) that their critique does not directly address — but the "quantitatively preferred" characterization above should be read as conditional on the slope-direction debate being resolved in van Binsbergen and Koijen's favor, which is not settled.

5.8 The impossibility is a class property: the rank-1 no-go *(verified structure)*

The no-go of §5.2 shut *one* channel — a single spectral ρ bounded by measured concentration. But that result was never really about concentration. This section proves it is a property of the resolution's *dimensionality*: **no rank-1 mechanism — any resolution driven by a single scalar parameter through a monotone amplifier — can clear the premium once that parameter is empirically bounded** (rank1_nogo_proof.py, 12 theorems, 0 axioms).

The unification is exact. Long-run risk amplifies through one persistence ϕ (amplifier $1/(1-\phi)$); external habit through one surplus-consumption sensitivity; the concentration channel through one Perron root λ_1 (amplifier $1/(1-\lambda_1)$). Persistence and concentration are *literally the same increasing map* $r = 1/(1-x)$ — the kernel proves this common form is positive (reciprocal_amplifier_positive) and monotone (reciprocal_amplifier_monotone) once, covering both channels at a stroke. For *any* such amplifier, the empirical bound on its single parameter transports to a ceiling on the premium (monotone_amplifier_bounded, amplifier_ceiling_caps_premium), and if that ceiling is below target the premium is provably unreachable — the grand rank1_class_nogo_theorem, which subsumes §5.2's concentration result and the long-run-risk and habit channels as *instances of one theorem*.

One clarification forestalls an obvious objection. Bansal–Yaron long-run risk is not a one-parameter model — it also carries an IES, a leverage ratio, and volatility-of- volatility, and its calibrated persistence ($\kappa_1\phi \approx 0.98$) sits far above the AR(1) figure §4.4 bounds. What the rank-1 characterization

captures is the model's *dominant amplification channel*: fixing the other preference parameters, the premium amplification runs through the single reciprocal map $1/(1-\kappa_1\phi)$, and the no-go binds that channel *once its persistence input is held to the empirically admissible range*. The claim is therefore not that LRR has one free parameter, but that its premium-generating amplification is rank-1 in the persistence it feeds on; the model's remaining dimensions (IES, leverage) rescale the map without adding a second independent amplification axis. §9.3 develops why this projection is the right level of description for a class no-go.

This turns a familiar objection into the paper's point. One could dismiss the latent amplifier as "just a reparametrization of long-run risk or habit" (§9.3). Precisely — and *that is why the no-go bites*: because all three are rank-1, the impossibility that sinks one sinks all of them together. The escape is then pinned exactly. The kernel proves the minimal second factor needed to reach target *strictly exceeds one* (escape_factor_exceeds_one), so no single amplifier can be rescaled into feasibility; only a genuinely two-dimensional mechanism — two independent parameters composed — reaches the premium (rank2_composition_reaches_target). That is not a new escape hatch: it is exactly the disciplined composition of §5.5 (concentration × persistence) and the combined-amplification clause of the capstone.

How large is the gap? The no-go is not a knife-edge miss, and the kernel now bounds the shortfall. Writing the premium's demand as an amplification $\rho^* \approx 4.6$ and the admissible concentration ceiling as $\rho_{\max} \approx 1.4$ (§5.2), the best-case rank-1 resolution — the amplifier run right to its empirical ceiling — reaches at most the fraction $\rho_{\max}/\rho^* \approx 0.30$ of the observed premium (rank1_premium_ceiling_fraction, division-free form premium $\cdot\, \rho^* \le$ target $\cdot\, \rho_{\max}$). Rank-1 therefore attains only about a *third* of the premium; the missing two-thirds is a multiplicative shortfall of $\rho^*/\rho_{\max} \approx 3.3$ that a second axis must supply (rank1_shortfall_strictly_positive). The dimensional deficit is thus large and quantified, not a rounding correction — which is also why a rank-1 mechanism cannot be rescued by re-tuning its single parameter.

How many axes, exactly? A data-anchored minimum rank. The abstract verdict is "rank ≥ 2," but the one measured fraction above pins a sharper count. The Ky Fan ladder (§9) makes the rank-r *additive* ceiling — a linear stack of r independent priced factors — at most r times the top eigenvalue, $\sum_{i\le r}\lambda_i \le r\,\lambda_{\max}$ (kyfan_top_r_sum_below_r_times_max); clearing the target therefore needs $r\,\lambda_{\max} \ge$ target, i.e. $r \ge$ target$/\lambda_{\max}$ (additive_rank_lower_bound) — the machine-checked minimum-rank count of rank1_nogo_proof.py §10 (5 theorems, 0 axioms). With the measured single-axis strength $\lambda_{\max}/\text{target} = \rho_{\max}/\rho^* \approx 0.30$ this ratio is $\rho^*/\rho_{\max} \approx 3.3$: a rank-3 linear stack undershoots ($3\lambda_{\max} \approx 0.9\,\text{target} <$ target, rank3_additive_insufficient_at_ceiling) while rank-4 clears (rank4_ceiling_clears_at_measured_bracket), so the *additive* minimum rank is exactly four (minimum_additive_rank_is_four). The margin is real but not huge — it takes $\rho_{\max} \ge \rho^*/3 \approx 1.53$ (a $\sim 9\%$ under-measurement of the 1.4 ceiling) to drop the count from four to three.

This does **not** contradict the two-axis resolution of §5.5, and the gap between "four" and "two" is itself the point. §5.5's composition is *multiplicative* ($p_0\,x\,y$), and multiplying amplifiers is *adding their logarithms* — so a single strong axis (cash-flow persistence priced by recursive preferences, $\approx 3.3\times$) carries a large share that no *equal* additive axis can. Two multiplicative axes thus clear what four comparable linear factors would. Read together, the measured 30% ceiling says the premium needs **either ≥ 4 comparable linear factors, or ≥ 2 axes with strong multiplicative interaction** — and the quantitatively successful resolutions in the literature (stochastic volatility × growth in Bansal–Yaron; disaster intensity × disaster size in Wachter) are all the second kind. The minimum-rank count is therefore not just a bigger number than two; it is the precise sense in

which the working models had to be *interactive* rather than a linear stack of priced factors.

How robust is this to the ceiling? The no-go rests on the measured concentration ceiling $\rho_{\max} \approx 1.4$, and the obvious objection is that the true ceiling might be higher. The kernel bounds exactly how much higher it would have to be: a rank-1 amplifier reaches the target *only if* the admissible ceiling is at least the full demand, $\rho_{\max} \geq \rho^* \approx 4.6$ (nogo_survives_unless_ceiling_reaches_demand). Equivalently, the measured ceiling would have to be under-stated by the factor $\rho^*/\rho_{\max} \approx 3.3$ for the result to fail. The no-go therefore does not hinge on the precise value 1.4 — only on 1.4 sitting comfortably below 4.6, a 3.3× margin that no plausible re-measurement of network concentration closes.

Risk aversion and dimension are substitutes. In the Mehra–Prescott parametrization the un-amplified premium is $\gamma\sigma_c^2$, so an amplifier scales it to $\gamma\sigma_c^2\rho$. Two things follow. First, the no-go is *risk-aversion-conditional*: it holds at any γ for which the rank-1 ceiling $\gamma\sigma_c^2\rho_{\max}$ stays below target (rank1_nogo_is_gamma_conditional) — a statement about *plausible* risk aversion, not a claim that no γ works. Second, and more usefully, risk aversion and mechanism amplification are *exchangeable inputs*: a rank-1 amplifier at its ceiling divides the risk aversion needed to hit target by exactly $\rho_{\max}$, and a rank-k mechanism divides it by the product of its axis ceilings (amplifier_divides_required_risk_aversion). This puts a number on the trade-off. The classic Mehra–Prescott calculation needs a coefficient of order $\gamma \approx 40$–50 with no amplifier; the rank-1 concentration channel ($\rho_{\max} \approx 1.4$) lowers that by only its own factor, to ≈ 30–35 — still far outside the plausible range — whereas a genuine second dimension is what brings the required risk aversion down toward single digits. The dimensional escape and the high-risk-aversion escape are therefore the same ledger read two ways: the literature buys the extra dimension precisely because the price in risk aversion is fixed at implausible levels.

The whole of §5 therefore collapses to a single dichotomy: **the single-ρ door is provably shut for the entire rank-1 class (§5.2, §5.4, §5.8); the only open door is rank ≥ 2 (§5.5).** The impossibility is not a fact about one channel — it is a fact about how many dimensions a resolution needs, and the answer is: more than one.

Here "rank" means **mechanism-axis dimension**, not matrix rank. That distinction is now formalized rather than left as rhetoric. The axis-dimension theorem (axis_dimension_proof.py, 10 theorems, 0 axioms) writes the composed premium as p_0xy, with x the cross-sectional axis and y the time-series axis, each neutral at one. If the space-only ceiling $p_0x_{\max}$ is below target, moving the space axis alone fails (space_axis_alone_fails_under_ceiling); if the time-only ceiling $p_0y_{\max}$ is below target, moving the time axis alone fails (time_axis_alone_fails_under_ceiling). The grand theorem axis_dimension_theorem then proves the genuine two-axis necessity: if a composed mechanism reaches target while both one-axis ceilings are below target, then **both** axes must be strictly active, $x > 1$ and $y > 1$. This is the precise formal content behind the phrase "rank ≥ 2": not a metaphor, but a verified statement that success cannot be obtained by sliding along either axis alone. The theorem presumes the two axes are genuinely distinct dimensions rather than one force in disguise; §5.5.2 confirms this empirically — across 18 economies network concentration and cash-flow persistence are uncorrelated, so the "two axes" the theorem counts are, in the data, two.

A resolvability criterion — necessity *and* sufficiency in one line. The no-go (nothing rank-1 clears the premium) and the construction (§5.5, the two measured axes do) are, until now, two separate results. They collapse into a single criterion. On the empirically admissible box $0 < x \leq x_{\max}$, $0 < y \leq y_{\max}$, the premium p_0xy is increasing in each axis, so it is maximized at the *corner* $(x_{\max}, y_{\max})$ (box_premium_bounded_by_corner). The premium is therefore resolvable

within admissible bounds **if and only if the joint (product) ceiling $p_0 x_{\max} y_{\max}$ clears the target**: below threshold, no admissible point reaches it (subthreshold_box_infeasible); at or above threshold, the corner itself is a witness (corner_reaches_target_when_joint_ceiling_clears). This is the whole argument in one sentence — the *marginal* ceilings $p_0 x_{\max}$ and $p_0 y_{\max}$ each fall short (the no-go), while the *joint* ceiling of the two measured axes clears it (the construction). Resolvability is thus not a coincidence of the particular channels we picked but a property of the admissible box: the premium is resolvable *exactly* when — and because — the economy affords two independent axes whose ceilings multiply past the target. Rank-2 is the criterion, not merely a mechanism that happens to work.

Representation independence — the ceiling is a property of the SDF, not of our parametrization. A reader may object that the product form $p_0 xy$ (and the amplifier form $p_0 \cdot \text{amp}$) *builds in* the result. It does not: the same ceiling follows from the Hansen–Jagannathan bound alone, for *any* stochastic discount factor (rank1_nogo_proof.py §8, 6 theorems). The HJ bound (§6, hj_is_sharpe_bound) states that for every SDF m and every excess return, $\pi^2 \le (\sigma^2(m)/E[m]^2)\,\sigma^2(R)$; writing $S^2 := \sigma^2(m)/E[m]^2$ for the squared price of risk gives $\pi^2 \le S^2\,\sigma^2(R)$ with no assumption on how m is parametrized (sdf_price_of_risk_bounds_premium). Now read "rank" as the dimension of the SDF's *priced-factor* structure: for orthogonal factors the price-of-risk variance is additive, $S^2 = a_1 + a_2 + \cdots$, with $a_j \ge 0$ the variance contribution of axis j — and §5.5.2 is precisely what licenses the sum, since the two axes we use are empirically uncorrelated. A **rank-1** SDF prices a single axis, $S^2 = a_1$; if that axis is admissibly bounded, $a_1 \le c_1$, then $\pi^2 \le \sigma^2(R)\,c_1$ *regardless of the amplifier's functional form* (rank1_premium_squared_ceiling). The bound references only $(\sigma^2(R), c_1)$, so it is identical for every rank-1 SDF — the representation- independence claim, now a theorem. If the target exceeds it, no rank-1 SDF resolves the premium (rank1_nogo_representation_free); a second orthogonal axis strictly lifts the ceiling (rank2_variance_ceiling_exceeds_rank1); and the resolvability criterion reappears in SDF-native form — the premium is resolvable *iff* the summed axis-variance ceiling clears the target, $\sigma^2(R)\,(c_1 + c_2) \ge \text{target}^2$ (sdf_resolvability_necessity, sdf_resolvability_sufficiency). The product parametrization of §5.5–§5.8 is one instance of a statement that holds for the whole SDF. This is the sense in which "rank ≥ 2" is *structural* rather than parametric: it is a fact about how much variance a bounded, low-dimensional SDF can carry, read off the same Hansen–Jagannathan frontier every consumption-based model must respect.

The spectral ceiling: "rank" is the eigenvalue count, and the ceiling is an eigenvalue sum (Ky Fan). The additive decomposition $S^2 = a_1 + a_2 + \cdots$ above is not a modelling choice — it is the spectral decomposition of the SDF's price-of-risk matrix. Write the priced risk as a symmetric second-moment matrix M; by the spectral theorem (cited as external linear algebra) M has orthogonal principal axes with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$, and these *are* the axis variance contributions a_j. A mechanism of algebraic **rank** r loads on at most r independent directions, so by **Ky Fan's maximum principle** its attainable price-of-risk variance is bounded by the sum of the r largest eigenvalues, $S^2_{\text{rank } r} \le \lambda_1 + \cdots + \lambda_r$, hence $\pi^2 \le \sigma^2(R)\,(\lambda_1 + \cdots + \lambda_r)$. We machine-verify the content this rests on (rank1_nogo_proof.py §9, 6 theorems): the **Rayleigh enclosure** $\lambda_{\min}\|b\|^2 \le b^\top M b \le \lambda_{\max}\|b\|^2$ for a symmetric 2×2 M, proved from the sum-of-squares witnesses of the PSD form with *no diagonalization assumed* (rayleigh_form_bounded_by_eigenvalue_cap, rayleigh_form_above_min_eigenvalue); that the **rank-1 ceiling is the top eigenvalue** $\pi^2 \le \sigma^2(R)\,\lambda_{\max}$ (rank1_ceiling_is_top_eigenvalue, so the abstract c_1 of the previous paragraph is concretely $\lambda_{\max}$); that the **rank-2 ceiling is the top-two sum** (rank2_ceiling_is_top_two_eigenvalues, Ky Fan's selection inequality); that the ceiling is **monotone in the rank** (rank_ceiling_monotone_in_premium); and that **full rank hits the**

trace, $\pi^2 \leq \sigma^2(R)\,\mathrm{tr}(M)$, the entire Hansen–Jagannathan variance (full_rank_ceiling_is_trace). The eigenvalues are sorted, so the marginal gain from the k-th dimension is λ_k, *nonincreasing* in k: there are diminishing returns to dimension, and the first two modes carry the most amplification — the spectral reason rank-2 is not only necessary but, in the economies we measure, enough. "Rank" is thus the honest linear-algebraic rank of the SDF's priced-factor loading, and the whole rank ladder — rank-1 $= \lambda_{\max}$, rank-r = top-r sum, full rank = trace = the HJ bound — is a theorem about eigenvalues, not about our parametrization.

The rank-≥ 2 verdict, from three directions. This is the paper's central result, and it is over-determined — three independent arguments converge on the same dimension count. *Theory* (this section): the class no-go caps every single-parameter resolution below the premium and forces a second factor strictly greater than one. *Data* (§5.7): the equity term structure, which Gormsen (2021) shows no single priced factor can match, demands exactly two factors — the same rank arrived at from an entirely different observable, and one the level of the premium is blind to. *Construction* (§5.5): the two axes we actually measure — cross-sectional concentration and time-series persistence — compose, by a derived (not assumed) multiplicativity law (§5.5.1), to clear the premium with no fitted parameter. A skeptic can reject the constructive leg (the persistence channel is weakly identified) and the theory still stands; can reject the concentration story and the term structure still says rank ≥ 2. The claim that survives all the negatives in this paper is not "ρ resolves the puzzle" but the sturdier, more general one: *whatever* resolves it cannot be one-dimensional.

5.9 A forward prediction: concentration should order the term structure across countries *(verified structure + preliminary confrontation)*

The results above are established. This section states the framework's sharpest *forward* prediction — one the literature has not tested as a cross-country structural link, and one that only the concentration reading makes — and then confronts it with the sparse cross-region data that exist, honestly. If the downward equity term structure is driven by a priced network mode decaying at the Perron root λ_1 (§5.7), then, *holding the business-cycle state fixed*, a country's term structure should be *ordered by its production-network concentration*: the higher λ_1, the slower dividend-strip betas converge to one, and the more drawn-out the downward slope. The beta gap is $g(n) = c\,\lambda_1^{n-1}$, so the whole curve's steepness is monotone in λ_1 — a comparative static, not a level statement (and, as the confrontation below makes clear, a *within-regime* one).

This is a genuinely new implication on two counts. First, the existing literature (van Binsbergen, Brandt and Koijen 2012; Gormsen 2021) documents *that* the term structure slopes downward and varies over time, but not that its *cross-country magnitude* tracks production-network structure — that link is unique to this framework. Second, the relevant spectral quantity here, the Perron root λ_1 (the *convergence rate*), is *distinct* from the concentration ratio $\rho_{\text{gap}} = \lambda_1/\lambda_2$ that carries the cross-country *premium* signal (§5.3). The framework therefore makes *two independent* spectral predictions from *two different* features of the same network — a demanding, falsifiable structure, not a single knob fit twice.

The comparative static is machine-verified (term_structure_proof.py §6, 4 theorems, 0 axioms): for economies sharing a short-maturity gap c but differing in λ_1, the higher-λ_1 economy has a strictly larger beta gap at every maturity (higher_lambda_larger_gap_m2, higher_lambda_larger_gap_m3) and a strictly larger cumulative steepness (cumulative_steepness_monotone, steeper_term_structure_from_higher_perron_root). Feeding the 16-economy production net-

work Perron roots (OECD ICIO, the same instrument as §5.3) through the verified law yields a concrete predicted ranking (term_structure_fit.py::cross_country_prediction):

Predicted term-structure steepness	Economies (by network λ_1)
steepest / slowest-converging	GBR (0.50), PRT (0.49), AUS (0.46), ITA (0.45)
middle	CHE, JPN, FRA, ESP, FIN, SWE, USA, DEU
flattest / fastest-converging	NOR (0.41), BEL (0.41), NLD (0.38), DNK (0.37)

The predicted cumulative-slope spread across the panel is $\approx 1.7\times$ (GBR versus DNK).

A preliminary confrontation — and what it teaches about the right test. Cross-country dividend-strip data exist for exactly four markets: the Euro Stoxx 50, the Nikkei 225, the FTSE 100, and the S&P 500 (van Binsbergen, Brandt and Koijen 2012; van Binsbergen and Koijen 2017). Their *unconditional* average downward slopes do **not** line up with our λ_1 ranking. The downward slope is strongest for the Euro Stoxx 50 and statistically insignificant for the Nikkei, FTSE, and S&P individually (van Binsbergen and Koijen 2017); Bansal, Miller, Song and Yaron (2021) find the U.S. term structure is if anything mildly *upward* while Europe and Japan are mildly downward. Against our network roots that gives a mixed scorecard: the U.S. (low $\lambda_1 \approx 0.41$) being the flattest is consistent with the prediction, but the Euro area (a low-to-middling $\lambda_1 \approx 0.42$ for its constituents) being the *steepest* and the U.K. (the *highest* $\lambda_1 \approx 0.50$) being among the *weakest* run against it. The crude unconditional ranking is not supported.

We read this not as a refutation of the mechanism but as evidence that the *unconditional* cross-country ranking is the **wrong test** — for three concrete reasons, each independent of our framework. (i) *The slope is regime-dependent.* Bansal, Miller, Song and Yaron (2021) show the term structure is downward in recessions and upward in expansions, so a country's *average* slope is dominated by how many recessions its short sample happens to contain — a country sampled in more recessionary years masquerades as "steeper" irrespective of its network structure. (ii) *The Euro Stoxx 50 has no single country* λ_1*.* It is a multi-country index (France, Germany, the Netherlands, Spain, Italy, …), so matching its slope to any one national input–output Perron root is a category error; its constituents' λ_1 average low, yet the index is the steepest — exactly the kind of aggregation mismatch that a single-country test avoids. (iii) *The individual-market slopes are statistically insignificant* (only the pooled world portfolio is significant — van Binsbergen and Koijen 2017), so a four-point unconditional ranking has almost no power to begin with.

The prediction, correctly stated, is regime-conditional. The verified comparative static says that *holding the business-cycle state fixed*, a higher-λ_1 economy has a more drawn-out downward structure. The clean test is therefore within-regime and single-country: for economies observed in the *same* state, rank their strip-beta convergence by λ_1. That test does not yet exist in the data — it needs single-country strip series long enough to condition on the state, which currently exist only for the U.S. and, marginally, Europe and Japan. **Status: an open, riskable prediction whose first crude confrontation fails the naive unconditional form and thereby pins down the form a decisive test must take.** The one clean data point we do have — the U.S. being the flattest, with the lowest λ_1 — is consistent; everything else awaits regime-conditioned, single-country strip data. We report the non-confirmation plainly rather than presenting the four-market snapshot as support it does not provide.

---

6. Markowitz / Hansen–Jagannathan bridge *(verified algebra)*

Returning to verified-fact territory: the bridge file (sdf_markowitz_bridge_proof.py) connects the SDF view to mean–variance analysis. These are classical identities, now machine-checked.

- **CAPM from the SDF** (beta_pricing_identity, market_beta_is_one, beta_portfolio_additivity): the beta-pricing relation is recovered from $E[mR] = 1$.
- **HJ–Sharpe duality** (hj_sharpe_duality, max_sharpe_from_frontier, premium_equals_sr_times_vol): the Hansen–Jagannathan bound is the SDF-side dual of the maximal Sharpe ratio on the efficient frontier.
- **Capital market line** (cml_slope_is_sdf_ratio, two_fund_sdf_decomposition, combined_vol_linear): the CML slope equals $\sigma(m)/E[m]$.
- **Amplification on the frontier** (amplified_sharpe_ratio, amplified_beta_pricing, frontier_shift_steeper_cml): *if* ρ amplifies the SDF volatility, the frontier steepens accordingly — again a conditional statement, with the conditioning made explicit.

Structural lemmas (ratio_eq_from_numerator_eq, monotone_ratio, product_ratio_cancel, squared_ratio_cleared, ratio_ordering) support the above; they are pure real-arithmetic facts.

---

7. Formal framework

Modeling hypotheses (assumed, clearly marked as such in the kernel):

- H_sdf_pricing — the SDF pricing relation $E[mR] = 1$.
- H_rf_def — the risk-free rate definition $R_f = 1/E[m]$.
- H_cov_decomp — the covariance-decomposition bookkeeping identity.

Established facts (standard mathematics, used as building blocks):

- F_var_nonneg — variance is non-negative.
- F_cauchy_schwarz_cov — Cauchy–Schwarz for covariances.

Everything else is a derived theorem. The kernel reports **0 axioms**, **0 sorry**, and **0 derivation debt**: no theorem leans on an unproven escape hatch.

A terminological note for the formally-minded reader: in strict proof-theoretic usage the five items above (three hypotheses, two facts) *are* axioms of the formal system — unproven premises from which the theorems are derived. We reserve the word "axiom" for an *undisclosed* escape hatch inside the deductive machinery, and use "hypothesis"/ "fact" for these five because they are economic modeling choices and standard mathematical lemmas, explicitly listed and cleanly separable from the derived mathematics. "0 axioms" thus means "no hidden premises," not "no premises at all": the premises are exactly the five disclosed here, and nothing else.

A second honesty point about *scope of composition*. The kernel verifies the SDF-side relations (H_sdf_pricing, H_rf_def, H_cov_decomp) and the Mehra–Prescott impossibility inequality as *independent components*, not as a single end-to-end deduction that runs from a probability space through the SDF to the numerical bound. There is no measure-theoretic layer here, so

H_cov_decomp is a bookkeeping identity rather than a derived covariance decomposition, and the step from "SDF pricing" to "premium = $\gamma\sigma_c^2$" is the standard economic interpretation supplied by us, not machine-derived. What the formalization guarantees is that *each algebraic link is exact*; the economic reading that chains them is disclosed and conventional, not part of the certificate. This is the honest boundary of what "machine-checked" buys in a kernel without probability primitives.

---

8. Proof architecture

All proofs live in elysium/ fields /equity_premium/ and are verified in the formal proof kernel (exportable to Lean 4). Combined: **163 theorems proved, 0 axioms, 0 sorry, 0 derivation debt** across 11 proof files.

The count is a hygiene figure, not the contribution: most of the 163 are elementary arithmetic and identity checks that exist so that no step is taken on faith. The paper's weight rests on three load-bearing results — the class no-go (rank1_nogo_proof.py::rank1_class_nogo_theorem, §5.8), the axis-dimension necessity (axis_dimension_proof.py::axis_dimension_theorem, §5.8), and the derived multiplicativity of the two channels (composition_proof.py, §5.5.1) — with the term-structure discriminator (term_structure_proof.py, §5.7) as the empirical counterpart. A reader auditing the argument should start there; the rest is scaffolding that makes those four machine-checkable end to end.

File	Role	Epistemic class
equity_premium_proof.py	SDF pricing, premium decomposition, HJ bound, Mehra–Prescott impossibility, and the latent-premium derivation	§2 verified fact; §3 conditional
sdf_markowitz_bridge_proof.py	CAPM-from-SDF, HJ–Sharpe duality, CML, frontier amplification	verified fact (amplification clause conditional)

File	Role	Epistemic class
multiperiod_sdf_proof.py	Geometric-sum identities, convergence, spectral amplification $\rho/(\rho-1)$, channel comparison; and the *recursive (Epstein–Zin) pricing of the time axis* (§7) — the long-run-risk price $\lambda_e = (\gamma - 1/\psi)\kappa_1/(1-\kappa_1\phi)$ is zero under power utility (the $\rho_{\text{eff}} \approx 1$ of §4.5), positive/increasing/unbounded in persistence under EZ, and in S^2 terms a strong second eigenvalue $a_2 = k^2 a_1$ — the constructive complement of the rank-1 no-go (§4.5)	conditional on the spectral model (EZ block verified structure, 5 thm, 0 axioms)
unified_rho_capstone_proof.py	Cross-domain unification of ρ (premium contagion anomalies)	framework hypothesis
spectral_concentration_proof.py	Cross-sectional channel: spectral-gap concentration, monotone amplification, channel composition, premium bracketing (§4.6)	verified structure (7 thm, 0 axioms)
premium_nogo_proof.py	Single-ρ no-go: a concentration-bounded ρ cannot clear the premium (§5.2)	verified structure (6 thm, 0 axioms)
composition_proof.py	The composition law: multiplicativity of the two orthogonal channels is *derived*, not assumed — a separable joint response factorizes, the product solves the joint geometric equation and is the unique solution, and the multiplicative composition dominates the additive one (§5.5.1)	verified structure (6 thm, 0 axioms)

File	Role	Epistemic class
rank1_nogo_proof.py	The rank-1 *class* no-go and its extensions (statements in formal_statements.md): any single-parameter monotone amplifier (long-run risk, habit, concentration — one shared map $1/(1-x)$) is capped below the premium, so escape requires rank ≥ 2 (§5.8); the quantitative ceiling ($\leq \rho_{\max}/\rho^* \approx 30\%$) and its robustness bound (3.3× under-measurement to break); the γ-bridge (risk aversion and dimension are substitutes); the *representation-independent* ceiling from the HJ bound for an arbitrary SDF (§6); the *spectral rank ceiling* (Ky Fan — Rayleigh enclosure, rank-r = sum of the r largest eigenvalues); the *data-anchored minimum rank* (a linear stack needs ≥ 4 axes, a multiplicative 2-axis mechanism suffices — §5.8); and the *puzzle-agnostic generalization* to any puzzle with a large multiplicative gap and a bounded single channel (excess volatility, value premium; §9.3, §11)	verified structure (32 thm, 0 axioms)
axis_dimension_proof.py	Axis-dimension theorem plus the resolvability criterion: under individual space-axis and time-axis ceilings, any composed mechanism reaching the target must move both axes above neutral; and the premium is resolvable within admissible bounds *iff* the joint (product) ceiling clears the target — necessity and sufficiency in one criterion (§5.8)	verified structure (10 thm, 0 axioms)

File	Role	Epistemic class
impossibility_frontier_proof.py	The impossibility frontier: Pareto boundary, truncated endpoints, dimensional collapse — the full feasible set of the single-ρ class (§5.4)	verified structure (9 thm, 0 axioms)
term_structure_proof.py	Equity term-structure discriminator: concentration front-loads (downward slope), persistence back-loads (upward), opposite signs, downward data selects concentration; the quantitative $\beta(n) = 1 - c\lambda_1^{n-1}$ beta-convergence law (§5.7); and the cross-country comparative static — higher network λ_1 steeper term structure (§5.9)	verified structure (14 thm, 0 axioms)
empirical_rho.py	Estimates naive AR(1) and latent-component (state-space MLE) persistence on FRED consumption data; computes implied $\widehat{\rho}$ (§§4.4–4.5), the variance-weighted ρ_{eff}, the disciplined two-channel composition (§5.5), and the Epstein–Zin recursive amplification k, $S^2_{EZ}/S^2_{\text{pow}}$ under the BY calibration (epstein_zin_amplification, §4.5)	empirical estimate (not a kernel proof)
crossdomain_rho.py	Cross-domain shared-ρ test: premium-implied vs. Fama–French-implied ρ (Section 5.1)	empirical estimate (not a kernel proof)
spectral_concentration.py	Cross-sectional ρ probe: eigenvalue gap of FF industry-portfolio return correlations (§4.6)	empirical estimate (not a kernel proof)
bea_network_rho.py	Real production-network ρ: spectral gap and Leontief multiplier of the BEA input–output Direct Requirements matrix (§4.6)	empirical estimate (not a kernel proof)

File	Role	Epistemic class
crosscountry_rho.py	Cross-country panel: network-concentration ρ_{gap} vs. equity premium across 16 economies (OECD ICIO + Jordà–Schularick–Taylor); §5.3. Also the time-series test $\rho_{\text{gap}}(t)$ vs. premium (§5.6, via bea_network_rho.py)	empirical estimate (not a kernel proof)
term_structure_fit.py	Quantitative term-structure calibration: the measured network λ_1 and the observed short-maturity beta reproduce the strip-beta convergence to 1 over 5–7 years, with no premium fit (§5.7); plus the §5.9 prediction — the 16-economy λ_1 steepness ranking and its honest confrontation with the four strip markets (unconditional ranking not supported; reframed as regime-conditional)	empirical estimate (not a kernel proof)
orthogonality_test.py	Orthogonality test of the two axes: across 18 economies, network concentration ($\lambda_1/\rho_{\text{gap}}$, OECD ICIO) is uncorrelated with cash-flow persistence (AR(1) of consumption/GDP growth, JST) — direct support for the independence premise behind multiplicativity (§5.5.1–5.5.2) and the axis-dimension theorem (§5.8)	empirical estimate (not a kernel proof)

File	Role	Epistemic class
concentration_factor_pricing.py	Direct asset-pricing test of the concentration-factor price §5.7 posits: BEA network centrality (Perron eigenvector *and* Leontief influence vector) crosswalked to FF48 industries, a centrality-mimicking long-short factor, priced four ways (premium, FF5 alpha, Fama–MacBeth λ_{conc}, characteristic slope). Robust **null** across both centralities. A decisive mode adds a well-powered broad cross-section (75 FF size/value/profitability/investment portfolios, std $\beta_{\text{conc}} >$ std β_{mkt}) with Fama–MacBeth + a GRS joint-alpha test: the two-factor price is sign-unstable across centralities, vanishes under FF5, and adds nothing to GRS — confirming the null is real, not a power artifact (§5.7). A second_axis mode answers the complementary question — the Barillas–Shanken max-squared-Sharpe decomposition of FF5 shows the market spans only 23% of the achievable price of risk and the leading second axis is investment/profitability (CMA/RMW), not concentration (§5.7)	empirical estimate (null + attribution)

---

9. Discussion

9.1 What stands on its own

Two results stand on their own, independent of everything speculative in the paper, because both are axiom-free verified algebra. The **Section 2 impossibility** — "the standard consumption-CAPM provably cannot generate the observed premium, certified with zero axioms" — is the safe core, and it is deliberately placed first. The **Section 5.8 class no-go** is the second such result and

the paper's central one: it holds for *any* single-parameter monotone amplifier, so it survives the rejection of every specific ρ-story in the paper. Even if the concentration channel, the persistence channel, and the cross-domain identity are all wrong, "no rank-1 resolution can clear an empirically-bounded premium" remains a theorem. The empirical legs (§§4.4–4.6, 5.1, 5.3, 5.7) decorate and corroborate these two verified spines; they do not carry them.

9.2 What is conditional, and what would make it empirical

Sections 3–5 prove implications of the form "$\rho = r \Rightarrow$ premium $= p$." To turn any of this into an empirical resolution of the puzzle one would need, at minimum:

1. **An estimator for ρ.** *Done, with a mixed verdict (Sections 4.4–4.5).* We read ρ off consumption data through the persistence channel ($\rho = 1/(1-\hat{\phi})$), independently of the premium. The naive version fails ($\hat{\rho} \approx 1$); the latent-component (state-space) version is admissible but weakly identified ($\hat{\rho}$ between 2.6 and 4.4 depending on the estimator, against the $\rho \approx 4.6$ required) — and once variance-weighted (§4.5), the persistence channel's *effective* amplification collapses to $\rho_{\text{eff}} \approx 1$ under power utility, so any amplification it delivers must come from recursive-preference pricing, not from the variance of the persistent component. The cross-country panel of Section 5.3 adds a sample dimension as *corroboration* (not a load-bearing test): in a small, window-sensitive panel of 16 economies, network-concentration ρ tracks the premium ($r \approx 0.5$) specifically through concentration — the one direction the reading could most easily have failed — while the sharper empirical anchor is the term structure (point 4).
2. **An out-of-sample test.** *Done (Section 5.1).* We tested the shared-ρ claim against Fama–French factor data and rejected its strong form: the premium-implied ρ (≈ 4.6) and the anomaly-implied ρ (≈ 1.5) differ threefold. What remains is the contagion leg, which needs a proportionality constant public data do not fix.
3. **A standard-errors story.** *Done (Sections 4.4–4.5, 5.1).* The 6.48%-vs-6% illustration remains a point output of a point input, but every *estimate* now carries uncertainty: the naive $\hat{\rho}$ has a delta-method 95% CI of [0.84, 1.04] (quarterly), which excludes the required 4.6; the latent $\hat{\phi}$ has a profile-likelihood CI spanning the whole stationary range, which cannot exclude it; and the cross-domain ρ_{anom} has a bootstrap CI of [1.43, 1.67], far below the premium-implied 4.6. The error bars are what turn each verdict from an assertion into a measurement.
4. **A discriminating observable.** *Done (Section 5.7), with a contested premise flagged.* The premium's *level* cannot separate the persistence and concentration channels — they are observationally equivalent there. The equity *term structure* can, on two counts. Its *sign*, on the van Binsbergen, Brandt and Koijen (2012) reading, is downward, which concentration predicts and persistence gets wrong — but Bansal, Miller, Song and Yaron (2021) argue this sign is a short-sample artifact and the unconditional slope is upward, a live debate we engage directly in Section 5.7 rather than resolve. Its *rate* — dividend-strip betas rising from ≈ 0.5 to ≈ 1 over five-to-seven years, a separate empirical fact the 2021 critique does not directly address — is reproduced by the independently measured network $\lambda_1 \approx 0.48$ with no premium fit (Section 5.7). A direct FF48 asset-pricing test of the concentration factor's *price* — the magnitude the downward slope posits — returns a robust null (§5.7), so the discriminator rests on the *sign* and *rate*, not on a calibrated price. That the full term structure needs *two* priced factors (Gormsen 2021) is the same rank-≥ 2 verdict as the class no-go (Section 5.8). This is the one place the concentration reading is, on the sign question, only as strong as the van Binsbergen–Koijen reading of the data, but on the rate question genuinely *quantitatively*

preferred over the long-run-risk alternative — the paper's sharpest falsifiable prediction, resolved in its favor on the point that survives the debate. Section 5.9 turns this into a forward test: the same reading predicts, via a verified comparative static, that production-network concentration should *order the equity term structure across countries* (higher λ_1, steeper slope) — a riskable prediction, distinct from the premium signal, awaiting country-level strip data.

Sections 3–5 are therefore a *consistency proof for a candidate mechanism* whose naive empirical form is rejected, whose latent form is admissible-but-unidentified, and whose strong cross-domain form is rejected (Sections 4.4–4.5, 5.1) — not a claim that the mechanism operates. The one constructive empirical result is Section 5.5's disciplined composition: the two independently *measured* channels (concentration × persistence) bring the required amplification back inside the admissible band without any fitted parameter — sufficient to close the wedge, though not a confirmation that the mechanism does.

9.3 Relation to existing resolutions

The latent amplifier is, structurally, a compact way of writing what habit formation, long-run risk, and disaster models achieve by other means: raise the effective volatility (or price of risk) of the SDF. The contribution here is not a new economic story but a *minimal verified scaffold* — one parameter, two channels that agree, zero axioms — into which such stories could be poured and then tested. Section 4.4 makes the relationship concrete: the data reject ρ as the persistence of *measured* consumption growth, which is exactly why the long-run-risk resolution works through a *latent* persistent component instead. In that reading ρ is a reduced-form sufficient statistic for the effective persistence such models inject, and the empirical task is to recover it from a state-space fit rather than a naive AR(1).

Section 5.8 turns this relationship from a caveat into a result. If the amplifier is a sufficient statistic for what long-run risk and habit inject, then a bound on the amplifier is a bound on *all* of them at once: the rank-1 class no-go proves that no resolution of this shape — long-run risk, habit, or the concentration channel, all the same map $1/(1-x)$ — can clear an empirically-bounded premium, and that escape requires a second independent dimension. The scaffold's value is thus not that it is a *new* story but that, being the *minimal* one, it lets a single verified theorem speak about the whole family the literature has proposed.

A taxonomy of resolutions by axis-dimension. The resolvability criterion gives a lens on the whole literature: classify each proposed resolution by how many *independent* amplifying axes it moves. A single-channel amplifier — one scalar state variable, however elaborate its map — is rank-1 and therefore subject to the single-axis ceiling; the quantitatively successful resolutions are, without exception we are aware of, rank-2.

Resolution	Amplifying channel(s)	Axes	Criterion reading
Habit (Campbell–Cochrane 1999)	surplus-consumption ratio (one slow state)	1	rank-1: one ceiling-bounded channel
Long-run risk, growth-only	persistent expected-growth component	1	rank-1: one channel

Resolution	Amplifying channel(s)	Axes	Criterion reading
Long-run risk, calibrated (Bansal–Yaron 2004)	growth persistence × stochastic volatility	2	rank-2: escapes
Rare disasters, constant (Rietz 1988; Barro 2006)	fixed tail probability/size	1	rank-1: one channel
Time-varying disasters (Gabaix 2012; Wachter 2013)	tail size × time-varying intensity	2	rank-2: escapes
Concentration (this paper)	network spectral concentration	1	rank-1: measured ceiling $\rho_{\max} \approx 1.4 < \rho^*$, *provably* capped
Concentration × persistence (§5.5)	space × time	2	rank-2: clears the wedge

Two honesty caveats keep this from overreaching. First, we have *measured* a below-demand ceiling only for our own concentration channel (§5.2); calling habit or the growth-only long-run-risk channel "capped" asserts only that they are rank-1 and hence fall *within the scope* of the class no-go — whether each fails turns on its own admissible ceiling, which we do not estimate here. Second, Epstein–Zin recursive preferences are the *lens*, not an axis: they separate risk aversion from the elasticity of intertemporal substitution so that a second shock can be priced, but they add no amplification by themselves — the escape comes from the shock structure (volatility, intensity) they make priceable.

With those caveats, the organizing thesis is sharp and, we believe, new: **the quantitatively successful resolutions did not find a better stochastic discount factor — they found a second dimension.** Bansal–Yaron works because stochastic volatility is an axis independent of growth persistence; Wachter works because time-varying intensity is an axis independent of disaster size. The criterion says this is not incidental: rank-1 is capped, so *any* successful resolution must be rank-2, and the empirically-relevant question becomes which two axes a given economy actually affords. Our own contribution is one measured pair (concentration × persistence) that clears — but the more durable point is that "how many independent dimensions?" is the precise question the literature has been answering implicitly for forty years.

The no-go is not even specific to the equity premium. Stripped of its numbers, the class no-go is a statement about a *shape*: a puzzle is a triple (un-amplified baseline $b > 0$, single-axis amplifier ceiling c, demanded target τ); a rank-1 mechanism attains at most $b\,c$, so if $b\,c < \tau$ no admissible single amplifier $\rho \le c$ clears ($b\rho \le bc < \tau$, puzzle_rank1_nogo_general), while a second *multiplicative* axis strictly raises the ceiling to $b\,c^2$ whenever $c > 1$ (puzzle_rank2_strictly_exceeds_rank1) — the rank-≥ 2 escape, puzzle-agnostic. The equity premium is the instance in which we have measured all three numbers. Other quantitative asset-pricing puzzles share the shape. The cleanest is Shiller's (1981) *excess-volatility* puzzle: prices move some $K \approx 5$–$13\times$ more than a present-value model driven by realized dividend changes allows, and a single mean-reversion/persistence amplifier is bounded well below that gap — exactly the K-fold-gap template puzzle_kfold_gap_nogo ($c < K$ and $\tau \ge Kb$ force $bc < \tau$), which takes the gap factor K as a parameter so each puzzle is one

instance with its own measured (K, c). The value premium and the credit-spread puzzle plausibly fit the same mould. We state these as *structural* observations, not new calibrated results: what the three theorems (rank1_nogo_proof.py §11, 0 axioms) establish is that the "rank-1 is capped, escape needs a second axis" logic is a property of the puzzle *shape*, and measuring each puzzle's own ceiling is the natural way to extend the program.

9.4 Limitations

- The premium figures in Sections 3–5 are illustrative outputs of chosen inputs, not estimates; only Sections 4.4–4.5 and 5.1 report actual estimates.
- The time-series estimates (Sections 4.4–4.5) are single-country and in-sample; the cross-country panel (Section 5.3) broadens this to 16 economies but is itself small-n, so its significant post-1950 concentration–premium relation ($r \approx 0.5$, $p \approx 0.04$) is suggestive rather than decisive and is window-sensitive. We now quantify that fragility rather than assert it (§5.3): under leave-one-out the slope stays uniformly positive and tightly banded, and an exact 10^5-permutation test confirms $p \approx 0.04$, but 5% significance survives only 11 of 16 drops and dropping Denmark lifts p to 0.118 — a robust *sign*, a one-country-fragile 5%. Standard errors are now reported throughout (OLS delta-method for the naive AR(1), profile-likelihood for the latent component, bootstrap for the cross-domain ρ_{anom}), but the latent-component persistence remains weakly identified — its $\hat{\rho}$ is a wide range, not a number, and once variance-weighted its effective amplification is $\rho_{\text{eff}} \approx 1$ under power utility (§4.5) — and the anomaly leg's tangency Sharpe is an in-sample, upward-biased quantity.
- The "unified ρ" cross-domain identity is the strongest claim; its strong (single shared number) form is *rejected* by the Section 5.1 test, leaving only the weak qualitative version (each domain governed by some $\rho > 1$, with the proven tradeoff signs). The anomaly leg uses an in-sample tangency Sharpe and the contagion leg is not independently identified.
- The orthogonality test (§5.5.2) that supports the two-independent-axes premise is itself small-n (18 economies), summarizes persistence by a single AR(1) coefficient, and matches recently-measured network concentration to long-run growth series; it is corroboration that the axes are empirically independent, not proof that no shared latent force exists.
- The formalization works with first-order (log-linear) premium approximations consistent with the Mehra–Prescott setup; higher-order terms are out of scope.
- Reproducibility of the empirical legs requires local setup, not just the scripts: the estimators depend on numpy/scipy/openpyxl and on external datasets staged locally (OECD ICIO for the network matrices, the Jordà–Schularick–Taylor Macrohistory Database for cross-country premia and growth, FRED for U.S. consumption, and the Kenneth French library for the anomaly tangency). In a bare interpreter the scripts fail at import; a referee reproducing the numbers must install the dependencies and stage the source files first. Inferential p-values are computed with an exact Student-t tail (self-contained incomplete-beta), so they do not depend on whether SciPy is present.
- The term-structure discriminator's empirical anchor is itself contested: Bansal, Miller, Song and Yaron (2021) argue the downward sample-average slope van Binsbergen, Brandt and Koijen (2012) report is a short-sample, recession-overrepresentation artifact, and that the unconditional slope implied by standard models is upward. We engage this directly in §5.7 rather than treating the downward slope as settled; the concentration channel's sign prediction and the separate beta-convergence law survive this critique better than the "quantitatively preferred" framing did before. The §5.9 cross-country prediction, confronted with the only four strip markets that exist, is *not* supported in its naive unconditional form (only the low-

λ_1 U.S. being flattest fits; the Euro area and U.K. run against it); we reframe it as an open, regime-conditional, single-country test rather than presenting the four-market snapshot as support it does not provide.

- **The one remaining concentration test, pre-registered.** The industry-level pricing test (§5.7) is a robust null, but it cannot see a premium that lives *within* an industry, across individual firms of differing supply-chain centrality — the last place a concentration price could hide. That test is well-defined but needs firm-level data this paper does not use, and we state it as a pre-registration rather than run an underpowered proxy. (1) *Centrality construction:* for each firm-quarter, build a firm-level network-centrality score from customer–supplier links — Cohen–Frazzini (2008) customer links, Compustat segment data, or FactSet Revere — as the firm's Katz/Perron centrality in the realized production graph, *not* its industry's centrality (the industry-level score has no within-industry variation, which is exactly why the FF48 test cannot substitute). (2) *Sort/regression:* within each 4-digit industry and quarter, form centrality quintiles from CRSP-covered firms and run a Fama–MacBeth cross-sectional regression of excess returns on centrality beta, industry fixed effects absorbing the (already-null) between-industry variation. (3) *Prediction:* given the between-industry null and the concentration channel's posited *positive* price, the honest prior is a weak or null within-industry premium; a clean null would close the concentration channel as an empirical pricing factor, while a positive λ_{conc} would revive the §5.7 magnitude leg. (4) *Power:* anchored to the FF48 test's HAC standard error, detecting a 2%/yr centrality premium at 80% power needs roughly the full CRSP cross-section ($\gtrsim$ 3000 firms) over 30+ years — feasible with the data above, out of reach for the public portfolio series used here. This keeps the concentration claim exactly as strong as the evidence: a *discriminator* (sign + independent decay rate), never a *confirmed factor price*.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used Cursor and its integrated AI models (Claude and GPT families) in order to assist with manuscript drafting, structural editing, and language polishing. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

- Mehra, R. and Prescott, E. C. (1985). *The Equity Premium: A Puzzle.* Journal of Monetary Economics, 15(2), 145–161.
- Hansen, L. P. and Jagannathan, R. (1991). *Implications of Security Market Data for Models of Dynamic Economies.* Journal of Political Economy, 99(2), 225–262.
- Fan, K. (1949). *On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations I.* Proceedings of the National Academy of Sciences, 35(11), 652–655.
- Campbell, J. Y. and Cochrane, J. H. (1999). *By Force of Habit.* Journal of Political Economy, 107(2), 205–251.

- Bansal, R. and Yaron, A. (2004). *Risks for the Long Run.* Journal of Finance, 59(4), 1481–1509.
- Beeler, J. and Campbell, J. Y. (2012). *The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment.* Critical Finance Review, 1(1), 141–182.
- Epstein, L. G. and Zin, S. E. (1989). *Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework.* Econometrica, 57(4), 937–969. (Recursive preferences separating risk aversion from the EIS.)
- Rietz, T. A. (1988). *The Equity Risk Premium: A Solution.* Journal of Monetary Economics, 22(1), 117–131. (Constant rare-disaster resolution.)
- Barro, R. J. (2006). *Rare Disasters and Asset Markets in the Twentieth Century.* Quarterly Journal of Economics, 121(3), 823–866.
- Gabaix, X. (2012). *Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance.* Quarterly Journal of Economics, 127(2), 645–700. (Time-varying disaster intensity as a second axis.)
- Wachter, J. A. (2013). *Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?* Journal of Finance, 68(3), 987–1035.
- van Binsbergen, J. H., Brandt, M. and Koijen, R. S. J. (2012). *On the Timing and Pricing of Dividends.* American Economic Review, 102(4), 1596–1618. (Downward-sloping equity term structure.)
- van Binsbergen, J. H. and Koijen, R. S. J. (2017). *The Term Structure of Returns: Facts and Theory.* Journal of Financial Economics, 124(1), 1–21. (Strip-beta convergence to one with maturity; Sharpe ratios declining with maturity.)
- Gormsen, N. J. (2021). *Time Variation of the Equity Term Structure.* Journal of Finance, 76(4), 1959–1999. (The average and cyclicality of the slope require two priced factors — a single factor cannot match them.)
- Bansal, R., Miller, S., Song, D. and Yaron, A. (2021). *The Term Structure of Equity Risk Premia.* Journal of Financial Economics, 142(3), 1209–1228. (Argues the sample-average downward slope in van Binsbergen and Koijen's 2004–2017 window is a short-sample artifact — Europe and Japan over-sample recessions — and that the *unconditional* slope implied by standard models, and by a bias-corrected estimate, is positive; only Japan is statistically significantly downward.)
- Cochrane, J. H. (2005). *Asset Pricing* (Revised ed.). Princeton University Press.
- Acemoglu, D., Carvalho, V. M., Ozdaglar, A. and Tahbaz-Salehi, A. (2012). *The Network Origins of Aggregate Fluctuations.* Econometrica, 80(5), 1977–2016.
- Gabaix, X. (2011). *The Granular Origins of Aggregate Fluctuations.* Econometrica, 79(3), 733–772.
- Laloux, L., Cizeau, P., Bouchaud, J.-P. and Potters, M. (1999). *Noise Dressing of Financial Correlation Matrices.* Physical Review Letters, 83(7), 1467–1470.
- Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M. and Taylor, A. M. (2019). *The Rate of Return on Everything, 1870–2015.* Quarterly Journal of Economics, 134(3), 1225–1298. (Jordà–Schularick–Taylor Macrohistory Database, R6.)
- OECD (2023). *OECD Inter-Country Input–Output (ICIO) Tables, 2023 edition.* http://oe.cd/icio.