Complexity Breaks Arbitrage Pricing Theory *

Carter Davis[1] and Alejandro Lopez-Lira[2]

[1] *The Ohio State University*

[2] *University of Florida*

This Version: March 19, 2026

Abstract

Arbitrage Pricing Theory predicts pricing errors vanish as markets grow. Three findings reverse this prediction. First, under estimation uncertainty, equilibrium pricing errors scale as N/t (number of assets over sample size), with a persistent component from time-varying fundamentals that no estimator can eliminate. Second, the stochastic discount factor loads primarily on estimation errors rather than fundamental risk factors, explaining why risk-based factor models fail to price assets. Third, multiple strategies achieve high Sharpe ratios with low cross-correlations as they exploit distinct dimensions of the growing estimation error space. When investors disagree about expected returns, mispricing persists even with informed investors because risk aversion prevents them from taking sufficiently large positions to eliminate complexity-driven pricing errors. Empirically, individual stock returns require 145 principal components to explain half their variance. Systematic hedge funds' active positions exhibit near-zero correlations (average 0.07), consistent with exploitation of distinct estimation-error dimensions.

*We thank multiple seminar participants at the Chicago Booth Asset Pricing Conference, Carnegie Mellon University, the University of Florida, CUNEF University, ITAM, Seoul National University, NOVA University, the Theory-based Empirical Asset Pricing Research Workshop, the European Finance Association, Michigan State University, the Korean University and the Korea Advanced Institute of Science & Technology for helpful comments and suggestions. Correspondence: ckd1@iu.edu, alejandro.lopez-lira@warrington.ufl.edu. There are no financial disclosures or conflicts of interest.

For decades, asset pricing theory has sought parsimonious models linking expected returns to systematic risk factors. The capital asset pricing model (CAPM), the Arbitrage Pricing Theory (APT), and their multi-factor successors produce large pricing errors in empirical tests. Machine-learning strategies generate high Sharpe ratios with low pairwise correlations, and hundreds of predictive signals have little connection to traditional risk factors. Why do factor models fail to price assets, and what drives the proliferation of profitable, weakly correlated anomalies?

Complexity resolves this puzzle. APT relies on the fact that idiosyncratic return risk diversifies away in large portfolios. But idiosyncratic estimation uncertainty does not. In high-dimensional markets, agents must estimate expected returns for each of N assets from limited data. Each additional asset adds a new source of estimation error, and unlike idiosyncratic return shocks, these errors do not diversify away. Any portfolio, however diversified, spans a single direction in weight space, while estimation errors accumulate across N dimensions. Because agents' demands reflect these undiversified errors, equilibrium prices inherit them: rational risk-averse agents facing uncertainty about expected returns and covariances generate pricing errors that grow with market size.

The paper makes three findings. First, as complexity increases through more assets or predictive signals, equilibrium pricing errors grow linearly with the number of assets. The scaling has two components. A learnable part, from uncertainty about permanent expected returns, shrinks with sample size. A persistent part, from time-varying fundamentals, cannot be learned away regardless of sample size. This bound requires only minimal conditions on the risk-adjusted uncertainty structure and no assumptions on preferences or learning algorithms; Section 1 provides the intuition; the Appendix establishes the bound formally. APT predicts pricing errors vanish as the number of assets grows; the model predicts the opposite. Under perfect information, systematic factors span the Sharpe ratio of the oracle mean-variance portfolio as $N \to \infty$; under estimation uncertainty, the fraction captured by systematic factors goes to zero.

Second, the stochastic discount factor (SDF) loads primarily on estimation errors rather than fundamental risk factors, so traditional risk-based models fail to price assets. Third, multiple

strategies achieve high Sharpe ratios with low cross-correlations as they exploit distinct dimensions of the growing estimation error space. When investors disagree about expected returns, mispricing persists even with informed investors because risk aversion limits arbitrage.

The three findings provide a unified explanation for major empirical puzzles. Parsimonious factor models fail because agents' estimation errors, not systematic risk exposures, dominate the equilibrium SDF. Traditional risk factors cannot span an SDF whose variation comes from estimation uncertainty rather than fundamental risk. The weak beta-return relationship arises because large estimation errors obscure underlying risk premia. The proliferation of uncorrelated anomalies emerges as investors exploit different dimensions of the vast estimation error space. All three patterns follow from rational learning under complexity.

U.S. equity markets exhibit precisely this high-dimensional structure. The ratio of assets to sample size ranges from 30 to 120 for typical estimation windows, and individual stock returns require 145 principal components to explain half their variance. Aggregate pricing errors covary positively with the number of stocks, and the relationship strengthens with richer covariance estimators, consistent with estimation errors in expected returns driving the pattern. Section 4 documents these patterns.[1]

Related Literature

The paper builds on three studies of complexity in asset pricing.

Martin and Nagel (2022) establish that high-dimensional learning creates statistical illusions: in-sample return predictability is uninformative about market efficiency when the number of characteristics is comparable to the number of assets. In their framework, prices are correct given investors' information, and the apparent predictability is an artifact of the econometrician's hindsight advantage. We show that high-dimensional learning does more than create illusions: it generates real mispricing. Equilibrium pricing errors are nonzero and scale with the number of

1. See Harvey, Liu, and Zhu (2016), McLean and Pontiff (2016), Yan and Zheng (2017), Chen (2021) on declining return predictability. Missing data further exacerbates estimation challenges (Bryzgalova et al. 2022).

assets.

The distinction arises from three modeling departures. First, risk aversion transforms estimation uncertainty from a statistical nuisance into priced risk. Risk-averse agents hold finite portfolios that reflect their forecast errors, so estimation uncertainty enters equilibrium prices and generates true alphas. Under risk neutrality, no unique optimal portfolio exists, zero risk premia prevent SDF decomposition, and the equilibrium pricing consequences that follow from complexity do not arise. Second, covariance uncertainty generates variance-based risk premia absent from the known-covariance setting of Martin and Nagel (2022). Third, heterogeneous beliefs allow studying persistence with informed investors and strategy proliferation. Martin and Nagel (2022) show that learning affects how we test for market efficiency; learning also breaks the APT.

Da, Nagel, and Xiu (2022) examine the statistical limits of arbitrage in a partial equilibrium framework where alphas are exogenous. In general equilibrium, pricing errors, asset prices, and the SDF emerge endogenously from rational learning. Pricing errors grow with market complexity, reversing APT; the equilibrium SDF depends on estimation errors, explaining factor model failures; and multiple strategies achieve high Sharpe ratios with low correlations.

Kozak and Nagel (2024) demonstrate that characteristics-based factors can theoretically span the stochastic discount factor. In equilibrium, characteristics-based spanning fails: when agents face estimation uncertainty, the SDF loads primarily on estimation errors rather than observable risk factors, generating the pricing errors documented in Section 4.

Two additional studies address related puzzles from different angles. Andrei, Cujean, and Wilson (2023) explain the flat security market line as a measurement problem: investors have dispersed private information, so the empiricist mismeasures betas. In their framework, the CAPM holds for investors but appears to fail empirically. Pricing errors in our framework are objective, not measurement artifacts. The baseline uses homogeneous agents who observe identical data but face parameter uncertainty; complexity alone generates pricing errors that scale with market size. Beyond CAPM failures, multi-factor models fail because the SDF loads on estimation errors rather than risk factors, and the proliferation of uncorrelated high-Sharpe strategies follows from the same

complexity mechanism.

Contemporaneous work by Dello Preite et al. (2024) reinterprets pricing errors as compensation for unsystematic risk, grounded in underdiversification (Merton 1987). The reinterpretation does not resolve the puzzle but restates the question: why is the compensation large and persistent? Estimation uncertainty under complexity is the mechanism: pricing errors grow linearly with the number of assets, with a persistent component that no estimation procedure can eliminate because fresh innovations to expected returns each period have not yet appeared in any realized return. Underdiversification may amplify pricing errors, but complexity alone suffices to explain their magnitude and persistence.[2]

1 Mechanism

Pricing errors grow linearly with the number of assets. The mechanism does not depend on specific preferences or distributional assumptions; a tractable model in the next section embeds the mechanism in equilibrium.

1.1 The Mechanism

Consider a market with many assets whose returns have a factor structure: each asset's return equals its exposure to common factors plus an idiosyncratic component. Even when agents correctly specify this structure, they must estimate asset-level parameters from each asset's own history. The parameters are features of the data-generating process that no agent in the economy observes; the uncertainty is intrinsic, not a consequence of informational asymmetry.

The estimation error is the gap between true expected returns and the best forecast given public

2. Our work also relates to Merton (1987) on incomplete information, Molavi, Tahbaz-Salehi, and Vedolin (2024) on cognitive constraints, Bossaerts, Plott, and Zame (2007) on heterogeneous portfolios with efficient pricing, and empirical work on high-dimensional factor models (Kelly, Malamud, and Zhou 2022; Bryzgalova, Pelger, and Zhu 2020a; Didisheim et al. 2024; Bryzgalova, Huang, and Julliard 2023). Our model assumes return predictability is supported by Fama and MacBeth (1973), Jacobs and Levy (1988), Lewellen (2015), Gu, Kelly, and Xiu (2020), and Freyberger, Neuhierl, and Weber (2020a), among others. See Alti and Titman (2019) for an alternative model justifying return predictability.

information. The maximum squared Sharpe ratio achievable by exploiting the gap is a model-free measure of pricing errors. The measure is a population object, the ex-ante squared Sharpe ratio given equilibrium prices, not an in-sample optimizer that mechanically grows with the number of assets. Under perfect information the market is mean-variance efficient and the pricing error measure is exactly zero.

The mechanism is geometric. The best implementable mean-variance portfolio spans a single direction in return space. Estimation uncertainty, however, disperses across all coordinates. The optimal portfolio direction absorbs at most one dimension of uncertainty; the remaining dimensions constitute irreducible pricing error.

The gap between one implementable direction and N dimensions of uncertainty is the dimensional mismatch at the heart of the linear scaling. With many assets, there are equally many dimensions of uncertainty but only one implementable direction. The unspanned dimensions generate pricing errors that scale linearly with the number of assets. When idiosyncratic expected-return shocks are independent across firms, pricing errors grow at least proportionally to N. A weaker sufficient condition is that idiosyncratic expected-return uncertainty has bounded eigenvalues after factor extraction. When the firm-specific events that make expected returns difficult to forecast also make realized returns volatile, the risk-adjusted uncertainty per firm remains bounded.

Errors in estimating factor risk premia do not drive this bound. Such errors shift expected returns proportionally to factor loadings, moving them within the factor span; they affect the level of the implementable portfolio's Sharpe ratio but not the component orthogonal to it. Only genuinely firm-specific uncertainty, which cannot be absorbed by any portfolio direction, drives the linear scaling.

1.2 Undiversifiable Uncertainty

We call the result the Undiversifiable Uncertainty Theorem. Unlike idiosyncratic return risk, uncertainty about expected returns does not diversify. This distinction explains why APT's diversification argument fails. Idiosyncratic return shocks cancel in a large portfolio: if each asset

contributes independent variance σ^2, a portfolio of N assets has idiosyncratic variance $\sigma^2/N \to 0$. With vanishing variance and positive expected return, arbitrage becomes possible, forcing pricing errors to zero. But the diversification logic applies to return realizations, not return expectations. Each asset adds one more unknown expected return to estimate. These estimation errors do not cancel in a portfolio; they accumulate. An investor seeking to exploit pricing errors must estimate N expected returns from finite data. The estimation errors span N dimensions while any implementable portfolio spans just one. APT's arbitrage requires knowing which direction to trade; estimation uncertainty prevents precise identification of that direction.

The Appendix establishes the result formally:

Theorem (Undiversifiable Uncertainty: Informal Statement)**.** *Consider a market with many assets and a small number of factors. Decompose each firm's idiosyncratic expected return into a permanent component and a time-varying component. The gap between the squared Sharpe ratio of the oracle portfolio and that of the best implementable portfolio is bounded below by:*

1. *A learnable term that shrinks with sample size, reflecting uncertainty about permanent expected returns.*

2. *A persistent term that remains bounded away from zero regardless of sample size, reflecting time-varying expected returns.*

Both terms scale linearly with the number of assets, provided estimation uncertainty is spread across assets rather than concentrated in a few portfolio directions.

Under minimal assumptions on the risk-adjusted uncertainty structure (Assumptions A.1–A.3 in the Appendix), both terms scale linearly with the number of assets. The learnable term is proportional to the number of assets divided by sample size; the persistent term is proportional to the number of assets times the variance of innovations to expected returns. Idiosyncratic return variance enters the denominator: mispricing is smaller relative to return risk for more volatile assets.

The persistent component arises because expected returns are time-varying. Each period, new shocks shift firm-level expected returns. Even with infinite data and optimal estimation, pricing errors that scale linearly with the number of assets persist. The fixed-parameter case where pricing errors vanish with sample size is an optimistic lower bound that holds only when fundamentals are constant. The learnable and persistent decomposition applies to uncertainty about expected returns. The equilibrium model (Section 2) identifies a separate source of pricing errors from covariance estimation uncertainty, which is learnable under static parameters; whether time-varying covariances generate an analogous persistent component is left for future work.

This result requires no assumptions on preferences (CARA, Epstein-Zin, behavioral), learning mechanism (Bayesian, frequentist, machine learning), or estimator properties (biased or unbiased). The bound is estimator-agnostic: suboptimal estimators can only increase error covariance, so the result applies to any forecasting method.

1.3 What the Model Adds

A tractable CARA-normal model in the next section serves three purposes: first, to show the mechanism operates when prices clear markets and the SDF emerges endogenously from optimization; second, to characterize how the SDF loads on estimation errors versus fundamental risk factors (explaining why factor models fail); and third, to derive results on strategy proliferation, the role of informed investors, and cross-sectional return patterns. The model is deliberately simple (CARA utility, normal distributions, known factor loadings) to isolate the complexity mechanism. The scaling result extends beyond this specific setting (Appendix).

2 Model

We work with economies with an increasing number of assets (N) and a sample of t observations, where N and t grow jointly with N/t bounded away from zero. Let d_t represent the N dimensional column vector of dividends in period t. Similarly, p_t denotes the N-dimensional column vector

of asset prices at time t. Share excess returns, dollar excess returns, or excess dollar payoffs are defined as $r_{t+1} = d_{t+1} - (1 + r_f)p_t$ where r_f is a risk-free exogenous rate.[3]

Because dividends drive payoffs, the return covariance matrix $\Lambda_t = \mathbb{V}\text{ar}_t[d_{t+1}] = \mathbb{V}\text{ar}_t[r_{t+1}]$ is independent of endogenous prices, a simplification that aids tractability. We define the expected dividends under the objective data-generating process as $a_t = E_t[d_{t+1}]$ and the expected excess returns as $\mu_t = E_t[r_{t+1}]$. For notational simplicity, we omit time subscripts where the context permits.

A unit measure of identical CARA agents has risk aversion γ and faces per-asset supply $\frac{1}{N}$ as in Kozak, Nagel, and Santosh (2018). Let ι denote the N-dimensional column vector of ones.[4] Section 3 extends the framework to heterogeneous agents.

All proofs are in the Appendix.

2.1 Statistical Assumptions

For each asset i, the dividend process is given by:

$$d_{i,t+1} = b'_{i,t} x_{t+1} + c_{i,t} z_{i,t+1} \tag{1}$$

where x_{t+1} is an S-dimensional column vector of state variables, and $b_{i,t}$ is a vector of loadings. These loadings $b_{i,t}$ can be interpreted as firm-specific characteristics (such as size, age, capital structure, productivity, or management quality) that influence dividend outcomes when interacting with productivity shocks captured by x_{t+1}.[5]

Each element of $b_{i,t}$ represents the sensitivity or weight of asset i's dividends to these systematic shocks. For example, $b'_{i,t} x_{t+1}$ can be considered a reduced form specification of the production

3. The terms share excess returns, dollar excess returns, and excess dollar payoffs are all used in the literature. Following convention, we adopt this formulation rather than percentage returns.

4. This assumption ensures that under perfect information and common dividend variation, the maximum Sharpe ratio of the economy remains well-behaved as N increases. Specifically, with per-asset supply $\frac{1}{N}$, we have $\mu = \gamma_N \Lambda_t \iota = \frac{\gamma}{N} \Lambda_t \iota$, so $SR(w^*)^2 = \mu' \Lambda_t^{-1} \mu = \gamma_N^2 \iota' \Lambda_t \Lambda_t^{-1} \Lambda_t \iota = \frac{\gamma^2}{N^2} \iota' \Lambda_t \iota$. With standard factor structure where $\iota' \Lambda_t \iota \in O(N^2)$, this yields $SR(w^*)^2 \in O(1)$, keeping the maximum Sharpe ratio bounded as N grows.

5. Alternatively, x_{t+1} can represent the coefficients of OLS regressions as in Martin and Nagel (2022).

function, where firm production potentially covaries with each other via common shocks. Moreover, there is a firm-specific state variable $z_{i,t+1}$ with loading $c_{i,t}$. Without the idiosyncratic component, all N assets would be spanned by S factors, and estimation complexity would not grow with N. With idiosyncratic risk, each asset requires separate parameter estimation, generating N independent estimation problems (see Appendix for formal treatment).

Stacking the previous equation across all assets, we express the dividend vector as:

$$d_{t+1} = B_t X_{t+1} \tag{2}$$

where d_{t+1} is the N dimensional vector of dividends, and X_{t+1} is the $S + N$ dimensional column vector defined as $X_{t+1} = (x_{1,t+1}, ..., x_{S,t+1}, z_{1,t+1}, ..., z_{N,t+1})'$ and B_t is the $N \times (S+N)$ matrix defined as

$$B_t = \begin{bmatrix} b'_{1,t} & c_{1,t} & 0 & \dots & 0 \\ b'_{2,t} & 0 & c_{2,t} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b'_{N,t} & 0 & 0 & \dots & c_{N,t} \end{bmatrix}. \tag{3}$$

Under the objective data-generating process, conditional on nature's information set at time t, denoted $\mathcal{F}_t$, the state variables follow:

$$X_{t+1} \mid \mathcal{F}_t \sim N(\pi, V), \tag{4}$$

where π is the mean vector and (V) is the covariance matrix of (X_{t+1}). Consequently, dividends d_{t+1} follow a multivariate normal distribution with mean $B_t\pi$ and covariance $B_t V B'_t$.

We use "objective" and "true" interchangeably for quantities from the data-generating process, as opposed to agents' estimates.

We assume investors perfectly observe B_t at time t.[6] For the general case, they lack exact knowledge of π and V, relying instead on prior beliefs and Bayesian updates using observed

6. In the Internet Appendix, we extend the model to allow unknown factor loadings determined by firm characteristics. The baseline $O(N/t)$ result remains the dominant term under standard conditions.

dividends up to time t. Subsequent sections analyze specific scenarios where investors have perfect knowledge of π, V, or both, to understand how different types of estimation uncertainty affect equilibrium outcomes.

2.2 Agents' Estimation

We begin with the general case where agents observe neither the mean vector π nor the covariance matrix V of the state variables. Given the dividend process and state variable structure, agents must form beliefs about future dividends based on historical observations. With t observations of the state variables, agents update their beliefs using Bayesian inference.

2.2.1 Prior Distributions

For joint estimation of both the mean vector and covariance matrix, agents use a normal-inverse-Wishart prior, which is the conjugate prior for multivariate normal distributions with unknown mean and covariance. Specifically, the prior is:

$$V \sim \text{Inverse-Wishart}(\nu_0, V_0^{-1}) \tag{5}$$

$$\pi | V \sim N(\pi_0, V/\kappa_0) \tag{6}$$

where:

- π_0 is the prior mean vector for π
- $\kappa_0 > 0$ represents the strength of the prior belief about π
- $\nu_0 > S + N - 1$ is the degrees of freedom parameter for the inverse-Wishart distribution[7]
- V_0 is the prior scale matrix for V, under the parametrization Inverse-Wishart(ν, Ψ^{-1}) with $\mathbb{E}[V] = \Psi/(\nu - (S+N) - 1)$, so V_0 plays the role of Ψ

7. To ensure the posterior variance remains well-defined as N increases, we set $\nu_0 = N + S + \delta$ for some $\delta > 1$. This guarantees the posterior degrees of freedom $\nu_t = \nu_0 + t - (S+N) + 1 = t + \delta + 1$ exceeds 2 for all $t \geq 0$, necessary for the variance of the Student-t distribution to be defined.

2.2.2 Posterior Distributions

Let $\mathcal{I}_t$ denote the investors' information set at time t, consisting of observed state variables up to t. Variables denoted with a hat (e.g., $\hat{a}_t$, $\hat{\Lambda}_t$) represent agents' subjective estimates based on $\mathcal{I}_t$. The posterior mean of π is:

$$\hat{\pi}_t = \frac{\kappa_0 \pi_0 + t\bar{X}_t}{t + \kappa_0}, \tag{7}$$

where $\bar{X}_t = \frac{1}{t}\sum_{\tau=1}^{t} X_\tau$ is the sample mean. The scale matrix for the predictive t-distribution is defined as:

$$\hat{\Sigma}_t = \frac{\kappa_0 + t + 1}{(\kappa_0 + t)(\nu_0 + t - (S+N) + 1)} \left(V_0 + \sum_{\tau=1}^{t} (X_\tau - \bar{X}_t)(X_\tau - \bar{X}_t)' + \frac{\kappa_0 t}{\kappa_0 + t} (\bar{X}_t - \pi_0)(\bar{X}_t - \pi_0)' \right), \tag{8}$$

with degrees of freedom updated as $\nu_t = \nu_0 + t - (S+N) + 1 = t + \delta + 1$ (using our prior specification).[8] Together, these imply a posterior predictive distribution for X_{t+1} given by:

$$X_{t+1} \mid \mathcal{I}_t \sim t_{\nu_t}(\hat{\pi}_t, \hat{\Sigma}_t), \tag{9}$$

where t_{ν_t} denotes a multivariate t-distribution with ν_t degrees of freedom, location $\hat{\pi}_t$, and scale matrix $\hat{\Sigma}_t$.

For tractability in portfolio optimization, agents use a normal approximation to this t-distribution. This approximation becomes increasingly accurate as the degrees of freedom $\nu_t = t + \delta + 1$ grow with the sample size, since the t-distribution converges to normal as $\nu_t \to \infty$. Agents approximate it as:

$$X_{t+1} \mid \mathcal{I}_t \sim N(\hat{\pi}_t, \hat{V}_t), \tag{10}$$

where the covariance matrix is adjusted as $\hat{V}_t = \frac{\nu_t}{\nu_t - 2}\hat{\Sigma}_t$ to match the variance of the t-distribution

8. $\hat{\Sigma}_t$ as defined here is not the posterior mean of V from the Inverse-Wishart distribution. Rather, it is the scale parameter appearing in the predictive Student-t distribution for X_{t+1}. These differ by factors involving $\nu_0 + t$ and $\kappa_0 + t$. The additional factor $\frac{\kappa_0 + t + 1}{(\kappa_0 + t)}$ is precisely the standard correction that appears in predictive distribution formulas when integrating out both π and V.

under this normal approximation.

This belief about state variables translates directly to beliefs about dividends. Since $d_{t+1} = B_t X_{t+1}$, investors perceive dividends as normally distributed:

$$d_{t+1} \mid \mathcal{I}_t \sim N(B_t \hat{\pi}_t, B_t \hat{V}_t B_t'), \tag{11}$$

with expected dividends $\hat{a}_t = B_t \hat{\pi}_t$ and covariance matrix $\hat{\Lambda}_t = B_t \hat{V}_t B_t'$.

Consequently, investors perceive excess returns as:

$$r_{t+1} \mid \mathcal{I}_t \sim N(\hat{\mu}_t, \hat{\Lambda}_t), \tag{12}$$

where $\hat{\mu}_t = \hat{a}_t - (1 + r_f)p_t$ represents their expected excess returns and $\hat{\Lambda}_t = B_t \hat{V}_t B_t'$ is the perceived covariance matrix of returns. These subjective beliefs shape agents' demands and, in turn, equilibrium prices.

For the special cases where agents know either π, V, or both, the estimation problem simplifies accordingly. When π is known, agents use the true mean directly ($\hat{\pi}_t = \pi$) with no estimation error. When V is known, the covariance estimation becomes exact ($\hat{V}_t = V$).[9]

2.2.3 Special Case: Known Covariance Matrix

When the covariance matrix V is known, agents only estimate the mean vector π. The posterior mean remains:

$$\hat{\pi}_t = \frac{\kappa_0 \pi_0 + t\bar{X}_t}{t + \kappa_0}, \tag{13}$$

but the predictive distribution simplifies to:

$$X_{t+1} \mid \mathcal{I}_t \sim N\left(\hat{\pi}_t, V + \frac{V}{t + \kappa_0}\right), \tag{14}$$

9. These complete information scenarios can be formally represented as limiting cases of our Bayesian framework. For known π, set the prior mean $\pi_0 = \pi$ with $\kappa_0 \to \infty$, which forces $\hat{\pi}_t = \pi$ for all t. For known V, set $V_0 = \nu_0 V$ with $\nu_0 \to \infty$, ensuring $\hat{V}_t = V$. When both parameters are known, apply both adjustments simultaneously to eliminate all estimation uncertainty, resulting in agents' beliefs matching the true distribution: $d_{t+1} \mid \mathcal{I}_t \sim N(B_t \pi, B_t V B_t')$.

where the second term represents parameter uncertainty.

2.2.4 Special Case: Known Mean Vector

With known mean vector π, agents estimate only the covariance matrix V. The prior for V is Inverse-Wishart(ν_0, V_0^{-1}). The posterior distribution for V after observing t samples is Inverse-Wishart$(\nu_{post,t}, (V_0 + S_\pi)^{-1})$, where $\nu_{post,t} = \nu_0 + t$ and $S_\pi = \sum_{\tau=1}^{t}(X_\tau - \pi)(X_\tau - \pi)'$.

The posterior predictive distribution for X_{t+1} follows a multivariate t-distribution:

$$X_{t+1} \mid \mathcal{I}_t, \pi \sim t_{\nu_{pred,t}}(\pi, \hat{\Sigma}_t), \tag{15}$$

where the degrees of freedom are $\nu_{pred,t} = \nu_0 + t - (S + N) + 1$, and the scale matrix is

$$\hat{\Sigma}_t = \frac{1}{\nu_{pred,t}} (V_0 + S_\pi) . \tag{16}$$

Using a normal approximation to this t-distribution by matching the mean and variance:

$$X_{t+1} \mid \mathcal{I}_t \sim N(\pi, \hat{V}_t), \tag{17}$$

where the approximate covariance matrix $\hat{V}_t$ is defined as

$$\hat{V}_t = \frac{\nu_{pred,t}}{\nu_{pred,t} - 2}\hat{\Sigma}_t \quad \text{for } \nu_{pred,t} > 2. \tag{18}$$

2.3 Agents' Demand

Given that CARA agents perceive the world as normally distributed, and there are no transaction costs, their demand for assets takes the mean-variance form:

$$x_t = \frac{1}{\gamma}\hat{\Lambda}_t^{-1}\hat{\mu}_t = \frac{1}{\gamma}\hat{\Lambda}_t^{-1}(\hat{a}_t - (1 + r_f)p_t). \tag{19}$$

where $\hat{\Lambda}_t$ is the agents' estimated covariance matrix of returns, $\hat{\mu}_t$ is their expected excess returns, $\hat{a}_t$ is their expected dividend forecast, and p_t is the asset price vector. This demand reflects agents' optimal portfolios under their subjective beliefs, which typically deviate from the true optimal portfolio due to imperfect knowledge of the actual covariance Λ_t and expected dividends a_t. In the limiting case of perfect information, agents' estimates align with reality: $\hat{\Lambda}_t = \Lambda_t$ and $\hat{a}_t = a_t$.[10]

2.4 Equilibrium

Given a per-asset supply of $\frac{1}{N}$, market clearing requires $x_t = \frac{1}{N}\iota$, where ι is the N-dimensional vector of ones. Substituting the demand function into this condition and solving for prices yields the equilibrium price vector:

$$p_t = \frac{1}{1+r_f}\left(\hat{a}_t - \frac{\gamma}{N}\hat{\Lambda}_t\iota\right). \tag{20}$$

For convenience, define $\gamma_N = \frac{\gamma}{N}$ as the scaled risk aversion. The true expected excess returns under the data-generating process are then:

$$\text{Expected Returns} = \mu_t \equiv \mathbb{E}_t[d_{t+1}] - (1+r_f)p_t \tag{21}$$

$$= a_t - (1+r_f)p_t = (a_t - \hat{a}_t) + \gamma_N\hat{\Lambda}_t\iota \tag{22}$$

This expression can be decomposed into three components:

$$\mu_t = \underbrace{(a_t - \hat{a}_t)}_{\text{mean estimation error}} + \underbrace{\gamma_N(\hat{\Lambda}_t - \Lambda_t)\iota}_{\text{covariance estimation error}} + \underbrace{\gamma_N\Lambda_t\iota}_{\text{risk compensation}} \tag{23}$$

Economic Intuition Each component has a distinct economic interpretation:

- Mean estimation error $(a_t - \hat{a}_t)$: When agents underestimate expected dividends for certain

10. Given the CARA utility structure with normally distributed returns, price taking, and absence of transaction costs, the demand function is the same if we consider single-period agents that observe the entire history of dividends or infinitely lived agents.

assets, those assets become underpriced, offering higher expected returns to compensate.

- Covariance estimation error $\gamma_N(\hat{\Lambda}_t - \Lambda_t)\iota$: When agents overestimate risk (variance), they demand excessive risk premia, causing assets to be underpriced relative to their true risk. This term enters the risk compensation because agents' portfolio choices depend on their perceived covariance $\hat{\Lambda}_t$, but the actual returns they earn depend on the true covariance Λ_t.

- Risk compensation $\gamma_N \Lambda_t \iota$: The fundamental premium for bearing systematic risk that would exist even under perfect information, scaled by risk aversion and per-capita asset supply.

The covariance estimation error affects expected returns through the risk premium channel: agents price assets based on perceived risk $\hat{\Lambda}_t$, but the objective expected returns must compensate for both the true risk and the pricing errors induced by incorrect risk perceptions.

In contrast, agents' subjective expected returns are simpler: $\hat{\mu}_t = \hat{a}_t - (1 + r_f)p_t = \gamma_N \hat{\Lambda}_t \iota$. This reflects their perceived risk premia, driven solely by their estimated covariance.

If taken as a primitive, the demand function in equation 19 can accommodate various estimation methods beyond the Bayesian approach detailed in the previous section. Alternative preference structures or belief formation mechanisms, including regression-based forecasts or sophisticated machine-learning algorithms, can be viewed as different ways to generate $\hat{\Lambda}_t$ and $\hat{a}_t$. Regardless of how beliefs are formed, equilibrium prices and expected returns reflect both estimation errors and risk premia, as long as agents act as mean-variance optimizers using their subjective estimates of means and covariances.

2.5 Objective Optimal Portfolio

In equilibrium, market clearing requires that agents are satisfied holding the market portfolio given their subjective beliefs. That is, prices adjust until agents' optimal portfolios (based on their subjective estimates of means and covariances) align with the available supply. Hence, the market is subjectively optimal. However, the market portfolio will generally be different from the optimal portfolio under complete knowledge of the economy's parameters.

Given the vector of expected returns in equilibrium, the weights of this objective optimal portfolio are given by

$$w_t^* = \frac{1}{\gamma}\Lambda_t^{-1}\mu_t = \underbrace{\frac{1}{\gamma}\Lambda_t^{-1}(a_t - \hat{a}_t)}_{\text{errors in means}} + \underbrace{\frac{1}{N}\Lambda_t^{-1}(\hat{\Lambda}_t - \Lambda_t)\iota}_{\text{errors in covariances}} + \underbrace{\frac{1}{N}\iota}_{\text{market portfolio}} \tag{24}$$

Given equilibrium prices formed under imperfect beliefs, the optimal portfolio exploits these estimation errors arising from both the covariance and expected returns while also having a market portfolio component exploiting the positive risk premium.

Remark 1 (SDF Composition)**.** *In the CARA-normal economy, the minimum-variance stochastic discount factor (SDF) is affine in the return on the tangency portfolio: $m_t^{mv*} \propto 1 - w_t^{*\prime} r_{t+1}$. The decomposition above therefore implies that the SDF loads on both estimation-error components and the market return. The squared Sharpe ratio of the market component is $SR(w_m)^2 = O(1)$, while the total squared Sharpe ratio satisfies $SR(w^*)^2 = SR(w_m)^2 + \alpha'\Lambda_t^{-1}\alpha$, where $\alpha'\Lambda_t^{-1}\alpha = O(N/t)$ by Propositions 3–5. As N/t grows, the fraction of SDF variance attributable to estimation errors, $\alpha'\Lambda_t^{-1}\alpha / SR(w^*)^2$, approaches one. The SDF is dominated by agents' estimation errors, not by fundamental risk factors.*

2.6 Pricing Errors

Let $\sigma_{m,t}^2 = \iota'\Lambda_t\iota$ denote the market variance and $\hat{\sigma}_{m,t}^2 = \iota'\hat{\Lambda}_t\iota$ the perceived variance. Further, let $\varepsilon = a_t - \hat{a}_t$ denote the mean estimation error. Unlike expected returns, market variance depends only on dividends (exogenous), not on equilibrium prices, since all investors hold identical portfolio shares.

To measure the magnitude of pricing errors, we define the vector of pricing errors (alphas) relative to the market portfolio as the deviations of expected returns from their predictions under a market model. For the market portfolio with weights $w_m = \frac{1}{N}\iota$, the market beta for asset i is $\beta_i = \frac{(\Lambda_t\iota)_i}{\iota'\Lambda_t\iota/N} = \frac{N(\Lambda_t\iota)_i}{\iota'\Lambda_t\iota}$, and the market's expected return is $\mu_{m,t} = \frac{\iota'\mu_t}{N}$. The vector of pricing errors

is then:

$$\alpha = \mu_t - \beta\mu_{m,t} = \mu_t - \frac{\Lambda_t\iota \cdot \iota'\mu_t}{\iota'\Lambda_t\iota} \tag{25}$$

where $\beta = \frac{N\Lambda_t\iota}{\iota'\Lambda_t\iota}$ is the vector of market betas. This represents the component of expected returns that cannot be explained by assets' covariances with the market portfolio.

Proposition 1 (Decomposition of Pricing Errors)**.** *The weighted sum of squared pricing errors in the economy, relative to the market portfolio, is given by the difference in squared Sharpe ratios (SR) of the optimal portfolio and the market portfolio. This sum can be decomposed into mean and covariance estimation errors plus an interaction term.*

$$\begin{aligned}
\alpha'\Lambda_t^{-1}\alpha &= SR(w^*)^2 - SR(\tfrac{1}{N}\iota)^2 \\
&= \underbrace{\left[\varepsilon'\,\Lambda_t^{-1}\,\varepsilon - \frac{(\iota'\varepsilon)^2}{\sigma_{m,t}^2}\right]}_{\textit{Mean estimation errors}} \\
&\quad + \underbrace{2\gamma_N\left[\varepsilon'\Lambda_t^{-1}\hat{\Lambda}_t\iota - \frac{(\iota'\varepsilon)\hat{\sigma}_{m,t}^2}{\sigma_{m,t}^2}\right]}_{\textit{Interaction term}} \\
&\quad + \underbrace{\gamma_N^2\left[\iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota - \frac{(\hat{\sigma}_{m,t}^2)^2}{\sigma_{m,t}^2}\right]}_{\textit{Covariance estimation errors}}
\end{aligned} \tag{26}$$

This decomposition builds on the insight from Gibbons, Ross, and Shanken (1989) that pricing errors can be expressed as differences in squared Sharpe ratios, but extends their framework to the equilibrium case where agents have uncertainty about means and covariances. The first term captures errors in estimating means, comparing how the correct weighting exploits mean estimation errors versus market weights. The second term represents the interaction between mean estimation errors and differences in perceived versus true covariance weighting. The third term isolates the pure covariance pricing-error effect, measuring how differences between estimated and true covariances affect the Sharpe ratio.

The market portfolio is the correct factor model from agents' perspective; they perceive it as mean-variance efficient given their beliefs. An econometrician analyzing historical returns would

observe the market portfolio is not truly mean-variance efficient under the objective data-generating process. The econometrician would detect apparent pricing errors and might conclude additional factors beyond the market are needed. These "factors" capture agents' estimation errors.

Proposition 1 decomposes pricing errors into mean estimation, covariance estimation, and interaction components. We now analyze how these scale under different informational assumptions: Proposition 2 establishes the perfect-information benchmark (zero errors), Propositions 3 and 4 isolate covariance and mean estimation errors respectively, and Proposition 5 combines both in our main result.

2.7 Perfect Information

As a benchmark, we first consider the case of perfect information, where agents know the true mean π and covariance V of the state variables. This case serves as the baseline: any deviation from perfect information will generate pricing errors relative to this frictionless ideal.

Proposition 2 (Perfect Information)**.** *When agents know the economy's parameters, CAPM and Arbitrage Pricing Theory (APT) hold. In particular,*

1. *Mean returns are given by* $\mu_t = \gamma_N \Lambda_t \iota$.

2. *The mean-variance efficient portfolio is the market portfolio with weights of* $\frac{1}{N}$.

3. *Pricing errors with respect to the market portfolio are zero,* $\alpha_t = 0$.

4. *For* $N \to \infty$*, and rotating the factors to make them orthogonal, the price of risk for factor* k *is given by* $\gamma \bar{b}_{t,k} \sigma^2_{t,k}$*, where* $\bar{b}_{t,k}$ *is the average loading on factor* k *and* $\sigma_{t,k}$ *is the factor's volatility.*

5. *Systematic factors asymptotically span the maximum Sharpe ratio:* $SR_S^2 / SR_{all}^2 \to 1$ *as* $N \to \infty$*, where* $SR_{all}^2 = \mu_t' \Lambda_t^{-1} \mu_t$ *is the maximum squared Sharpe ratio using all* N *assets and* SR_S^2 *uses only* S *factor-mimicking portfolios. This requires that the systematic factor*

loadings $B^{sys} = [b_{1,t}, \ldots, b_{N,t}]'$ *(the* $N \times S$ *submatrix of* B_t*) satisfy the regularity condition* $\frac{1}{N}(B^{sys})'B^{sys} \to G$ *for some positive definite matrix* G *as* $N \to \infty$.

The intuition for CAPM: mean-variance optimizers choose only optimal portfolios. Hence, prices adjust so that agents are satisfied with holding the market portfolio by making the market portfolio mean-variance efficient.

For the APT, consider a setting with uncorrelated normal factors[11]. Further, assume N going to infinity, so that idiosyncratic risk is completely diversifiable. For asset i, the risk compensation simplifies to:

$$(\gamma_N \Lambda_t \iota)_i = \gamma \sum_{k=1}^{S} b_{i,t,k} \bar{b}_{t,k} \sigma^2_{t,k}, \tag{27}$$

Hence, APT holds with a price of risk given by:

$$\gamma \bar{b}_{t,k} \sigma^2_{t,k}, \tag{28}$$

where $\bar{b}_{t,k} = \frac{1}{N} \sum_{j=1}^{N} b_{j,t,k}$ is the average loading (or beta) on factor k, $b_{i,t,k}$ is asset i's loading on factor k, and $\sigma_{t,k}$ is the factor's volatility.

For instance, a pure-play portfolio with a beta of 1 on factor m and zero on all others has an expected excess return of:

$$\gamma \bar{b}_{t,m} \sigma^2_{t,m}, \tag{29}$$

where $\gamma \bar{b}_{t,m} \sigma^2_{t,m}$ is the price of risk for factor m. This term quantifies the compensation per unit of exposure to factor m, determined by the market's average loading, factor volatility, and scaled risk aversion. Moreover, a factor's exposure is only priced if its average beta across the economy is non-zero. When the average loading is zero, positive and negative exposures across assets cancel out in the aggregate, leaving no systematic risk for the market to price.

For example, consider decompositions such as Kozak, Nagel, and Santosh (2018), where the first factor is an equal-weighted portfolio (i.e., $q_1 = \sqrt{\frac{1}{N}}\iota$) and the rest are long-short portfolios

11. Achievable via orthogonal rotation, albeit the interpretation of betas would change.

orthogonal to it ($q_k'\iota = 0$ for $k > 1$). Factors with $\bar{b}_{t,k} = 0$, akin to their long-short PCs, should command no premium, as they lack systematic impact. In the case of uncorrelated factors, the optimal weights $w_f = Cov(F)^{-1}E(f)$ should have zero weight on all the factors but the first. This result explains why in Lopez-Lira and Roussanov (2022), it is optimal to hedge most factors. If the factor betas are net zero across the economy, then the Sharpe ratio will be increasing in the number of hedged factors.

2.8 Known Means and Unknown Covariances

The case where agents know the true expected value of the state variables π (and hence expected dividends a_t) but must estimate the covariance matrix V isolates the effects of covariance estimation errors.

Proposition 3 (Known Means, Unknown Covariances)**.** *When agents know the expected value of dividends, a, but have uncertainty about the covariance, Λ_t, neither CAPM nor APT hold. Further, under standard factor structure ($\sigma_m^2 = \iota'\Lambda_t\iota \sim O(N^2)$), expected pricing errors scale as $O(N/t)$.*[12] *In particular:*

1. *Mean returns are given by $\mu_t = \gamma_N(\hat{\Lambda}_t - \Lambda_t)\iota + \gamma_N\Lambda_t\iota$, and hence APT fails.*
2. *CAPM fails, and the mean-variance efficient portfolio weights are given by $\frac{1}{N}\Lambda_t^{-1}(\hat{\Lambda}_t - \Lambda_t)\iota + \frac{1}{N}\iota$.*
3. *Pricing errors are given by $\alpha'\Lambda_t^{-1}\alpha = (\gamma_N)^2\left[\iota'\,\hat{\Lambda}_t\,\Lambda_t^{-1}\,\hat{\Lambda}_t\,\iota \;-\; \frac{(\hat{\sigma}^2_{m,t})^2}{\sigma^2_{m,t}}\right]$.*
4. *The expected pricing errors (where $\mathbb{E}$ is over the distribution of observed data under the true DGP) satisfy the following lower bound, assuming $\nu_t = \nu_0 + t - (S+N) + 1 > 2$:*

$$\mathbb{E}[\alpha_t'\Lambda_t^{-1}\alpha_t] \geq \frac{\gamma^2 t(N-1)}{N^2(\nu_t - 2)^2}(\iota'\Lambda_t\iota) \geq c\,\frac{N}{t} \tag{30}$$

12. Throughout, $f = O(g)$ means $f \leq c\,g$ for a positive constant c; "scales as N/t" means $c_1\,N/t \leq f \leq c_2\,N/t$ for positive constants c_1, c_2; $f \sim O(g)$ means f scales as g.

where $\iota'\Lambda_t\iota = \sigma^2_{m,t}$ is the market variance and $c > 0$ is a constant. Under standard factor structures where assets share common exposures, $\sigma^2_{m,t}$ scales as N^2 because the N^2 pairwise covariances sum constructively, yielding pricing errors that scale as N/t. Equality holds if and only if $B_tV_0B_t'\iota \propto \Lambda_t\iota$ (which holds in particular when $V_0 = V$). The tightest bound obtains with minimal prior degrees of freedom ($\nu_0 = S + N$, giving $\nu_t - 2 = t - 1$).

Covariance estimation uncertainty creates a wedge between what investors achieve (the market portfolio) and the objective optimum. As N grows with limited data, the $N \times N$ covariance matrix requires $O(N^2)$ parameters from the same t observations. An investor who knew the true Λ_t could construct portfolios exploiting pricing errors that grow linearly with N. The market portfolio's Sharpe ratio remains bounded, reflecting investors' conservative positions based on uncertain estimates. The gap between feasible and optimal Sharpe ratios is the pricing error that scales with market complexity.

2.9 Unknown Means and Known Covariances

The complementary case where agents know the true covariance matrix V but must estimate the mean vector π isolates the effects of mean estimation errors. Estimation uncertainty in expected returns alone generates systematic pricing errors even when agents have perfect knowledge of the covariance structure.

Proposition 4 (Unknown Means, Known Covariances)**.** *If investors have perfect information about covariances but only the correct prior about the expectation, neither CAPM nor APT hold. In particular:*

1. *Mean returns are given by $\mu_t = (a_t - \hat{a}_t) + \gamma_N\Lambda_t\iota$, and hence APT fails.*
2. *CAPM fails, and the mean-variance efficient portfolio weights are given by $\frac{1}{\gamma}\Lambda_t^{-1}(a_t - \hat{a}_t) + \frac{1}{N}\iota$.*
3. *Pricing errors are given by $\alpha'\Lambda_t^{-1}\alpha = \varepsilon'\Lambda_t^{-1}\varepsilon - \frac{1}{\sigma^2_m}(\iota'\varepsilon)^2$, where $\varepsilon = a_t - \hat{a}_t$.*

Moreover, the average pricing errors relative to the market portfolio increase linearly with the number of assets N and decrease with the number of observations t:

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] \geq \frac{t(N-1)}{(t+\kappa_0)^2} \sim O\left(\frac{N}{t}\right). \tag{31}$$

This bound holds exactly for the case with a correct prior ($\pi_0 = \pi$) and serves as a lower bound when the prior is incorrect ($\pi_0 \neq \pi$).

Extension: Time-Varying State Means Proposition 4 assumes the state mean π is constant, so estimation errors vanish as $O(N/t)$ with accumulating data. When fundamentals evolve over time, this learning benefit disappears. Time-varying state means generate persistent $O(N)$ pricing errors.

Suppose the state mean follows an AR(1) process:

$$\pi_{t+1} = \rho \pi_t + \eta_{t+1}, \quad \eta_{t+1} \sim N(0, \Sigma_\eta), \tag{32}$$

where $\rho \in [0, 1)$ is the persistence parameter and agents observe $X_t = \pi_t + \varepsilon_t^{obs}$ with $\varepsilon_t^{obs} \sim N(0, V)$. Maintaining the assumption that V is known (as in Proposition 4), the optimal estimator is the Kalman filter.

Corollary 1 (Time-Varying Means: Persistent Pricing Errors)**.** *Under the AR(1) specification (32) with known V and scalar proportionality ($\Sigma_\eta = \sigma_\eta^2 V$ for some $\sigma_\eta^2 > 0$):*

1. *The Kalman filter posterior variance converges to a steady state: $P_t \to P_\infty = \sigma_p^2 V \succ 0$, where σ_p^2 solves the scalar Riccati equation*

$$\rho^2 (\sigma_p^2)^2 + (1 - \rho^2 + \sigma_\eta^2)\sigma_p^2 - \sigma_\eta^2 = 0. \tag{33}$$

2. *Expected pricing errors converge to a positive constant that scales linearly with N:*

$$\lim_{t \to \infty} \mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] = \kappa_\infty (N-1), \tag{34}$$

where $\kappa_\infty := \rho^2\sigma_p^2 > 0$.

The contrast with Proposition 4:

$$\begin{cases} \text{Static } \pi: & \mathbb{E}[\alpha'\Lambda^{-1}\alpha] \sim O(N/t) \to 0 \\ \text{Time-varying } \pi_t: & \mathbb{E}[\alpha'\Lambda^{-1}\alpha] \to \kappa_\infty N > 0 \end{cases} \tag{35}$$

Even with optimal Bayesian learning (Kalman filtering) and infinite data, pricing errors persist because π_t shifts each period. The fresh innovations η_{t+1} have not appeared in any observed data at time t, making the resulting pricing errors irreducible regardless of sample size or estimation method. This corollary provides the equilibrium foundation for the persistent $O(N)$ component in the Undiversifiable Uncertainty Theorem (Section 1). The scalar proportionality assumption simplifies the Kalman filter analysis; Appendix 6.11 establishes persistent $O(N)$ scaling under weaker conditions that do not require proportionality.

Propositions 4 (with Corollary 1 showing persistence under time-varying fundamentals) and 3 isolated mean and covariance estimation errors separately. Proposition 5 combines both.

2.10 Unknown Means and Covariances

In the main case, agents must estimate both π and V simultaneously. The joint estimation case forms the foundation for the baseline homogeneous-agent economy. Each source of estimation uncertainty independently generates $O(N/t)$ pricing errors (Propositions 3 and 4). When both parameters must be estimated jointly, these errors combine to preserve the same complexity-driven scaling, though the interaction introduces additional structure.

Under correctly specified priors ($\pi_0 = \pi$), the interaction term between mean and covariance estimation errors vanishes exactly ($\mathbb{E}[B] = 0$ by a sign-symmetry argument; see proof of Proposition

5), so the expected pricing errors decompose as:

$$E[\alpha' \Lambda_t^{-1} \alpha] = \underbrace{E\left[\varepsilon' \Lambda_t^{-1} \varepsilon - \frac{(\iota' \varepsilon)^2}{\sigma_m^2}\right]}_{\alpha \text{ from unknown means}} + \underbrace{E\left[\gamma_N^2 \left(\iota' \hat{\Lambda}_t \Lambda_t^{-1} \hat{\Lambda}_t \iota - \frac{(\hat{\sigma}_m^2)^2}{\sigma_m^2}\right)\right]}_{\alpha \text{ from unknown covariances}} \tag{36}$$

Under misspecified priors ($\pi_0 \neq \pi$), the interaction term is $O(N/t^2)$ and the decomposition holds to first order. Both components scale as $O(N/t)$.

Proposition 5 (Pricing Errors with Unknown Means and Covariances)**.** *When agents estimate both π and V using the Bayesian framework (Equations 5-10), neither CAPM nor APT hold. Estimating both mean and covariance leads to pricing errors scaling with market complexity. In particular:*

1. *Mean returns are given by the full decomposition including mean and covariance estimation errors (Equation 23):*

$$\mu_t = (a_t - \hat{a}_t) + \gamma_N (\hat{\Lambda}_t - \Lambda_t)\iota + \gamma_N \Lambda_t \iota.$$

   Hence, APT fails as returns depend on estimation errors.

2. *CAPM fails, and the objectively optimal mean-variance efficient portfolio weights deviate from the market portfolio due to both estimation errors (Equation 24):*

$$w_t^* = \frac{1}{\gamma} \Lambda_t^{-1} (a_t - \hat{a}_t) + \frac{1}{N} \Lambda_t^{-1} (\hat{\Lambda}_t - \Lambda_t)\iota + \frac{1}{N}\iota.$$

3. *Pricing errors relative to the market portfolio are given by the full decomposition including mean, covariance, and interaction terms (Equation 26):*

$$\alpha' \Lambda_t^{-1} \alpha = SR(w^*)^2 - SR(\frac{1}{N}\iota)^2.$$

4. *Under correctly specified priors ($\pi_0 = \pi$, $V_0 = V$) with bounded strength ($\kappa_0 \leq K$) and standard factor structure ($\sigma_m^2 = \iota' \Lambda \iota$ of order N^2), the expected weighted squared pricing*

errors scale as N/t:

$$c_1 \frac{N}{t} \;\leq\; \mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] \;\leq\; c_2 \frac{N}{t} \tag{37}$$

for constants $c_1, c_2 > 0$, establishing N/t as a fundamental benchmark that cannot be improved through better prior specification within this class.

The proof involves analyzing the three components of the pricing error decomposition (Equation 26). The mean estimation error term (A) and covariance estimation error term (C) are both non-negative and scale as N/t. The interaction term (B) satisfies $|\mathbb{E}[B]| \leq 2\sqrt{\mathbb{E}[A]\mathbb{E}[C]} = O(N/t)$, and vanishes exactly ($\mathbb{E}[B] = 0$) when the prior mean is correctly centered ($\pi_0 = \pi$) due to the symmetry properties of the normal distribution. Under correctly specified priors, since $\mathbb{E}[B] = 0$ and $C \geq 0$, we have $\mathbb{E}[A + B + C] \geq \mathbb{E}[A] \geq c\, N/t$, establishing the lower bound. Combined with the $O(N/t)$ upper bound, this gives $\sim O(N/t)$ scaling for the best-case scenario. Even with perfectly centered priors and optimal Bayesian learning, expected pricing errors grow linearly with market size N and decay only as $1/t$. Under regularity conditions (bounded loadings, bounded eigenvalues, bounded prior parameters; see Appendix), the $\sim O(N/t)$ rate persists for misspecified priors once t exceeds a fixed threshold, so no prior specification within this class overcomes the curse of dimensionality. A lower bound of order N/t holds for any estimator (Appendix 6.11), including shrinkage and machine learning methods, and worsens to order N if parameters are time-varying rather than fixed (see Corollary 1).

Remark (Why Complexity Scales with N, Not S) Although investors estimate $(S + N)$-dimensional parameters (the mean vector π and covariance matrix V of the state variables), pricing errors collapse to the N-dimensional asset return space where pricing occurs. The loading matrix B_t (of size $N \times (S + N)$) projects the estimation uncertainty from the latent state space to observable returns: state means π map to asset expected returns $a_t = B_t \pi$ (N-dimensional), and the state covariance V maps to the return covariance matrix $\Lambda_t = B_t V B_t'$ ($N \times N$). A central result (see Appendix) is the trace identity $\text{trace}(B_t' \Lambda_t^{-1} B_t V) = \text{trace}(\Lambda_t^{-1} \Lambda_t) = N$, which shows that the effective complexity for pricing, despite the higher-dimensional parameter space, collapses to the

cross-sectional dimension N of observable assets. Economically, investors care about estimating N expected returns and $N \times N$ return covariances in the asset space; this is what determines portfolio weights and equilibrium prices. The number of systematic factors S governs the structure of these return moments, but the cross-sectional dimension N determines the magnitude of pricing errors because it measures the number of distinct securities that must be priced. This explains why Propositions 3–5 establish $O(N/t)$ scaling: complexity grows with the number of observable assets, not with the latent factor dimension.[13]

Remark (Structured Unknown Loadings) For analytical tractability, our main analysis assumes investors perfectly observe the factor loading matrix B_t. A natural extension relaxes this assumption to allow loadings to be determined by observable firm characteristics: $b_{i,t} = \Theta' z_{i,t}$, where $z_{i,t}$ is a K-vector of characteristics (e.g., size, book-to-market) and Θ is a $K \times S$ coefficient matrix that must be estimated from data. This introduces KS additional parameters beyond the state parameters (π, V). Under standard cross-sectional normalization of characteristics (ensuring the information matrix grows at rate Nt), each parameter has posterior variance declining at rate $O(1/(Nt))$ under Bayesian updating. Total pricing errors then scale as $O(N/t + KS/(Nt))$ compared to our baseline $O(N/t)$. In fixed-N asymptotics (classical time-series setting), this simplifies to $O((N + KS)/t)$; for large cross-sections with K proportional to N, the mean estimation term $O(N/t)$ dominates but with an effective complexity multiplier. This confirms that the baseline $O(N/t)$ isolates asset-space complexity and provides a conservative lower bound. The Internet Appendix formalizes this extension through Bayesian regression for Θ given observable state realizations.

3 Heterogeneous Beliefs

In the homogeneous agent baseline (Section 1), all investors share the same information and update their beliefs identically from observed data. Why do pricing errors matter if nobody has

13. This dimensional reduction relies on B_t being known. With structured unknown loadings $b_{i,t} = \Theta' z_{i,t}$, the KS parameters in Θ directly determine asset returns and do not benefit from this projection-based compression, yielding $O(N/t + KS/(Nt))$ complexity in general, or $O((N + KS)/t)$ in fixed-N asymptotics, as noted in the next remark.

perfect information? Two reasons. First, an econometrician testing asset pricing models ex-post will find large pricing errors relative to standard factor models. Specifically, while agents at time t make optimal decisions given their beliefs, an econometrician analyzing the realized returns at time $t+1$ finds the market portfolio was not mean-variance efficient under the objective data-generating process. These apparent anomalies arise not from irrationality but from estimating high-dimensional parameters with limited data. Second, in real markets, there may be a subset of agents that have partial or complete information about the data-generating process, yet mispricing persists.

Mispricing persists in equilibrium even when a positive fraction of investors knows the true data-generating process. To formalize this persistence, the model extends to heterogeneous beliefs, where some investors are endowed with knowledge of π as a modeling device representing the limit of extensive experience or superior data. Complexity-driven pricing errors persist even when some investors know the true parameters (Proposition 6). The extension also generates predictions about the proliferation and correlation structure of profitable trading strategies, connecting our theory to the factor zoo and hedge fund evidence. In the data, (1) many uncorrelated anomaly portfolios generate alpha, and (2) specialized quantitative funds exploit distinct inefficiencies with limited overlap.

Analytical Scope: Known Covariances For analytical tractability, we analyze heterogeneous beliefs in the setting where all investors know the true covariance matrix V but differ in their beliefs about the mean vector π, paralleling Proposition 4's informational environment. This restriction simplifies the equilibrium analysis considerably: when investors differ only in their mean estimates, their perceived covariance matrices $\hat{\Lambda}_{i,t}$ are scalar multiples of the true covariance Λ_t, allowing demands to aggregate cleanly. In contrast, if investors also held heterogeneous beliefs about V, they would perceive different covariance structures (not merely different scales), making analytical solutions intractable without additional restrictions.[14]

14. This restriction preserves the main results. The fundamental mechanism (pricing errors persist because risk aversion limits arbitrage) operates through mean disagreement. Adding covariance uncertainty (as in Proposition 5) would contribute additional $O(N/t)$ errors but preserve the $O(N/t)$ scaling and qualitative predictions. Covariance

Notation for Heterogeneous Setting We introduce investor-specific subscripts to distinguish individual beliefs from aggregate quantities. Variables with subscript i (e.g., $\hat{\pi}_{i,t}$, $\hat{a}_{i,t}$) denote investor i's subjective beliefs, while aggregate or economy-wide quantities use subscript E (e.g., $\hat{a}_{E,t}$ for the weighted-average expected dividend). This extends our earlier notation where, in the homogeneous case, $\hat{\pi}_t$ and $\hat{a}_t$ represented the single representative agent's beliefs.

We maintain the following standard assumptions:

Assumption 1 (H.1: Agent Behavior)**.** *Investors are price takers (treat prices p_t as exogenous), agree to disagree (do not update beliefs based on others' actions), and do not learn from prices (do not update beliefs about parameters like π from p_t).*

The agree-to-disagree assumption is standard in models with heterogeneous beliefs (Harrison and Kreps 1978; Scheinkman and Xiong 2003) and serves two roles here. First, it prevents the no-trade theorem: if agents updated from prices, rational agents with common priors and common knowledge of rationality would infer others' information and converge to identical beliefs, eliminating trade and making heterogeneous equilibria trivial. Second, in our setting, learning from prices is informationally demanding because the mapping from π to p_t depends on the wealth-weighted average of all agents' beliefs, which no individual agent observes. We relax the agree-to-disagree assumption in Section 3.2, where a positive mass of sophisticated investors infers aggregate beliefs from equilibrium prices and trades optimally against the resulting mispricing. The $O(N/t)$ scaling of pricing errors persists in that rational expectations extension, with an attenuation factor that depends on the effective mass of informed capital.

3.1 Two-Type Economy: Informed and Uninformed Investors

We begin with a two-type economy that captures the main economic forces. Let $\theta \in (0, 1)$ denote the population share of informed investors who know the true mean π. The remaining share $(1 - \theta)$ consists of uninformed investors who must estimate π from data using Bayesian updating.

uncertainty affects both error magnitude and effective investor weights through parameter uncertainty factors $\varphi_{i,t}$, but the $O(N/t)$ scaling remains robust.

Informed Investors (Type I) These investors know π with certainty. In our framework, this corresponds to taking $\kappa_{0,I} \to \infty$ (infinite prior precision) with $\pi_{0,I} = \pi$. Consequently:

- Their forecast equals the truth: $\hat{a}_{I,t} = B_t \pi = a_t$
- No parameter uncertainty: $\varphi_{I,t} = 1$
- Effective belief weight equals population share: $\vartheta_{I,t} = \theta$

Uninformed Investors (Type U) These investors have prior $\pi \sim N(\pi_{0,U}, V/\kappa_{0,U})$ with finite precision $\kappa_{0,U}$. After observing t periods with sample mean $\bar{X}_t$, their posterior mean is:

$$\hat{\pi}_{U,t} = \frac{\kappa_{0,U}}{\kappa_{0,U} + t} \pi_{0,U} + \frac{t}{\kappa_{0,U} + t} \bar{X}_t. \tag{38}$$

Their parameter uncertainty inflates perceived risk:

$$\varphi_{U,t} = \frac{\kappa_{0,U} + t + 1}{\kappa_{0,U} + t} > 1, \tag{39}$$

leading to effective belief weight $\vartheta_{U,t} = (1 - \theta)/\varphi_{U,t} < (1 - \theta)$.

Equilibrium Each investor type's demand follows mean-variance optimization with perceived covariance $\hat{\Lambda}_{i,t} = \varphi_{i,t} \Lambda_t$:

$$x_{i,t} = \frac{1}{\gamma \varphi_{i,t}} \Lambda_t^{-1} \left(\hat{a}_{i,t} - (1 + r_f) p_t \right). \tag{40}$$

Market clearing ($\theta x_{I,t} + (1 - \theta) x_{U,t} = \iota/N$) yields the equilibrium price:

$$p_t = \frac{1}{1 + r_f} \left(\hat{a}_{E,t} - \tilde{\gamma}_{N,t} \Lambda_t \iota \right), \tag{41}$$

where $\Omega_t = \theta + (1-\theta)/\varphi_{U,t}$ is the total effective weight, $\tilde{\gamma}_{N,t} = \gamma_N/\Omega_t$ is the aggregate risk aversion, and the economy's aggregate expectation is:

$$\hat{a}_{E,t} = \frac{\theta \cdot a_t + \vartheta_{U,t} \cdot \hat{a}_{U,t}}{\Omega_t}. \tag{42}$$

Why Mispricing Persists The aggregate estimation error is:

$$\tilde{\varepsilon}_t \triangleq a_t - \hat{a}_{E,t} = \frac{\vartheta_{U,t}}{\Omega_t}(a_t - \hat{a}_{U,t}) = \frac{\vartheta_{U,t}}{\Omega_t} B_t(\pi - \hat{\pi}_{U,t}). \tag{43}$$

Even though informed investors know the truth, the aggregate belief $\hat{a}_{E,t}$ deviates from a_t because uninformed investors have positive effective weight $\vartheta_{U,t} > 0$. Informed investors cannot fully eliminate this estimation error: their optimal response is to tilt portfolios toward underpriced assets, but risk aversion limits position sizes. The aggregate error inherits the uninformed investors' estimation error, scaled by their relative weight.

Since $\pi - \hat{\pi}_{U,t}$ has the same $O(N/t)$ variance structure as in the homogeneous case (Proposition 4), pricing errors persist at rate $O(N/t)$ with magnitude proportional to $(1-\theta)^2$.

General Case: I Types The two-type analysis extends immediately to I types with population shares θ_i and priors $(\pi_{0,i}, \kappa_{0,i})$. Each type has effective weight $\vartheta_{i,t} = \theta_i/\varphi_{i,t}$, the aggregate weight is $\Omega_t = \sum_i \vartheta_{i,t}$, and the aggregate expectation is $\hat{a}_{E,t} = \Omega_t^{-1} \sum_i \vartheta_{i,t}\hat{a}_{i,t}$. The equilibrium price takes the same form (41), and pricing errors persist as long as any type has finite prior precision.

Proposition 6 (Persistence of Complexity-Driven Pricing Errors)**.** *In the heterogeneous economy under Assumption H.1, where agents know the covariance matrix V but estimate the mean π using priors $(\pi_{0,i}, \kappa_{0,i})$, the aggregate mean estimation error $\tilde{\varepsilon}_t = a_t - \hat{a}_{E,t}$ leads to pricing errors relative to the market portfolio. If there is a positive mass $\theta_u > 0$ of uninformed investors (Type u) with finite prior precision $\kappa_{0,u}$, the expected squared pricing errors satisfy a lower bound analogous to*

the homogeneous case (Proposition 4):

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] \equiv \mathbb{E}[\tilde{\varepsilon}_t' \Lambda_t^{-1} \tilde{\varepsilon}_t - \frac{(\iota' \tilde{\varepsilon}_t)^2}{\iota' \Lambda_t \iota}] \geq \frac{\theta_u^2 t(N-1)}{(\kappa_{0,u} + t + 1)^2} \sim O\left(\frac{N}{t}\right), \tag{44}$$

where the constant θ_u^2 depends on the uninformed agents' population share θ_u, and the $(+1)$ in the denominator arises from the parameter uncertainty factor in the Normal-Normal conjugate posterior predictive distribution.

The market's aggregate expectation inherits the same relationship between complexity and estimation error as in the homogeneous case, provided some agents are learning.

Corollary 2 (Two-Type Economy)**.** *In the two-type economy with informed share θ and uninformed share $(1-\theta)$, Proposition 6 yields:*

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] \geq \frac{(1-\theta)^2 t(N-1)}{(\kappa_{0,U} + t + 1)^2} \sim O\left(\frac{N}{t}\right). \tag{45}$$

Remark: Relation to Joint Estimation Proposition 6 addresses the known-V case (paralleling Proposition 4). With heterogeneous covariance beliefs, investors would perceive structurally different covariance matrices (not merely different scales), and the clean demand aggregation used here would not apply. Whether the $O(N/t)$ scaling extends to that setting remains an open question.

3.2 Sophisticated Investors with Partial Information

Proposition 6 established that pricing errors persist in equilibrium when heterogeneous learning agents (some informed about π, others estimating it from data) set prices through market clearing. Market clearing generates an aggregate estimation error $\tilde{\varepsilon}_t = a_t - \hat{a}_{E,t}$ that scales as $O(N/t)$. A distinct group of sophisticated investors exploits these errors, generating uncorrelated high-Sharpe strategies observed in practice.

The model extends Proposition 6's equilibrium by introducing a group of sophisticated, atomistic investors who take prices as given. Unlike the Type 1 informed agents in Proposition 6 (who

know π but participate in price formation and thus cannot fully eliminate mispricing due to risk aversion and limited capital), these sophisticated investors are price takers who observe the realized aggregate mispricing $\tilde{\varepsilon}_t$ on a random fraction of assets. The equilibrium prices and the aggregate error $\tilde{\varepsilon}_t$ continue to be determined by the heterogeneous learning agents from Proposition 6; the sophisticated investors simply exploit these errors without affecting them. Under Proposition 6's persistent complexity-driven errors, sophisticated investors generate profitable, weakly correlated trading strategies consistent with the factor zoo and hedge fund evidence.

Assumption 2 (S.1: Sophisticated Investor Information and Behavior)**.** *Sophisticated investors (i) know the true covariance matrix Λ_t, (ii) receive perfect private signals about the aggregate estimation error $\tilde{\varepsilon}_t$ for a random fraction ϕ of assets via diagonal selector matrix $D_{i,t} = \text{diag}(d_{i1,t}, \ldots, d_{iN,t})$ where each $d_{ij,t} \sim$ Bernoulli(ϕ) independently across assets j, investors i, and time t,*[15] *(iii) are atomistic price takers whose demand does not influence equilibrium prices p_t or the aggregate error $\tilde{\varepsilon}_t$ (which are determined by the heterogeneous learning agents from Proposition 6), and (iv) have CARA utility with risk aversion γ.*

Error Covariance Structure (Consequence of Proposition 6) Under the Bayesian learning framework of Proposition 6, the aggregate estimation error satisfies $\tilde{\varepsilon}_t = c_t B_t(\pi - \bar{X}_t)$ where c_t depends on population shares and prior precisions. Since $\pi - \bar{X}_t \sim N(0, V/t)$ and $\Lambda_t = B_t V B_t'$, the error covariance is proportional to the return covariance:

$$\Lambda_{a,t} = \text{Cov}(\tilde{\varepsilon}_t) = \frac{c_t^2}{t}\Lambda_t \equiv \frac{1}{t + \kappa_E}\Lambda_t, \tag{46}$$

where $\kappa_E \geq 0$ is the effective aggregate prior precision. Two quantities govern the properties of partial-information strategies: the trace term $T \equiv \text{Tr}(\Lambda_t^{-1}\Lambda_{a,t})$ and the diagonal term $\Delta \equiv \sum_{k=1}^{N} [\Lambda_t^{-1}]_{kk}[\Lambda_{a,t}]_{kk}$. Under proportionality, $T = N/(t+\kappa_E)$ and the ratio Δ/T equals the diagonal

15. Knowing $a_{j,t}$ for an asset j, along with the publicly-observed price $p_{j,t}$ and the common risk-premium component of expected returns (related to $\tilde{\gamma}_{N,t}\Lambda_t\iota$), reveals the mispricing $\tilde{\varepsilon}_{j,t}$ for that asset. Thus, observing $(p_{j,t}, a_{j,t})$ for a subset of assets is informationally equivalent to our modeling device $D_{i,t}\tilde{\varepsilon}_t$.

overlap coefficient:

$$\chi_t \equiv \frac{1}{N}\sum_{j=1}^{N}(\Lambda_t^{-1})_{jj}(\Lambda_t)_{jj} \geq 1, \tag{47}$$

where the inequality follows from Cauchy-Schwarz applied to each diagonal element. The ratio $\Delta/T = \chi_t$ governs how the covariance structure of Λ_t affects the interaction between random asset selection and portfolio optimization.

With Assumption S.1 in place, we now characterize the optimal strategies of sophisticated investors and the properties of their returns. Conditional on their information (including $\mathcal{F}_t$ and their private signal $D_{i,t}\tilde{\varepsilon}_t$), these investors perceive an opportunity. Because they employ CARA utility and perceive normally distributed returns (conditional on their information), their optimal dollar demand to exploit their informational advantage regarding $\tilde{\varepsilon}_t$ is proportional to $\Lambda_t^{-1}D_{i,t}\tilde{\varepsilon}_t$.[16]

For analyzing the properties of such strategies, we define a scaled portfolio representing the core signal exploitation strategy for investor i. Incorporating the standard risk aversion factor and scaling positions by $1/N$ to keep typical dollar holdings stable as N grows, we define the portfolio weights vector as:

$$w_{i,t} = \frac{1}{\gamma N}\Lambda_t^{-1}D_{i,t}\tilde{\varepsilon}_t. \tag{48}$$

Here $w_{i,t}$ represents an individual sophisticated investor's strategy weights (dollar amounts), distinct from the objective optimal portfolio w_t^* or the belief weights $\vartheta_{i,t}$ defined earlier. We analyze the statistical properties of the dollar returns $R_{w_{i,t}}$ generated by holding this portfolio.

The aggregate estimation error $\tilde{\varepsilon}_t$ follows a normal distribution, conditional on $\mathcal{F}_t$, reflecting the residual uncertainty given the market's aggregate information:

$$\tilde{\varepsilon}_t \mid \mathcal{F}_t \sim N(0, \Lambda_{a,t}), \tag{49}$$

where $\Lambda_{a,t} = \frac{1}{t+\kappa_E}\Lambda_t$ by the proportionality established above.

16. These sophisticated investors are atomistic and their demand does not affect the equilibrium price p_t or the aggregate belief $\hat{a}_{E,t}$, which continue to be determined by the heterogeneous learning agents from Proposition 6. Their demand reflects exploiting the mispricing $\tilde{\varepsilon}_t$ they observe through their partial lens $D_{i,t}$. Because $D_{i,t}\tilde{\varepsilon}_t$ reveals the true mispricing for the selected assets, the signal is noise-free for those names and non-existent for the rest.

We characterize the properties of these partial-information strategies. In general, the expected return is $\mathbb{E}[R_{w_{i,t}}] = \phi T/(\gamma N)$, the expected variance is $[\phi^2 T + \phi(1-\phi)\Delta]/(\gamma N)^2$, and the squared Sharpe ratio is:

$$SR_{i,t}^2 = \frac{\phi T^2}{\phi T + (1-\phi)\Delta}. \tag{50}$$

Under the model's proportionality, $T = N/(t+\kappa_E)$ and $\Delta/T = \chi_t$, giving the expected return $\phi/[\gamma(t+\kappa_E)]$ and the Sharpe ratio $SR_{i,t} = \sqrt{N/(t+\kappa_E)} \cdot \sqrt{\phi}/\sqrt{\phi + (1-\phi)\chi_t} \sim O(\sqrt{N/(t+\kappa_E)})$. As markets become more complex (higher N), investors understanding a fixed fraction of assets (ϕ) can achieve higher risk-adjusted returns, though learning (increasing t) and covariance structure (χ_t) reduce the effect.

For the ex-ante correlation between different investors' strategies (see Appendix):

$$\text{Corr}^{ex\text{-}ante}(R_{w_{i,t}}, R_{w_{j,t}}) = \frac{\phi}{\phi + (1-\phi)\Delta/T}. \tag{51}$$

Under proportionality, $\Delta/T = \chi_t \geq 1$, so correlations are bounded above by ϕ. The correlation depends on the overlap in known assets (ϕ) and the covariance structure (χ_t), but is independent of N, t, and κ_E. Higher χ_t (more correlated assets) reduces correlation between investors by making their partial signals less informationally redundant.

Proposition 7 (Properties of Partial Information Strategies)**.** *Consider a continuum of sophisticated investors who know the true covariance matrix Λ_t and observe the aggregate estimation error $\tilde{\varepsilon}_t$ on a random fraction ϕ of assets, selected by $D_{i,t}$. Assume $\tilde{\varepsilon}_t \mid \mathcal{F}_t \sim N(0, \Lambda_{a,t})$ for a positive definite error covariance $\Lambda_{a,t}$. Investors form portfolios $w_{i,t} = \frac{1}{\gamma N}\Lambda_t^{-1} D_{i,t}\tilde{\varepsilon}_t$. Define $T \equiv Tr(\Lambda_t^{-1}\Lambda_{a,t})$ and $\Delta \equiv \sum_{k=1}^{N}[\Lambda_t^{-1}]_{kk}[\Lambda_{a,t}]_{kk}$. Then, the expected properties of the dollar returns $R_{w_{i,t}}$, averaged over the distribution of $\tilde{\varepsilon}_t$ and the independent random selectors $D_{i,t}$, $D_{j,t}$, are:*

1. *Expected Return:* $\mathbb{E}[R_{w_{i,t}}] = \frac{\phi T}{\gamma N}$.
2. *Expected Variance:* $\mathbb{E}[Var_t(R_{w_{i,t}})] = \frac{\phi^2 T + \phi(1-\phi)\Delta}{(\gamma N)^2}$.

3. *Squared Sharpe Ratio:* $SR^2_{i,t} = \frac{\phi T^2}{\phi T + (1-\phi)\Delta}$.[17]

4. *Ex-ante Correlation:* $Corr^{ex\text{-}ante}(R_{w_{i,t}}, R_{w_{j,t}}) = \frac{\phi}{\phi + (1-\phi)\Delta/T}$ *between two distinct investors* i *and* j*, where the ex-ante correlation is* $\mathbb{E}[Cov]/\sqrt{\mathbb{E}[Var_i] \cdot \mathbb{E}[Var_j]}$.

Under the model's Bayesian learning structure (Proposition 6), $\Lambda_{a,t} \propto \Lambda_t$*, so* $\Delta/T = \chi_t \geq 1$ *and* $Corr^{ex\text{-}ante} \leq \phi$*. The upper bound* ϕ *is achieved when* $\chi_t = 1$ *(diagonal* Λ_t*); there is no positive universal lower bound.*

While individual strategy returns and variances depend on learning time t, effective prior κ_E, market size N, and covariance structure χ_t, the achievable Sharpe ratio increases with market complexity (N) and information advantage (ϕ), offset by learning and covariance structure. The correlation between strategies is bounded above by ϕ and decreases with χ_t: markets with correlated assets ($\chi_t > 1$) support more numerous strategies with lower cross-correlations, enhancing diversification potential.

These theoretical predictions align with our empirical findings about the low correlations between systematic hedge fund strategies and the difficulties parsimonious factor models face. The predictions also yield several testable implications about how strategy relationships should evolve as markets become more complex. As investment universes expand (higher N), funds face a choice between maintaining deep knowledge of a fixed fraction of assets (constant ϕ) or spreading resources more thinly (declining ϕ). Successful funds maintain stable ϕ by scaling research capabilities with N; large quantitative funds invest in research infrastructure. Funds that fail to scale research appropriately should see declining Sharpe ratios relative to competitors (Eq. 50).

The model also makes specific predictions about cross-fund correlation patterns. Correlations between funds depend on research overlap (ϕ) and covariance structure (χ_t), not strategy similarity per se (Eq. 51). Funds using similar techniques but focusing on different market segments (different

17. The Sharpe ratio is defined as $SR^2 = (\mathbb{E}[R])^2/\mathbb{E}[\mathrm{Var}_t(R)]$, using the expected conditional variance. The unconditional variance $\mathrm{Var}(R) = \mathbb{E}[\mathrm{Var}_t(R)] + \mathrm{Var}(\mathbb{E}_t[R])$ is larger, so the reported SR is an upper bound on the unconditional Sharpe ratio.

draws of $D_{i,t}$) should maintain low correlations. The dependence on information overlap rather than strategy similarity explains why even "copycat" strategies can succeed without eroding the profits of the original strategy if they target sufficiently different information subsets. Furthermore, increasing market complexity (N) should lead to more, not fewer, profitable trading strategies (higher potential Sharpe ratios, Eq. 50). Markets with correlated assets ($\chi_t > 1$) enhance diversification by reducing cross-strategy correlations. The "factor zoo" should continue to expand with market complexity, with new factors showing less overlap with existing ones.

The predictions rationalize the proliferation of seemingly unrelated anomalies: the pattern is an outcome of rational investors responding to opportunities that market complexity creates. Propositions 6 and 7 together characterize the equilibrium: the heterogeneous learning agents (Proposition 6) generate persistent mispricing through rational but imperfect estimation, while the sophisticated partial-information investors (Proposition 7) exploit the resulting pricing errors in ways that generate the observed proliferation of weakly correlated profitable strategies. Rather than indicating market inefficiency, multiple successful but uncorrelated strategies reflect the difficulty of spanning high-dimensional markets with limited information.

3.2.1 Robustness: Sophisticated Investors with Positive Mass

The atomistic assumption in Proposition 7 raises a question: do the results persist when sophisticated investors collectively have non-negligible capital and participate in price formation?

Consider an economy with three groups: informed learning agents (mass θ_1, knowing π), uninformed learning agents (mass θ_2, estimating π from data), and sophisticated investors (total mass $\theta_s > 0$, each observing the true a_t on a random fraction ϕ of assets via independent selector $D_{i,t}$). In a rational expectations equilibrium, sophisticated investors infer the aggregate belief $\hat{a}_{E,t}^{(new)}$ from prices. For observed assets they use a_t; for unobserved assets they use $\hat{a}_{E,t}^{(new)}$. Aggregating via the Law of Large Numbers and solving the fixed point yields an attenuation factor $\rho^* = \Omega_t^{(L)}/(\Omega_t^{(L)} + \theta_s\phi) \in (0, 1)$, where $\Omega_t^{(L)}$ is the effective weight of learning agents from Proposition 6. The equilibrium aggregate error is $\tilde{\varepsilon}_t^{(new)} = \rho^* \tilde{\varepsilon}_t^{(L)}$.

Proposition 8 (Sophisticated Investors with Positive Mass)**.** *Under the setup above, with $\theta_2 > 0$ (positive mass of uninformed learning agents):*

1. *Pricing errors persist at rate $O(N/t)$:*

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t]^{(new)} = (\rho^*)^2 \cdot \mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t]^{(L)} \geq \frac{(\rho^*)^2 \theta_2^2 \, t(N-1)}{(\kappa_{0,2} + t + 1)^2} \sim O\left(\frac{N}{t}\right). \tag{52}$$

2. *Ex-ante correlations between sophisticated investors' strategies satisfy:*

$$Corr^{ex\text{-}ante}(R_{w_{i,t}}, R_{w_{j,t}}) = \frac{\phi}{\phi + (1-\phi)\Delta/T} \leq \phi, \tag{53}$$

   independent of θ_s, N, t, or ρ^*.

3. *Squared Sharpe ratios scale as $SR_{i,t}^2 = \phi T^2 / [\phi T + (1-\phi)\Delta] \sim O(N/(t+\kappa_E))$, where T and Δ are defined from the equilibrium error covariance $\Lambda_{a,t}^{(new)} = (\rho^*)^2 \Lambda_{a,t}^{(L)}$.*

The atomistic limit ($\theta_s \to 0$) recovers Proposition 7.

The attenuation depends on effective informed capital $\theta_s \phi$, not total sophisticated capital θ_s. A large fund with shallow coverage has less impact than a smaller fund with deep expertise, rationalizing the coexistence of specialized and generalist quantitative managers. Because sophisticated investors rationally use $\hat{a}_{E,t}^{(new)}$ for unobserved assets, their signals contain no common component across investors; cross-strategy comovement arises solely from overlap in directly observed assets.

The extension generates predictions about how sophisticated capital affects markets. As the quantitative investing industry grows (increasing $\theta_s \phi$), average alphas should decline (lower ρ^*) but remain positive. The number of successful low-correlation strategies should not decline because the correlation bound is independent of θ_s. Competition affects levels of mispricing but not its $O(N/t)$ scaling with market complexity. For empirical calibration, specialized quantitative funds and hedge funds collectively represent less than 5% of equity market capitalization, suggesting θ_s is small in practice.

4 Motivating Evidence

Three facts about individual stock pricing errors motivate the theoretical analysis. We focus on individual stocks rather than diversified portfolios because portfolios diversify away idiosyncratic variation, obscuring the asset-level estimation problem.

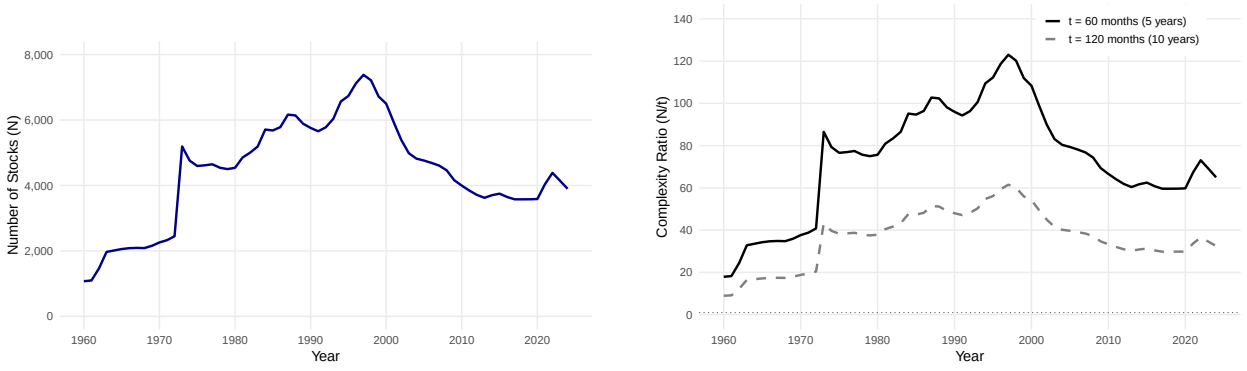
4.1 Data

We use monthly returns for all common stocks (share codes 10, 11) traded on NYSE, AMEX, and NASDAQ (exchange codes 1, 2, 3) from CRSP. For daily analysis, we use daily excess returns. Our sample spans 1962–2023. We merge with Fama-French factors for risk adjustment.

4.2 Fact 1: The Scale of the Estimation Problem

Investors pricing N assets with t observations face a complexity ratio N/t. Throughout the sample, $N/t \gg 1$.

Figure 1 plots the number of stocks over time. The cross-section ranges from 2,000 stocks in 1962 to over 7,000 in 1997, before declining to 4,000 by 2023. For $t = 60$ months (5 years), N/t ranges from 30 to over 120. Even with $t = 120$ months (10 years), N/t remains between 15 and 60.



(a) Number of Stocks

(b) Complexity Ratio N/t

Figure 1: The Scale of the Estimation Problem. Panel (a) plots common stocks (share codes 10, 11) on NYSE, AMEX, and NASDAQ from CRSP, 1962–2023. Panel (b) computes the complexity ratio N/t for estimation windows of 60 and 120 months.

Even with a decade of data, investors must estimate far more parameters than they have observations.

4.3 Fact 2: High Dimensionality of Individual Stock Returns

Factor structure does not reduce the effective dimensionality of individual stock expected returns.

Figure 2 plots cumulative variance explained from PCA on daily individual stock returns. We run PCA on non-overlapping 5-year windows from 1980–2019, keeping only stocks with complete data in each window (1,500–2,500 stocks per window, 1,250 trading days). The start date reflects daily return data availability in CRSP; the end date reflects the last complete 5-year window.

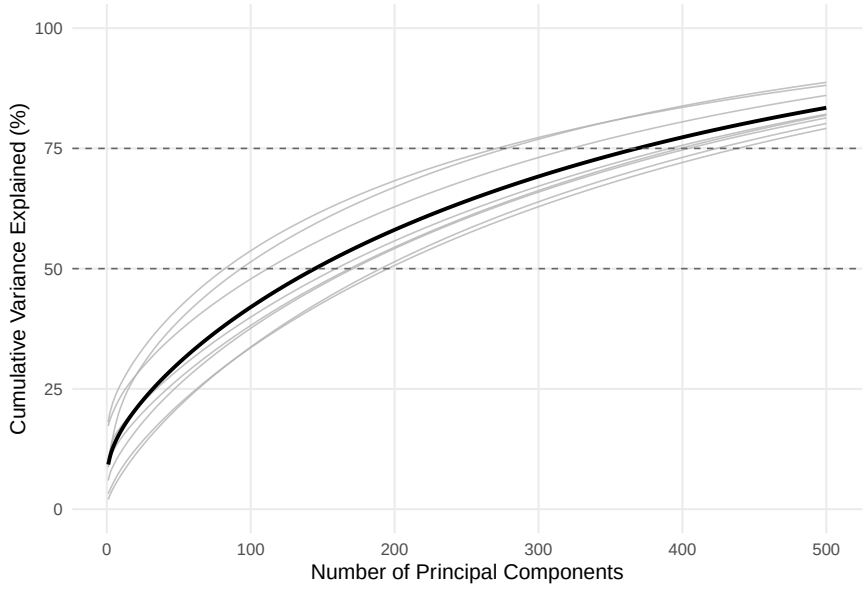


Figure 2: Cumulative Variance Explained by Principal Components: Individual Stocks. PCA is run on daily excess returns for stocks with complete data in each non-overlapping 5-year window from 1980–2019. Each gray line represents one window; the black line is the cross-window average.

Explaining 50% of variance in individual stock returns requires 145 principal components, and 370 for 75%. The first 5 principal components capture 15% of return variation. For diversified portfolios, 50 factors explain 50% of variance. Portfolios diversify away idiosyncratic variation, leaving the high-dimensional component of the estimation problem intact.

The PCA is on total return variance, not expected returns directly. But the high dimensionality of returns means that even if common factors were perfectly known, the N asset-level idiosyncratic parameters must be estimated separately from returns whose variation is spread across hundreds of dimensions.

4.4 Fact 3: Pricing Errors Scale with N

The aggregate squared pricing error, measured by $\alpha'\Lambda^{-1}\alpha$ (the squared Sharpe ratio improvement from exploiting pricing errors beyond the market portfolio), scales with the number of assets N.

For each year from 1970 to 2023, we estimate CAPM regressions for all stocks with at least 36 months of data in the preceding 5-year window:

$$r_{i,t} - r_f = \alpha_i + \beta_i(r_m - r_f) + \varepsilon_{i,t} \tag{54}$$

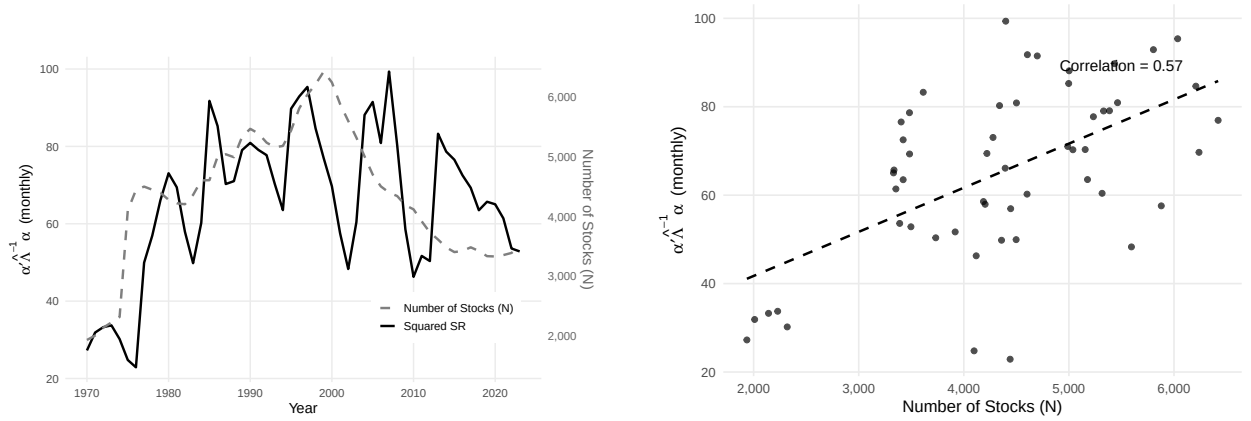
We then compute aggregate pricing errors $\alpha'\hat{\Lambda}^{-1}\alpha$, where $\hat{\Lambda}$ is the full return covariance matrix estimated using a factor model with $K = 10$ principal component factors plus a diagonal residual covariance, computed from excess returns in the same window.[18]

Figure 3 plots the results. $\alpha'\hat{\Lambda}^{-1}\alpha$ and N move together (Panel a); their correlation is 0.57 (Newey-West $t = 2.82$, $p = 0.007$; Panel b). The sample statistic $\hat{\alpha}'\hat{\Lambda}^{-1}\hat{\alpha}$ also scales with N mechanically under the null of zero true alphas, so the correlation documents consistency with the model rather than testing it against the null; see the identification caveat below.

The implied annualized Sharpe ratio, $\sqrt{12 \cdot \alpha'\hat{\Lambda}^{-1}\alpha}$, averages 28 (range 17–35). These Sharpe ratios are implausible as exploitable returns and point to estimation error in $\hat{\alpha}$.

Identification caveat. The correlation does not establish causation. U.S. stock listings follow an inverted-U shape peaking in the late 1990s, and many macro-finance variables (idiosyncratic

18. Using a diagonal matrix of CAPM residual variances instead of the full return covariance yields a higher correlation with N. We report the factor-model estimator because $\alpha'\Lambda^{-1}\alpha$ in the theory uses the full return covariance Λ, not the CAPM residual covariance.



(a) Time Series

(b) Scatter Plot

Figure 3: Pricing Errors and the Number of Assets. For each year from 1970 to 2023, CAPM regressions are estimated on stocks with 36+ months of data in the preceding 5-year window. Panel (a) plots $\alpha'\hat{\Lambda}^{-1}\alpha$ (solid line, left axis) and N (dashed line, right axis). Panel (b) plots $\alpha'\hat{\Lambda}^{-1}\alpha$ against N.

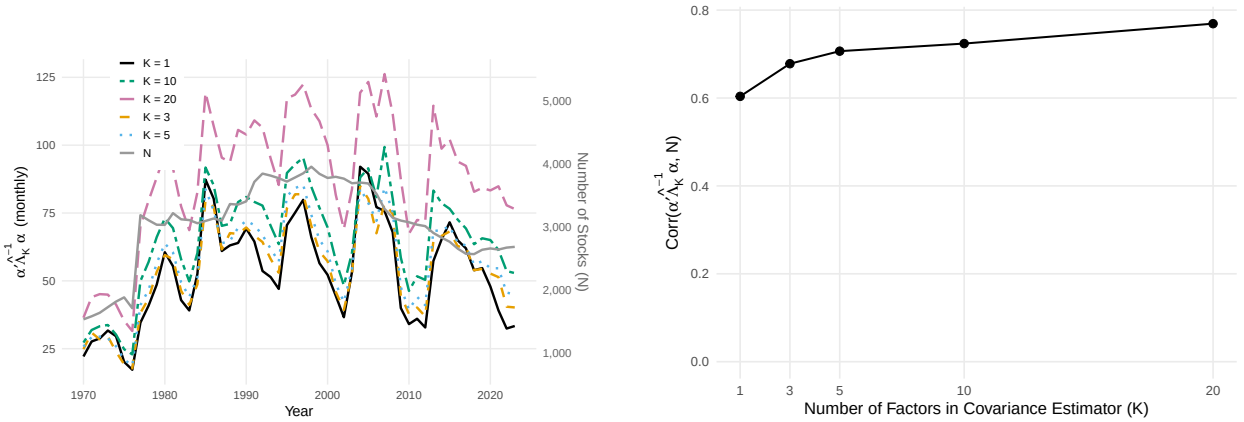
volatility, retail participation, turnover) share similar low-frequency trends that could generate spurious correlation. The covariance-adjusted statistic $\alpha'\hat{\Lambda}^{-1}\alpha$ absorbs some of the rise in idiosyncratic volatility, but shifts in the cross-sectional distribution of volatility could still affect the scaling. The robustness analysis below shows the correlation strengthens with richer covariance estimators, indicating the alphas rather than covariance misspecification drive the pattern, but low-frequency macro trends could confound the correlation. A further subtlety is that the model predicts scaling of true equilibrium pricing errors $\alpha'\Lambda^{-1}\alpha$, while the empirical measure $\hat{\alpha}'\hat{\Lambda}^{-1}\hat{\alpha}$ also reflects the econometrician's own estimation noise. Both equilibrium mispricing and econometric noise scale with N/t for the same reason: high-dimensional estimation from limited data. The empirical exercise documents the scaling pattern the theory predicts without separating equilibrium mispricing from econometric noise.

4.5 Robustness: Varying the Covariance Estimator

The scaling might reflect misspecification of $\hat{\Lambda}$ rather than the CAPM alphas. We hold the alphas fixed and vary K in the factor-model covariance estimator $\hat{\Lambda}_K = \hat{B}_K\hat{\Sigma}_f\hat{B}_K' + \hat{D}$. If the correlation were an artifact of the covariance estimate, it should vary with K; if the alphas drive

the scaling, the correlation should persist regardless of K.

Figure 4 plots the results.[19] The correlation between $\alpha'\hat{\Lambda}_K^{-1}\alpha$ and N ranges from 0.60 at $K = 1$ to 0.77 at $K = 20$, increasing monotonically with K.



(a) Time Series by K

(b) Correlation with N by K

Figure 4: Pricing Error Scaling Across Covariance Estimators. CAPM alphas are held fixed. The covariance matrix $\hat{\Lambda}_K$ is estimated using K principal component factors plus a diagonal residual covariance. Panel (a) plots $\alpha'\hat{\Lambda}_K^{-1}\alpha$ over time for different K. Panel (b) plots the correlation between $\alpha'\hat{\Lambda}_K^{-1}\alpha$ and N as a function of K.

The pattern is consistent with alphas rather than the covariance estimator driving the N-scaling. A richer factor structure in $\hat{\Lambda}$ brings the weighting matrix closer to the true Λ^{-1}.

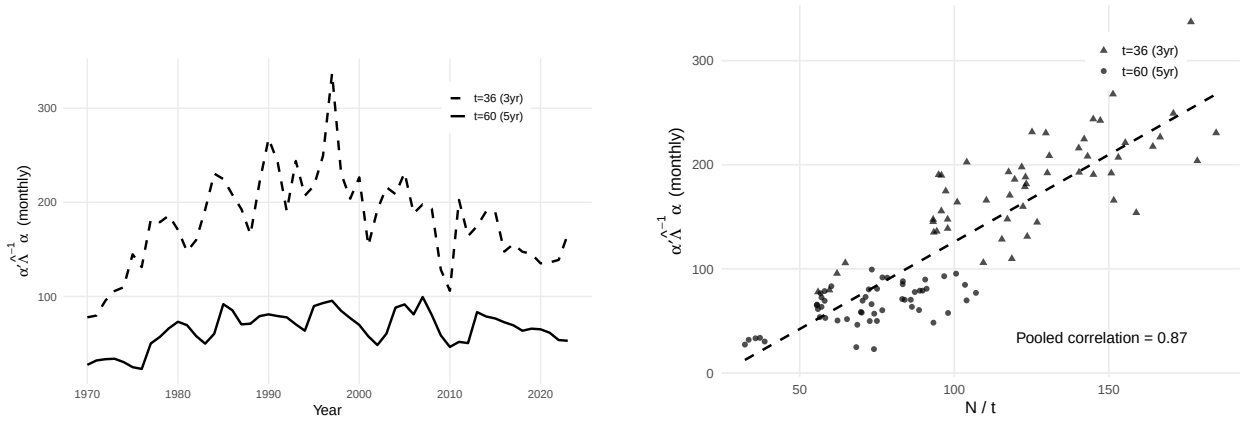
4.6 Robustness: Varying the Estimation Window

The theory predicts pricing errors proportional to N/t. Shortening the estimation window t increases N/t and should increase $\alpha'\hat{\Lambda}^{-1}\alpha$. We compare 3-year ($t = 36$) and 5-year ($t = 60$) estimation windows, holding the empirical procedure otherwise identical.

Figure 5 plots the results. The 3-year window produces larger pricing errors in every year, with a mean ratio of 2.89. The model predicts the qualitative direction but not the exact magnitude: the effective estimation horizon t in the model need not equal the window length in months, and the stock universes differ across windows (more stocks satisfy the 36-month filter than the 60-month

19. A balanced panel is required here, so the baseline correlation (0.72) differs from Figure 3 (0.57), which includes stocks with partial data.

filter). The pooled correlation between $\alpha'\hat{\Lambda}^{-1}\alpha$ and N/t across both windows is positive (Figure 5b), consistent with the N/t scaling.



(a) Time Series

(b) $\alpha'\hat{\Lambda}^{-1}\alpha$ vs. N/t

Figure 5: Pricing Errors Across Estimation Windows. CAPM alphas and factor-model covariance matrices ($K = 10$) are estimated using 3-year and 5-year windows ending in each year from 1970 to 2023. Panel (a) plots $\alpha'\hat{\Lambda}^{-1}\alpha$ for both windows over time. Panel (b) pools both windows and plots $\alpha'\hat{\Lambda}^{-1}\alpha$ against N/t.

4.7 Summary

First, N/t ranges from 30 to 120, placing the estimation problem in the high-dimensional regime. Second, 145+ principal components are needed for 50% of return variance, so the estimation problem agents face is high-dimensional. Third, aggregate pricing errors $\alpha'\Lambda^{-1}\alpha$ covary positively with N (correlation 0.57), and the scaling strengthens with richer covariance estimators, consistent with idiosyncratic alpha estimation errors driving the pattern. Sections 1–2 show that complexity generates all three patterns without behavioral biases, limits to arbitrage, or market frictions.

5 Strategy Proliferation with Low Correlations

Academic factor models and systematic hedge fund portfolios both achieve high Sharpe ratios with low pairwise correlations.

A factor model prices the cross-section if and only if the factor model can be collapsed to a mean-variance efficient portfolio (Huberman and Kandel 1987). Table 1 shows the factor models and collapse methods from the literature that are used here. Each strategy is named by combining a factor model with a collapse method (e.g., "GKX-BPZ" denotes the Gu, Kelly, and Xiu 2021 factor model collapsed using Bryzgalova, Pelger, and Zhu 2020a). Davis 2024 describes the implementation of these models. The analysis uses the Freyberger, Neuhierl, and Weber 2020b dataset of 62 asset characteristics. Following Kozak, Nagel, and Santosh 2020 and Kelly, Pruitt, and Su 2019, characteristics are converted to cross-sectional percentiles (between 0 and 1) and centered at zero to form long-short managed portfolios. Monthly data span July 1962 through December 2023. Parameters are estimated through December 1989 and re-estimated annually; the evaluation period runs January 1990 to December 2023, using strictly ex-ante information.

Panel (a) of Figure 6 shows the distribution of pairwise correlations across all strategies. The average pairwise correlation is 0.26 (implied R^2 of 7%).

Beyond academic strategies, we examine actual hedge fund portfolios using 13F holdings data from the SEC EDGAR database, covering the period 1999–2023. We follow the Koijen and Yogo 2019 methodology to correct investor types in the 13F data, and define a systematic hedge fund as an "investment advisor" that holds at least 500 stocks. Funds holding fewer than 500 stocks are excluded. The filter yields 80–120 funds per quarter. Among the well-known funds captured are AQR, Renaissance, Tower Research, Citadel, Two Sigma, S.A.C Capital, D.E. Shaw, Roll and Ross Asset Management, Paradigm Asset Management, Bridgewater Associates, Cubist Systematic Strategies, Martingale Asset Management, and Quantitative Systematic Strategies.

We isolate the active component of each fund's portfolio by subtracting the fund's market beta times the market portfolio weights. If the weights of fund i are w_i and its beta is β_i, the market-neutral weights are

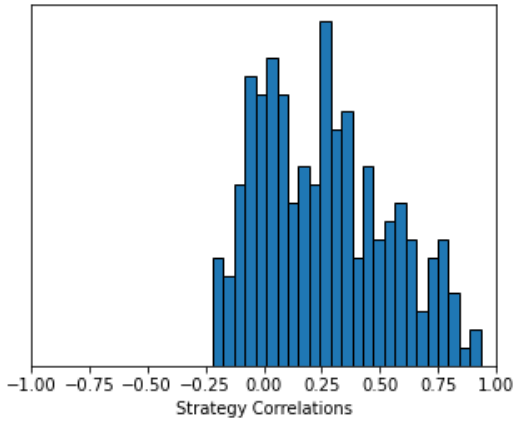
$$w_{i|M} = w_i - \beta_i w_M. \tag{55}$$

The correlation between the active components of funds i and j is then:

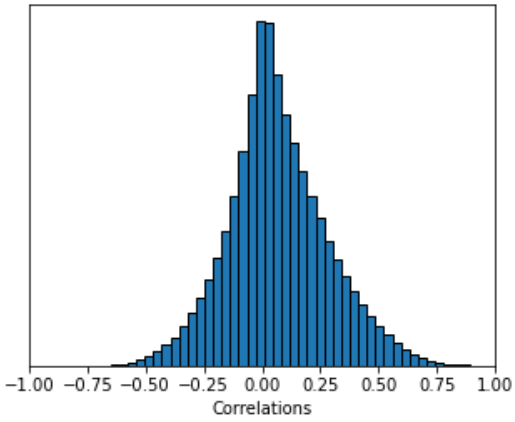
$$\text{corr}_{i,j} = \frac{(w_{i|M})'\Lambda(w_{j|M})}{\sqrt{(w_{i|M})'\Lambda(w_{i|M})}\sqrt{(w_{j|M})'\Lambda(w_{j|M})}} \tag{56}$$

where Λ is the return covariance matrix estimated using a factor model with $K = 10$ principal component factors plus a diagonal residual covariance, computed from daily returns over the preceding 12 months. Because 13F filings report only long positions, the correlations reflect the long book only.[20] Panel (a) reports return-based correlations across academic strategies; Panel (b) reports holdings-based correlations for hedge funds, using the covariance-based formula in Equation (56) because 13F holdings are observed only quarterly.

Panel (b) of Figure 6 shows the pairwise correlations. The average is 0.07. Koijen and Yogo 2019 similarly find that classic characteristics leave most demand unexplained.



(a) Trading Strategy Correlations



(b) Systematic Hedge Fund Cross Correlations

Figure 6: Correlations of Portfolios. Panel (a) computes pairwise return correlations across the trading strategies listed in Table 1 over January 1990 to December 2023. Panel (b) computes pairwise correlations of the market-neutral component of systematic hedge fund portfolios from 13F holdings data, using the covariance-based formula in Equation (56).

20. If short positions are correlated across funds (e.g., because many funds short the same overpriced stocks), the true active-portfolio correlations could be higher than what we estimate from long holdings alone.

Initialism	Factor Model
BPZF	Bryzgalova, Pelger, and Zhu 2020b
L	linear characteristic-weighted portfolios
CRW	Chen, Roussanov, and Wang 2021 linear portfolios
FF3	Fama and French 1993
FF6	Fama and French 2015 with Carhart 1997 momentum
GKX	Gu, Kelly, and Xiu 2021
HXZ	Hou, Xue, and Zhang 2014
KPS	Kelly, Pruitt, and Su 2019
NN	neural network
RF	random forest
Initialism	**Collapse Method**
BPZ	Bryzgalova, Pelger, and Zhu 2020a
KNS	Kozak, Nagel, and Santosh 2020
BSV	Brandt and Santa-Clara 2006
DGU	DeMiguel, Garlappi, and Uppal 2007

Table 1: Statistical Arbitrageur Demand Models. Initialisms and sources for the investment strategy models.

Table 2 reports CAPM regressions: alphas range from 18 to 51 basis points per month, with Sharpe ratios between 1.4 and 3.2.

Tables 3 and 4 regress each strategy's returns on other strategies' returns. Within similar strategy families (e.g., machine learning-based approaches), betas are high (0.5–0.8); across different strategy types, betas are near zero or negative. Market betas explain little of the variation in strategy returns. Strategies load on different return components rather than common risk factors.

Panel A: Weber CAPM Regressions Part 1

	NN	L-BPZ	KPS-BPZ	CRW-KNS	BPZF-BPZ	GKX-BPZ	GKX-BSV	FF6-DGU	HXZ-BSV	L-DGU	FF6-KNS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
α	51.10***	45.26***	45.01***	40.80***	27.12***	28.22***	27.29***	18.35***	22.79***	23.48***	20.42***
	(2.61)	(2.67)	(2.67)	(2.68)	(2.53)	(2.68)	(2.67)	(2.21)	(2.68)	(2.63)	(2.66)
β	-0.22***	0.08*	-0.06	-0.01	0.33***	-0.05	-0.07	0.57***	0.02	-0.18***	0.10**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
Observations	408	408	408	408	408	408	408	408	408	408	408
R^2	0.05	0.01	0.00	0.00	0.11	0.00	0.01	0.32	0.00	0.03	0.01
Sharpe Ratio	3.20	2.98	2.89	2.64	1.94	1.81	1.73	1.50	1.49	1.42	1.38
Note:											*p<0.1; **p<0.05; ***p<0.01

Panel B: Weber CAPM Regressions Part 2

	GKX-KNS	KPS-DGU	BPZF-DGU	HXZ-KNS	FF3-KNS	RF	FF3-DGU	CAPM	CRW-BPZ	GKX-DGU	BPZF-BSV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
α	21.58***	16.49***	10.74***	9.34***	12.93***	8.72***	5.18***		7.25***	4.00	-4.24
	(2.68)	(2.38)	(2.07)	(1.74)	(2.56)	(2.04)	(1.30)		(2.67)	(2.67)	(2.68)
β	-0.04	0.46***	0.63***	0.76***	0.29***	0.65***	0.87***		0.06	0.08*	-0.02
	(0.05)	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)	(0.02)		(0.05)	(0.05)	(0.05)
Observations	408	408	408	408	408	408	408	408	408	408	408
R^2	0.00	0.21	0.40	0.58	0.08	0.42	0.76		0.00	0.01	0.00
Sharpe Ratio	1.38	1.32	1.04	1.02	1.00	0.92	0.82	0.55	0.50	0.31	-0.29
Note:											*p<0.1; **p<0.05; ***p<0.01

Table 2: CAPM Regressions for Factor Model Strategies. CAPM regressions $R_{i,t} = \alpha + \beta R_{m,t} + \varepsilon_{i,t}$ for the factor model strategies in Table 1, January 1990 to December 2023. Panels A and B show separate subsets of strategies. Standard errors in parentheses.

Panel A: Weber Cross Regressions Beta Part 1

	Rows: Test Assets, Columns: Right-hand-side Portfolios										
	NN	L-BPZ	KPS-BPZ	CRW-KNS	BPZF-BPZ	GKX-BPZ	GKX-BSV	FF6-DGU	HXZ-BSV	L-DGU	FF6-KNS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
NN		0.62***	0.61***	0.65***	0.43***	0.42***	0.44***	0.04	0.21**	0.25**	0.12
		(0.05)	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)	(0.10)	(0.11)	(0.14)
L-BPZ	0.62***		0.54***	0.55***	0.41***	0.36***	0.35***	0.27***	0.16	0.22	0.19
	(0.08)		(0.06)	(0.10)	(0.11)	(0.11)	(0.13)	(0.09)	(0.14)	(0.14)	(0.21)
KPS-BPZ	0.61***	0.54***		0.74***	0.31***	0.56***	0.44***	0.18***	0.16**	0.30***	0.23**
	(0.05)	(0.09)		(0.07)	(0.05)	(0.06)	(0.07)	(0.07)	(0.06)	(0.08)	(0.09)
CRW-KNS	0.65***	0.55***	0.74***		0.50***	0.65***	0.66***	0.03	-0.02	0.05	-0.07
	(0.06)	(0.07)	(0.07)		(0.09)	(0.06)	(0.06)	(0.08)	(0.10)	(0.11)	(0.14)
BPZF-BPZ	0.43***	0.41***	0.31***	0.50***		0.40***	0.44***	0.27***	-0.05	-0.09	-0.02
	(0.11)	(0.11)	(0.07)	(0.10)		(0.10)	(0.11)	(0.09)	(0.11)	(0.12)	(0.16)
GKX-BPZ	0.42***	0.36***	0.56***	0.65***	0.40***		0.57***	-0.07	-0.03	-0.09	-0.19*
	(0.07)	(0.08)	(0.06)	(0.07)	(0.07)		(0.06)	(0.08)	(0.09)	(0.10)	(0.11)
GKX-BSV	0.44***	0.35***	0.44***	0.66***	0.44***	0.57***		-0.10	-0.05	-0.12	-0.21
	(0.09)	(0.12)	(0.08)	(0.08)	(0.10)	(0.09)		(0.09)	(0.11)	(0.12)	(0.14)
FF6-DGU	0.04	0.27**	0.18***	0.03	0.27**	-0.07	-0.10		0.57***	0.62***	0.81***
	(0.08)	(0.13)	(0.06)	(0.08)	(0.11)	(0.07)	(0.08)		(0.07)	(0.05)	(0.07)
HXZ-BSV	0.21**	0.16	0.16***	-0.02	-0.05	-0.03	-0.05	0.57***		0.60***	0.77***
	(0.10)	(0.15)	(0.06)	(0.10)	(0.11)	(0.09)	(0.11)	(0.06)		(0.06)	(0.06)
L-DGU	0.25**	0.22	0.30***	0.05	-0.09	-0.09	-0.12	0.62***	0.60***		0.83***
	(0.11)	(0.17)	(0.06)	(0.12)	(0.12)	(0.10)	(0.11)	(0.06)	(0.08)		(0.04)
FF6-KNS	0.12	0.19	0.23***	-0.07	-0.02	-0.19	-0.21	0.81***	0.77***	0.83***	
	(0.13)	(0.21)	(0.07)	(0.14)	(0.16)	(0.12)	(0.14)	(0.06)	(0.07)	(0.06)	
GKX-KNS	0.51***	0.35***	0.51***	0.66***	0.46***	0.56***	0.56***	-0.06	-0.05	-0.09	-0.17
	(0.06)	(0.08)	(0.07)	(0.05)	(0.07)	(0.06)	(0.06)	(0.08)	(0.09)	(0.10)	(0.11)
KPS-DGU	0.27***	0.37***	0.46***	0.51***	0.69***	0.44***	0.42***	0.33***	-0.07	-0.08	0.02
	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.07)	(0.08)	(0.09)	(0.10)
BPZF-DGU	0.14	0.30***	0.15**	0.32***	0.77***	0.26***	0.26**	0.45***	-0.08	-0.17*	0.03
	(0.12)	(0.10)	(0.08)	(0.11)	(0.04)	(0.08)	(0.11)	(0.08)	(0.09)	(0.10)	(0.14)
HXZ-KNS	-0.03	0.23***	0.03	0.11	0.55***	0.06	0.06	0.76***	0.27***	0.09	0.40***
	(0.09)	(0.06)	(0.07)	(0.08)	(0.06)	(0.06)	(0.08)	(0.05)	(0.08)	(0.07)	(0.12)
FF3-KNS	0.05	0.16	0.08	-0.03	0.16	0.01	-0.02	0.74***	0.71***	0.50***	0.68***
	(0.08)	(0.13)	(0.06)	(0.08)	(0.11)	(0.08)	(0.10)	(0.04)	(0.05)	(0.06)	(0.06)
RF	0.18	0.36***	0.14*	0.35**	0.75***	0.23**	0.26*	0.47***	-0.09	-0.13	0.02
	(0.15)	(0.13)	(0.08)	(0.14)	(0.07)	(0.11)	(0.15)	(0.10)	(0.12)	(0.13)	(0.18)
FF3-DGU	-0.10	0.20***	0.01	0.08	0.51***	0.08	0.05	0.72***	0.19**	-0.03	0.26***
	(0.08)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.08)	(0.07)	(0.10)
CAPM	-0.22***	0.08	-0.06	-0.01	0.33***	-0.05	-0.07	0.57***	0.02	-0.18***	0.10
	(0.06)	(0.05)	(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.06)	(0.07)	(0.06)	(0.07)
CRW-BPZ	0.17**	0.22***	0.24***	0.24***	0.31***	0.34***	0.26***	0.15**	0.19**	-0.02	0.07
	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)	(0.09)	(0.10)
GKX-DGU	0.24***	0.32***	0.27***	0.36***	0.53***	0.28***	0.36***	0.12	-0.05	-0.01	0.00
	(0.08)	(0.09)	(0.08)	(0.08)	(0.05)	(0.09)	(0.08)	(0.08)	(0.09)	(0.11)	(0.13)
BPZF-BSV	0.12*	0.11*	0.03	0.00	0.09	0.05	-0.02	-0.10*	-0.20***	-0.06	-0.16***
	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.05)	(0.05)	(0.07)	(0.06)
Observations	408	408	408	408	408	408	408	408	408	408	408
F-test	6.94	7.13	7.93	10.96	18.42	21.33	22.45	19.06	18.81	17.95	17.55
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Note:											*p<0.1; **p<0.05; ***p<0.01

Table 3: Cross-Sectional Beta Regressions Across Strategies (Part 1). Each strategy's returns regressed on all other strategies' returns, January 1990 to December 2023. Part 1 of 2; see Table 4 for Part 2. Standard errors in parentheses.

Panel B: Weber Cross Regressions Beta Part 2

	Rows: Test Assets, Columns: Right-hand-side Portfolios										
	GKX-KNS	KPS-DGU	BPZF-DGU	HXZ-KNS	FF3-KNS	RF	FF3-DGU	CAPM	CRW-BPZ	GKX-DGU	BPZF-BSV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
NN	0.51***	0.27***	0.14	-0.03	0.05	0.18	-0.10	-0.22***	0.17*	0.24***	0.12*
	(0.08)	(0.08)	(0.12)	(0.09)	(0.08)	(0.13)	(0.07)	(0.05)	(0.09)	(0.09)	(0.07)
L-BPZ	0.35***	0.37***	0.30***	0.23***	0.16	0.36***	0.20***	0.08	0.22**	0.32***	0.11
	(0.11)	(0.09)	(0.11)	(0.07)	(0.11)	(0.12)	(0.06)	(0.05)	(0.10)	(0.11)	(0.07)
KPS-BPZ	0.51***	0.46***	0.15**	0.03	0.08	0.14**	0.01	-0.06	0.24***	0.27***	0.03
	(0.05)	(0.06)	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)	(0.05)	(0.07)	(0.08)	(0.06)
CRW-KNS	0.66***	0.51***	0.32***	0.11	-0.03	0.35***	0.08	-0.01	0.24**	0.36***	0.00
	(0.07)	(0.08)	(0.11)	(0.09)	(0.09)	(0.11)	(0.07)	(0.05)	(0.10)	(0.10)	(0.07)
BPZF-BPZ	0.46***	0.69***	0.77***	0.55***	0.16	0.75***	0.51***	0.33***	0.31***	0.53***	0.09
	(0.10)	(0.07)	(0.07)	(0.07)	(0.10)	(0.05)	(0.06)	(0.04)	(0.10)	(0.09)	(0.08)
GKX-BPZ	0.56***	0.44***	0.26***	0.06	0.01	0.23**	0.08	-0.05	0.34***	0.28***	0.05
	(0.06)	(0.07)	(0.08)	(0.06)	(0.08)	(0.09)	(0.06)	(0.05)	(0.08)	(0.09)	(0.07)
GKX-BSV	0.56***	0.42***	0.26**	0.06	-0.02	0.26*	0.05	-0.07	0.26***	0.36***	-0.02
	(0.09)	(0.09)	(0.11)	(0.08)	(0.10)	(0.14)	(0.07)	(0.05)	(0.10)	(0.11)	(0.07)
FF6-DGU	-0.06	0.33***	0.45***	0.76***	0.74***	0.47***	0.72***	0.57***	0.15**	0.12	-0.10
	(0.07)	(0.08)	(0.10)	(0.05)	(0.04)	(0.15)	(0.05)	(0.06)	(0.07)	(0.08)	(0.06)
HXZ-BSV	-0.05	-0.07	-0.08	0.27***	0.71***	-0.09	0.19***	0.02	0.19**	-0.05	-0.20***
	(0.09)	(0.08)	(0.10)	(0.06)	(0.05)	(0.11)	(0.06)	(0.07)	(0.07)	(0.09)	(0.08)
L-DGU	-0.09	-0.08	-0.17	0.09	0.50***	-0.13	-0.03	-0.18***	-0.02	-0.01	-0.06
	(0.10)	(0.09)	(0.11)	(0.07)	(0.07)	(0.12)	(0.07)	(0.07)	(0.09)	(0.11)	(0.08)
FF6-KNS	-0.17	0.02	0.03	0.40***	0.68***	0.02	0.26***	0.10	0.07	0.00	-0.16*
	(0.12)	(0.10)	(0.13)	(0.06)	(0.07)	(0.18)	(0.07)	(0.06)	(0.10)	(0.13)	(0.08)
GKX-KNS		0.43***	0.32***	0.10	-0.07	0.32***	0.07	-0.04	0.30***	0.36***	0.04
		(0.08)	(0.09)	(0.08)	(0.08)	(0.09)	(0.06)	(0.05)	(0.09)	(0.10)	(0.07)
KPS-DGU	0.43***		0.76***	0.61***	0.21***	0.71***	0.60***	0.46***	0.34***	0.59***	0.04
	(0.06)		(0.04)	(0.05)	(0.08)	(0.07)	(0.05)	(0.05)	(0.07)	(0.05)	(0.07)
BPZF-DGU	0.32***	0.76***		0.82***	0.27***	0.93***	0.80***	0.63***	0.33***	0.50***	0.03
	(0.10)	(0.06)		(0.05)	(0.09)	(0.05)	(0.04)	(0.03)	(0.09)	(0.08)	(0.08)
HXZ-KNS	0.10	0.61***	0.82***		0.52***	0.79***	0.92***	0.76***	0.29***	0.36***	-0.08
	(0.07)	(0.05)	(0.06)		(0.07)	(0.11)	(0.03)	(0.03)	(0.07)	(0.06)	(0.06)
FF3-KNS	-0.07	0.21**	0.27***	0.52***		0.25*	0.59***	0.29***	0.23***	0.11	-0.15**
	(0.08)	(0.08)	(0.10)	(0.06)		(0.14)	(0.06)	(0.07)	(0.07)	(0.09)	(0.07)
RF	0.32**	0.71***	0.93***	0.79***	0.25**		0.79***	0.65***	0.29***	0.45***	0.08
	(0.13)	(0.08)	(0.06)	(0.06)	(0.11)		(0.05)	(0.03)	(0.11)	(0.11)	(0.08)
FF3-DGU	0.07	0.60***	0.80***	0.92***	0.59***	0.79***		0.87***	0.24***	0.29***	-0.05
	(0.06)	(0.06)	(0.07)	(0.04)	(0.07)	(0.12)		(0.03)	(0.06)	(0.06)	(0.06)
CAPM	-0.04	0.46***	0.63***	0.76***	0.29***	0.65***	0.87***		0.06	0.08	-0.02
	(0.05)	(0.05)	(0.09)	(0.05)	(0.07)	(0.14)	(0.03)		(0.05)	(0.07)	(0.05)
CRW-BPZ	0.30***	0.34***	0.33***	0.29***	0.23***	0.29***	0.24***	0.06		0.35***	-0.16**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)		(0.08)	(0.07)
GKX-DGU	0.36***	0.59***	0.50***	0.36***	0.11	0.45***	0.29***	0.08	0.35***		0.07
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)		(0.08)
BPZF-BSV	0.04	0.04	0.03	-0.08	-0.15***	0.08	-0.05	-0.02	-0.16**	0.07	
	(0.07)	(0.07)	(0.08)	(0.06)	(0.05)	(0.08)	(0.06)	(0.05)	(0.08)	(0.07)	
Observations	408	408	408	408	408	408	408	408	408	408	408
F-test	21.64	26.59	26.19	28.01	24.18	25.40	28.38	26.82	30.61	29.93	27.40
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

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

Table 4: Cross-Sectional Beta Regressions Across Strategies (Part 2). Coefficients from cross-sectional regressions of each strategy's returns (rows) on other strategies' returns (columns), January 1990 to December 2023. See Table 3 for Part 1. Standard errors in parentheses. F-test and p-value test joint significance.

6 Conclusion

Estimation uncertainty under market complexity reverses a prediction of Arbitrage Pricing Theory. APT predicts that pricing errors vanish as markets grow; instead, equilibrium pricing errors scale linearly with the number of assets and inversely with effective sample size. The bound has two components: a learnable part from uncertainty about permanent expected returns that shrinks with sample size, and a persistent part from time-varying fundamentals that remains irreducible regardless of estimation procedure or data availability. The bound is estimator-agnostic.

The equilibrium stochastic discount factor loads primarily on estimation errors rather than on fundamental risk factors. Because estimation uncertainty disperses across all N asset dimensions while any implementable portfolio spans a single direction, factor models cannot capture the $N - 1$ orthogonal dimensions of priced estimation risk. The weak beta-return relationship and the failure of parsimonious factor models both follow from the same dimensional mismatch. Multiple strategies achieve high Sharpe ratios with low cross-correlations because each strategy exploits different dimensions of the estimation error space. When investors disagree about expected returns, mispricing persists even with informed investors because risk aversion limits the positions that would otherwise eliminate complexity-driven pricing errors.

Individual stock returns require 145 principal components to explain half their variance, and aggregate pricing errors covary positively with the number of listed stocks. The active positions of systematic hedge funds exhibit near-zero pairwise correlations (average 0.07), and academic factor model strategies display an average pairwise correlation of 0.26, consistent with exploitation of distinct error dimensions.

Testable predictions. The $O(N/t)$ bound generates comparative statics beyond the baseline scaling.

Countries with more listed firms relative to available data history should exhibit larger pricing errors. Emerging markets with short reliable data histories face large N/t ratios even with fewer listed firms. Events that shift the data-generating process (financial crises, regulatory changes,

pandemics) effectively reduce the useful sample size t, since pre-break data is less informative about post-break parameters. Pricing errors should spike after regime changes even controlling for N.

Not all firms contribute equally to aggregate pricing errors. Younger firms have shorter histories (smaller firm-level t), so their expected returns are estimated less precisely; pricing errors should be concentrated among younger firms. Firms in rapidly changing industries (technology, biotechnology) have expected returns that shift more frequently, contributing a larger persistent component to pricing errors. Firms with high idiosyncratic return volatility relative to expected return variation have a worse signal-to-noise ratio, and sorting firms by idiosyncratic volatility, the high-volatility group should contribute disproportionately to aggregate pricing errors through the learnable channel.

The prior precision κ_0 maps to investor sophistication. Stronger priors reduce pricing errors when correctly specified ($\pi_0 = \pi$), but when $\pi_0 \neq \pi$, larger κ_0 can increase pricing errors: the agent is confident but wrong, and the bias does not shrink with t. Overconfident investors with misspecified models generate larger errors than investors with diffuse priors. Even perfectly calibrated Bayesians face $O(N)$ errors when fundamentals are time-varying (Corollary 1), implying diminishing returns to sophistication in complex, rapidly changing markets.

Pricing errors are convex in the fraction of uninformed investors (scaling as θ_u^2 in Proposition 6), so the marginal value of informed capital is highest when most capital is uninformed. Higher aggregate risk aversion amplifies the persistence of mispricing, so periods of elevated risk aversion should exhibit larger pricing errors even if the information environment is unchanged. Markets with more specialized investors (lower coverage ϕ per fund, more funds) should exhibit more low-correlation strategies; the total mispricing corrected depends on aggregate coverage $\theta_s\phi$, not on the number of funds alone.

Diversified markets where factor loadings have offsetting signs should have smaller covariance-driven pricing errors than markets where all assets load in the same direction. With unknown factor loadings determined by K characteristics, pricing errors scale as $O(N/t + KS/(Nt))$; if the

number of characteristics grows proportionally to N, the mean estimation term dominates but with an effective complexity multiplier. Short-sale constraints and low liquidity effectively reduce the informed share or coverage fraction for affected assets, so constrained assets should exhibit larger pricing errors.

References

Alti, Aydogan, and Sheridan Titman. 2019. "A Dynamic Model of Characteristic-Based Return Predictability." *Journal of Finance* 74, no. 6 (December 1, 2019): 3187–3216. ISSN: 15406261, accessed July 10, 2021. https://doi.org/10.1111/jofi.12839.

Andrei, Daniel, Julien Cujean, and Mungo Wilson. 2023. "The Lost Capital Asset Pricing Model." *The Review of Economic Studies* 90, no. 6 (November 8, 2023): 2703–2762. ISSN: 0034-6527, accessed September 8, 2024. https://doi.org/10.1093/RESTUD/RDAD013.

Bossaerts, Peter, Charles Plott, and William R. Zame. 2007. "Prices and Portfolio Choices in Financial Markets: Theory, Econometrics, Experiments." *Econometrica* 75 (4): 993–1038. https://doi.org/10.1111/j.1468-0262.2007.00780.x.

Brandt, Michael W, and Pedro Santa-Clara. 2006. "Dynamic portfolio selection by augmenting the asset space." *The Journal of Finance* 61 (5): 2187–2217.

Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard. 2023. "Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models." *The Journal of Finance* 78 (1): 487–557. ISSN: 1540-6261, accessed May 1, 2025. https://doi.org/10.1111/jofi.13197.

Bryzgalova, Svetlana, Sven Lerner, Martin Lettau, and Markus Pelger. 2022. "Missing Financial Data." Pre-published, May 11, 2022. SSRN Scholarly Paper. Accessed May 1, 2025. https://doi.org/10.2139/ssrn.4106794. Social Science Research Network: 4106794.

Bryzgalova, Svetlana, Markus Pelger, and Jason Zhu. 2020a. "Forest Through the Trees: Building Cross-Sections of Stock Returns." *working paper* (September 25, 2020). Accessed July 26, 2023. https://doi.org/10.2139/SSRN.3493458.

———. 2020b. "Forest through the trees: Building cross-sections of stock returns." *Available at SSRN 3493458*.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *The Journal of Finance* 52 (1): 57–82.

Chamberlain, Gary, and Michael Rothschild. 1983. "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets." *Econometrica* 51 (5): 1281–1304. ISSN: 00129682, 14680262. JSTOR: 1912275. http://www.jstor.org/stable/1912275.

Chen, Andrew Y. 2021. "Most Claimed Statistical Findings in Cross-Sectional Return Predictability Are Likely True." *SSRN Electronic Journal* (October 8, 2021). Accessed January 5, 2022. https://doi.org/10.2139/SSRN.3912915.

Chen, Qihui, Nikolai L. Roussanov, and Xiaoliang Wang. 2021. "Semiparametric Conditional Factor Models: Estimation and Inference." *SSRN Electronic Journal* (July 13, 2021). ISSN: 1556-5068, accessed August 8, 2022. https://doi.org/10.2139/ssrn.3984633.

Da, Rui, Stefan Nagel, and Dacheng Xiu. 2022. "The Statistical Limit of Arbitrage."

Davis, Carter. 2024. "The Elasticity of Quantitative Investment." *working paper*.

Dello Preite, Massimo, Raman Uppal, Paolo Zaffaroni, and Irina Zviadadze. 2024. "Cross-Sectional Asset Pricing with Unsystematic Risk." Working paper, available at SSRN: https://ssrn.com/abstract=4135146. https://doi.org/10.2139/ssrn.4135146.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2007. "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies* 22 (5): 1915–1953. ISSN: 08939454.

Didisheim, Antoine, Shikun (Barry) Ke, Bryan T. Kelly, and Semyon Malamud. 2024. "APT or "AIPT"? The Surprising Dominance of Large Factor Models." (Cambridge, MA) (September 30, 2024). Accessed October 4, 2024. https://doi.org/10.3386/W33012.

Fama, Eugene F., and Kenneth R. French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33 (1): 3–56. ISSN: 0304-405X.

———. 2015. "A five-factor asset pricing model." *Journal of Financial Economics* 116:1–22. ISSN: 0304-405X.

Fama, Eugene F., and James D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* 81 (3): 607–636. ISSN: 00223808, 1537534X. JSTOR: 1831028. http://www.jstor.org/stable/1831028.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. 2020a. "Dissecting Characteristics Nonparametrically." *The Review of Financial Studies* 33 (5): 2326–2377. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhz123.

———. 2020b. "Dissecting Characteristics Nonparametrically." *The Review of Financial Studies* 33, no. 5 (April): 2326–2377. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhz123.

Gibbons, Michael R, Stephen A Ross, and Jay Shanken. 1989. "A Test of the Efficiency of a Given Portfolio." *Econometrica* 57 (5): 1121–1152. ISSN: 00129682, 14680262. JSTOR: 1913625. http://www.jstor.org/stable/1913625.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. "Empirical Asset Pricing via Machine Learning." *The Review of Financial Studies* 33 (5): 2223–2273. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhaa009.

———. 2021. "Autoencoder asset pricing models." Annals Issue:Financial Econometrics in the Age of the Digital Economy, *Journal of Econometrics* 222 (1, Part B): 429–450. ISSN: 0304-4076.

Harrison, J. Michael, and David M. Kreps. 1978. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations." *The Quarterly Journal of Economics* 92 (2): 323–336.

Harvey, Campbell R, Yan Liu, and Heqing Zhu. 2016. "... and the Cross-Section of Expected Returns." *The Review of Financial Studies* 29 (1): 5–68. https://doi.org/10.1093/rfs/hhv059.

Hou, Kewei, Chen Xue, and Lu Zhang. 2014. "Digesting Anomalies: An Investment Approach." *The Review of Financial Studies* 28, no. 3 (September): 650–705. ISSN: 0893-9454.

Huberman, Gur, and Shmuel Kandel. 1987. "Mean-Variance Spanning." *The Journal of Finance* 42 (4): 873–888.

Jacobs, Bruce I, and Kenneth N Levy. 1988. "Disentangling Equity Return Regularities: New Insights and Investment Opportunities." *Financial Analysts Journal* 44 (3). ISSN: 0015-198X. https://doi.org/10.2469/faj.v44.n3.18.

Kelly, Bryan, Seth Pruitt, and Yinan Su. 2019. "Characteristics Are Covariances: A Unified Model of Risk and Return." *Journal of Financial Economics* 134 (3): 501–524.

Kelly, Bryan T., Semyon Malamud, and Kangying Zhou. 2022. "The Virtue of Complexity Everywhere." *SSRN Electronic Journal* (July 18, 2022). Accessed July 26, 2023. https://doi.org/10.2139/SSRN.4166368.

Koijen, Ralph S. J., and Motohiro Yogo. 2019. "A Demand System Approach to Asset Pricing." *Journal of Political Economy* 127 (4): 1475–1515. https://doi.org/10.1086/701683.

Kozak, Serhiy, and Stefan Nagel. 2024. "When Do Cross-Sectional Asset Pricing Factors Span the Stochastic Discount Factor?" *SSRN Electronic Journal* (May 29, 2024). Accessed November 18, 2024. https://doi.org/10.2139/SSRN.4443643.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. 2018. "Interpreting Factor Models." *The Journal of Finance* 73 (3): 1183–1223. ISSN: 15406261. https://doi.org/10.1111/jofi.12612.

———. 2020. "Shrinking the Cross Section." *Journal of Financial Economics* 135 (2): 271–292.

Lewellen, Jonathan. 2015. "The Cross-section of Expected Stock Returns." *Critical Finance Review* 4 (1): 1–44. ISSN: 2164-5744. https://doi.org/10.1561/104.00000024.

Lopez-Lira, Alejandro, and Nikolai L. Roussanov. 2022. "Do Common Factors Really Explain the Cross-Section of Stock Returns?" *SSRN Electronic Journal* (July 1, 2022). Accessed October 15, 2022. https://doi.org/10.2139/ssrn.3628120.

Martin, Ian W.R., and Stefan Nagel. 2022. "Market Efficiency in the Age of Big Data." *Journal of Financial Economics* 145, no. 1 (July 1, 2022): 154–177. ISSN: 0304-405X. https://doi.org/10.1016/J.JFINECO.2021.10.006.

McLean, R David, and Jeffrey Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *The Journal of Finance* 71 (1): 5–32. https://doi.org/10.1111/jofi.12365.

Merton, Robert C. 1987. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *The journal of finance* 42, no. 3 (July 1, 1987): 483–510. ISSN: 1540-6261, accessed November 18, 2024. https://doi.org/10.1111/J.1540-6261.1987.TB04565.X.

Molavi, Pooya, Alireza Tahbaz-Salehi, and Andrea Vedolin. 2024. "Model Complexity, Expectations, and Asset Prices." *The Review of Economic Studies* 91, no. 4 (July): 2462–2507. ISSN: 0034-6527. https://doi.org/10.1093/restud/rdad073. eprint: https://academic.oup.com/restud/article-pdf/91/4/2462/57335602/rdad073.pdf.

Muirhead, Robb J. 2005. *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Scheinkman, José A., and Wei Xiong. 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy* 111, no. 6 (December): 1183–1220. ISSN: 0022-3808, accessed December 18, 2025. https://doi.org/10.1086/378531.

Yan, Xuemin Sterling, and Lingling Zheng. 2017. "Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach." *The Review of Financial Studies* 30, no. 4 (April 1, 2017): 1382–1423. ISSN: 0893-9454, accessed November 13, 2022. https://doi.org/10.1093/RFS/HHX001.

Appendix

Lemma A.1 (Projection Identity for Pricing Errors)**.** *Let Λ be an $N \times N$ positive definite matrix and ι the N-vector of ones. Define the market variance $\sigma_m^2 = \iota'\Lambda\iota$, the whitened market direction $m = \Lambda^{1/2}\iota \in \mathbb{R}^N$, and the orthogonal projector*

$$P = I_N - \frac{mm'}{m'm}. \tag{57}$$

As an orthogonal projector onto the complement of m, P is symmetric, idempotent ($P^2 = P$), positive semidefinite, satisfies $Pm = 0$, and has $tr(P) = N - 1$.

Part (i): Quadratic Form Identity. *For any $x \in \mathbb{R}^N$, define $\tilde{x} = \Lambda^{-1/2}x$. Then:*

$$x'\Lambda^{-1}x - \frac{(\iota'x)^2}{\sigma_m^2} = \tilde{x}'P\tilde{x}. \tag{58}$$

Part (ii): Bilinear Extension. *For any $x, y \in \mathbb{R}^N$, with $\tilde{x} = \Lambda^{-1/2}x$ and $\tilde{y} = \Lambda^{-1/2}y$:*

$$x'\Lambda^{-1}y - \frac{(\iota'x)(\iota'y)}{\sigma_m^2} = \tilde{x}'P\tilde{y}. \tag{59}$$

By Cauchy-Schwarz: $|\tilde{x}'P\tilde{y}| \leq \sqrt{\tilde{x}'P\tilde{x}} \cdot \sqrt{\tilde{y}'P\tilde{y}}$.

Part (iii): Expectation Formula. *For a random vector x with $\mathbb{E}[x] = 0$ and $Cov(x) = \Sigma$:*

$$\mathbb{E}\left[x'\Lambda^{-1}x - \frac{(\iota'x)^2}{\sigma_m^2}\right] = tr\left(P \cdot \Lambda^{-1/2}\Sigma\Lambda^{-1/2}\right). \tag{60}$$

Part (iv): Scalar Proportionality. *If $\Sigma = c\Lambda$ for some scalar $c > 0$, then:*

$$\mathbb{E}\left[x'\Lambda^{-1}x - \frac{(\iota'x)^2}{\sigma_m^2}\right] = c(N-1). \tag{61}$$

Proof. **Part (i).** Since Λ is positive definite, $\Lambda^{1/2}$ and $\Lambda^{-1/2}$ exist and are symmetric. For

$\tilde{x} = \Lambda^{-1/2}x$:

$$x'\Lambda^{-1}x = x'\Lambda^{-1/2}\Lambda^{-1/2}x = \tilde{x}'\tilde{x},$$
$$\iota'x = \iota'\Lambda^{1/2}\Lambda^{-1/2}x = m'\tilde{x},$$
$$\sigma_m^2 = \iota'\Lambda\iota = m'm.$$

Therefore:

$$x'\Lambda^{-1}x - \frac{(\iota'x)^2}{\sigma_m^2} = \tilde{x}'\tilde{x} - \frac{(m'\tilde{x})^2}{m'm} \tag{62}$$
$$= \tilde{x}'\left(I_N - \frac{mm'}{m'm}\right)\tilde{x} = \tilde{x}'P\tilde{x}. \tag{63}$$

Part (ii). Apply Part (i) to $x + y$ and $x - y$:

$$(x+y)'\Lambda^{-1}(x+y) - \frac{(\iota'(x+y))^2}{\sigma_m^2} = (\widetilde{x+y})'P(\widetilde{x+y}) \tag{64}$$
$$(x-y)'\Lambda^{-1}(x-y) - \frac{(\iota'(x-y))^2}{\sigma_m^2} = (\widetilde{x-y})'P(\widetilde{x-y}). \tag{65}$$

Subtracting and dividing by 4, the left side becomes:

$$\frac{1}{4}\left[4x'\Lambda^{-1}y - \frac{4(\iota'x)(\iota'y)}{\sigma_m^2}\right] = x'\Lambda^{-1}y - \frac{(\iota'x)(\iota'y)}{\sigma_m^2}. \tag{66}$$

The right side becomes $\tilde{x}'P\tilde{y}$ by linearity of $\widetilde{(\cdot)} = \Lambda^{-1/2}(\cdot)$ and symmetry of P. The Cauchy-Schwarz bound follows from $P \succeq 0$.

Part (iii). For mean-zero u, expanding $u'Au = \sum_{i,j} A_{ij}u_iu_j$ and taking expectations gives $\mathbb{E}[u'Au] = \sum_{i,j} A_{ij}\mathbb{E}[u_iu_j] = \sum_{i,j} A_{ij}\Sigma_{ij} = \mathrm{tr}(A\Sigma)$. Applying this:

$$\mathbb{E}[\tilde{x}'P\tilde{x}] = \mathrm{tr}(P \cdot \mathrm{Cov}(\tilde{x})) = \mathrm{tr}(P \cdot \Lambda^{-1/2}\Sigma\Lambda^{-1/2}). \tag{67}$$

Part (iv). When $\Sigma = c\Lambda$:

$$\operatorname{tr}(P \cdot \Lambda^{-1/2}(c\Lambda)\Lambda^{-1/2}) = c \cdot \operatorname{tr}(P \cdot I_N) = c \cdot \operatorname{tr}(P) = c(N-1). \tag{68}$$

□

Corollary A.1 (Pricing Error Decomposition)**.** *Let $\mu = \varepsilon + \gamma_N \hat{\Lambda}\iota$ be the true expected excess return vector, where $\varepsilon = a - \hat{a}$ is the mean estimation error and $\hat{\Lambda}$ is the perceived covariance. Define $u = \Lambda^{-1/2}\varepsilon$ and $v = \Lambda^{-1/2}\hat{\Lambda}\iota$. Then the pricing error satisfies:*

$$\alpha'\Lambda^{-1}\alpha = (u + \gamma_N v)'P(u + \gamma_N v) = u'Pu + 2\gamma_N u'Pv + \gamma_N^2 v'Pv. \tag{69}$$

Under perfect information ($\hat{\Lambda} = \Lambda$), we have $v = \Lambda^{-1/2}\Lambda\iota = \Lambda^{1/2}\iota = m$, so $Pv = 0$ and all three terms vanish.

6.1 Decomposition of Pricing Errors

Proposition A.1 (Decomposition of Pricing Errors)**.** *The weighted sum of squared pricing errors in the economy, relative to the market portfolio, is given by the difference in squared Sharpe ratios (SR) of the optimal portfolio and the market portfolio. This sum can be decomposed into mean and covariance estimation errors plus an interaction term.*

$$\begin{aligned}
\alpha'\Lambda_t^{-1}\alpha &= SR(w^*)^2 - SR(\tfrac{1}{N}\iota)^2 \\
&= \underbrace{\left[\varepsilon'\,\Lambda_t^{-1}\,\varepsilon - \frac{(\iota'\varepsilon)^2}{\sigma_{m,t}^2}\right]}_{\textit{Mean estimation errors}} \\
&\quad + \underbrace{2\gamma_N\left[\varepsilon'\Lambda_t^{-1}\hat{\Lambda}_t\iota - \frac{(\iota'\varepsilon)\hat{\sigma}_{m,t}^2}{\sigma_{m,t}^2}\right]}_{\textit{Interaction term}} \\
&\quad + \underbrace{\gamma_N^2\left[\iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota - \frac{(\hat{\sigma}_{m,t}^2)^2}{\sigma_{m,t}^2}\right]}_{\textit{Covariance estimation errors}}
\end{aligned} \tag{70}$$

Proof. We show that $\alpha'\Lambda_t^{-1}\alpha = SR(w^*)^2 - SR(\frac{1}{N}\iota)^2$ and derive the three-term decomposition.

Step 1: Pricing Errors as Sharpe Ratio Difference By Lemma A.1(i), for any vector x:

$$x'\Lambda_t^{-1}x - \frac{(\iota'x)^2}{\sigma_{m,t}^2} = \tilde{x}'P\tilde{x}$$

where $\tilde{x} = \Lambda_t^{-1/2}x$ and $P = I_N - mm'/(m'm)$ with $m = \Lambda_t^{1/2}\iota$.

Taking $x = \mu_t$ and noting that $\alpha = \mu_t - \Lambda_t\iota(\iota'\mu_t)/\sigma_{m,t}^2$ is the residual from projecting μ_t onto $\Lambda_t\iota$, we have $\alpha'\Lambda_t^{-1}\alpha = \tilde{\mu}_t'P\tilde{\mu}_t$. The lemma identity then gives:

$$\alpha'\Lambda_t^{-1}\alpha = \mu_t'\Lambda_t^{-1}\mu_t - \frac{(\iota'\mu_t)^2}{\sigma_{m,t}^2}.$$

For the optimal portfolio $w^* \propto \Lambda_t^{-1}\mu_t$, standard mean-variance optimization yields $SR(w^*)^2 = \mu_t'\Lambda_t^{-1}\mu_t$. For the market portfolio $w_m = \frac{1}{N}\iota$, the squared Sharpe ratio is $SR(\frac{1}{N}\iota)^2 = (\iota'\mu_t)^2/\sigma_{m,t}^2$. Therefore:

$$\boxed{\alpha'\Lambda_t^{-1}\alpha = SR(w^*)^2 - SR\left(\frac{1}{N}\iota\right)^2}$$

Step 2: Decomposition into Mean, Covariance, and Interaction Terms Substituting $\mu_t = \varepsilon + \gamma_N\hat{\Lambda}_t\iota$ from Equation 21, where $\varepsilon = a_t - \hat{a}_t$ is the mean estimation error, into the projection identity gives:

$$\alpha'\Lambda_t^{-1}\alpha = (\varepsilon + \gamma_N\hat{\Lambda}_t\iota)'\Lambda_t^{-1}(\varepsilon + \gamma_N\hat{\Lambda}_t\iota) - \frac{(\iota'\varepsilon + \gamma_N\hat{\sigma}_{m,t}^2)^2}{\sigma_{m,t}^2}$$

where $\hat{\sigma}_{m,t}^2 = \iota'\hat{\Lambda}_t\iota$ is the perceived market variance.

Expanding and grouping by powers of γ_N:

$$\boxed{\alpha'\Lambda_t^{-1}\alpha = \underbrace{\varepsilon'\Lambda_t^{-1}\varepsilon - \frac{(\iota'\varepsilon)^2}{\sigma_{m,t}^2}}_{\text{Term 1 (mean errors)}} + \underbrace{2\gamma_N\left[\varepsilon'\Lambda_t^{-1}\hat{\Lambda}_t\iota - \frac{(\iota'\varepsilon)\hat{\sigma}_{m,t}^2}{\sigma_{m,t}^2}\right]}_{\text{Term 2 (cross term)}} + \underbrace{\gamma_N^2\left[\iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota - \frac{(\hat{\sigma}_{m,t}^2)^2}{\sigma_{m,t}^2}\right]}_{\text{Term 3 (covariance errors)}}}$$

By Lemma A.1(i), Term 1 equals $u'Pu$ where $u = \Lambda_t^{-1/2}\varepsilon$, and Term 3 equals $\gamma_N^2 v'Pv$ where

$v = \Lambda_t^{-1/2}\hat{\Lambda}_t\iota$. By Lemma A.1(ii), Term 2 equals $2\gamma_N u'Pv$. Under perfect information ($\hat{\Lambda}_t = \Lambda_t$), we have $v = \Lambda_t^{1/2}\iota = m$, so $Pv = 0$ and all terms vanish.

□

6.2 Perfect Information

Proposition A.2 (Perfect Information)**.** *When agents know the economy's parameters, CAPM and Arbitrage Pricing Theory (APT) hold. In particular,*

1. *Mean returns are given by $\mu_t = \gamma_N \Lambda_t \iota$.*
2. *The mean-variance efficient portfolio is the market portfolio with weights of $\frac{1}{N}$.*
3. *Pricing errors with respect to the market portfolio are zero, $\alpha_t = 0$.*
4. *For $N \to \infty$, and rotating the factors to make them orthogonal, the price of risk for factor k is given by $\gamma \bar{b}_{t,k}\sigma^2_{t,k}$, where $\bar{b}_{t,k}$ is the average loading on factor k and $\sigma_{t,k}$ is the factor's volatility.*
5. *Systematic factors asymptotically span the maximum Sharpe ratio: $SR_S^2/SR_{all}^2 \to 1$ as $N \to \infty$, where $SR_{all}^2 = \mu_t'\Lambda_t^{-1}\mu_t$ is the maximum squared Sharpe ratio using all N assets and SR_S^2 uses only S factor-mimicking portfolios. This requires that factor loadings satisfy the regularity condition $\frac{1}{N}(B^{sys})'B^{sys} \to G$ for some positive definite matrix G as $N \to \infty$.*

Proof. When agents have perfect information, their subjective estimates match the true values: $\hat{a}_t = a_t = B_t\pi$ and $\hat{\Lambda}_t = \Lambda_t = B_t V B_t'$.

From Equation 23, the general expression for expected returns is:

$$\mu_t = (a_t - \hat{a}_t) + \gamma_N(\hat{\Lambda}_t - \Lambda_t)\iota + \gamma_N \Lambda_t \iota \tag{71}$$

Under perfect information, $a_t = \hat{a}_t$ and $\hat{\Lambda}_t = \Lambda_t$, so both estimation error terms vanish:

$$\mu_t = 0 + 0 + \gamma_N \Lambda_t \iota = \gamma_N \Lambda_t \iota \tag{72}$$

From Equation 24, the objectively optimal portfolio is:

$$w_t^* = \frac{1}{\gamma}\Lambda_t^{-1}\mu_t = \frac{1}{\gamma}\Lambda_t^{-1}(\gamma_N \Lambda_t \iota) = \frac{\gamma_N}{\gamma}\iota = \frac{1}{N}\iota \tag{73}$$

This confirms that the optimal portfolio equals the market portfolio exactly, establishing mean-variance efficiency. From Proposition 1, since the optimal and market portfolios coincide, $\alpha'\Lambda_t^{-1}\alpha = SR(w^*)^2 - SR(\frac{1}{N}\iota)^2 = 0$, which implies $\alpha = 0$. Therefore, there are no pricing errors relative to the market portfolio.

Since the market portfolio is mean-variance efficient, CAPM follows by the standard equivalence between market efficiency and the security market line.

APT Verification. For APT, we need to show that expected excess returns are a linear function of factor loadings when idiosyncratic risks are diversified away as N grows large.

Without loss of generality, assume uncorrelated factors. This is achieved by redefining $\tilde{x}_{t+1} = Q'x_{t+1}$ where Q diagonalizes $\text{Var}(x_{t+1})$, and correspondingly $\tilde{b}_{i,t} = Q'b_{i,t}$. Under this rotation, the covariance matrix can be expressed as:

$$\Lambda_t = \sum_{k=1}^{S} \sigma_k^2 b_{.,k} b'_{.,k} + \text{diag}(c_{1,t}^2\sigma_{z1}^2, \ldots, c_{N,t}^2\sigma_{zN}^2) \tag{74}$$

where $b_{.,k} = (b_{1,t,k}, \ldots, b_{N,t,k})'$ is the vector of loadings on factor k, σ_k^2 is the variance of factor k, and $c_{i,t}^2\sigma_{zi}^2$ represents the idiosyncratic variance for asset i (where $\sigma_{zi}^2 = \text{Var}(z_{i,t+1})$).

For asset i, the risk compensation term becomes:

$$(\Lambda_t \iota)_i = N \sum_{k=1}^{S} \sigma_k^2 \bar{b}_{t,k} b_{i,t,k} + c_{i,t}^2\sigma_{zi}^2 \tag{75}$$

where $\bar{b}_{t,k} = \frac{1}{N}\sum_{i=1}^{N} b_{i,t,k}$ is the average loading on factor k.

The expected excess return is therefore:

$$\mu_{i,t} = \gamma_N(\Lambda_t \iota)_i = \gamma \sum_{k=1}^{S} \sigma_k^2 \bar{b}_{t,k} b_{i,t,k} + \frac{\gamma}{N} c_{i,t}^2 \sigma_{zi}^2 \tag{76}$$

As $N \to \infty$, if idiosyncratic variances remain bounded, the term $\frac{\gamma}{N} c_{i,t}^2 \sigma_{zi}^2 \to 0$, and:

$$\mu_{i,t} \approx \gamma \sum_{k=1}^{S} \sigma_k^2 \bar{b}_{t,k} b_{i,t,k} = \sum_{k=1}^{S} \lambda_k b_{i,t,k} \tag{77}$$

where $\lambda_k = \gamma \sigma_k^2 \bar{b}_{t,k}$ is the risk premium for factor k.

This confirms that APT holds in the limit as N increases: expected returns are a linear function of factor loadings, with risk premia determined by factor variances, average exposures, and risk aversion.

APT Spanning of Maximum Sharpe Ratio. We show that systematic factors asymptotically span the maximum Sharpe ratio: $SR_S^2 / SR_{all}^2 \to 1$ as $N \to \infty$.

Under perfect information, $\mu_t = \gamma_N \Lambda_t \iota$, so $SR_{all}^2 = \mu_t' \Lambda_t^{-1} \mu_t = \gamma_N^2 \iota' \Lambda_t \iota$. Decompose $\Lambda_t = B^{sys} V^{sys} (B^{sys})' + D$ where $D = \text{diag}(c_i^2 \sigma_{z_i}^2)$ is idiosyncratic variance.

Define factor-mimicking portfolios $Q = B^{sys} M^{-1}$ where $M = (B^{sys})' B^{sys}$. These have unit exposure to each factor: $(B^{sys})' Q = M M^{-1} = I_S$. The maximum SR using only these S portfolios is $SR_S^2 = \mu_t' Q [Q' \Lambda_t Q]^{-1} Q' \mu_t$.

For the covariance of factor-mimicking portfolios:

$$Q' \Lambda_t Q = M^{-1} (B^{sys})' (B^{sys} V^{sys} (B^{sys})' + D) B^{sys} M^{-1} \tag{78}$$

$$= V^{sys} + M^{-1} (B^{sys})' D B^{sys} M^{-1}. \tag{79}$$

The second term is $O(1/N)$: with bounded loadings and bounded idiosyncratic variances, $[(B^{sys})' D B^{sys}]_{kl} = \sum_i b_{i,k} b_{i,l} c_i^2 \sigma_{zi}^2 = O(N)$, and $\|M^{-1}\|_{\text{op}} = O(1/N)$ since $M = (B^{sys})' B^{sys}$ has eigenvalues $\sim O(N)$. The $\sim O(N)$ eigenvalue condition holds when $\frac{1}{N}(B^{sys})' B^{sys} \to G$ for some positive definite G as

$N \to \infty$, which requires that factor loadings do not concentrate on a lower-dimensional subspace, a standard regularity condition in the APT literature (e.g., Chamberlain and Rothschild 1983).

For the expected return projection:

$$Q'\mu_t = \gamma_N M^{-1}(B^{sys})'\Lambda_t\iota = \gamma_N V^{sys}(B^{sys})'\iota + \gamma_N M^{-1}(B^{sys})'D\iota. \tag{80}$$

The second term is $O(1/N)$: $(B^{sys})'D\iota$ is $O(N)$, M^{-1} contributes $O(1/N)$, and $\gamma_N = \gamma/N$.

Substituting into $SR_S^2 = (Q'\mu_t)'[Q'\Lambda_t Q]^{-1}(Q'\mu_t)$ and using $[V^{sys} + O(1/N)]^{-1} = (V^{sys})^{-1} + O(1/N)$ (valid for N large enough that $\|O(1/N)\| < \lambda_{\min}(V^{sys})$, by Neumann series):

$$SR_S^2 = \gamma_N^2 \iota' B^{sys} V^{sys} (B^{sys})'\iota + O(1/N). \tag{81}$$

Since $\Lambda_t = B^{sys}V^{sys}(B^{sys})' + D$, we have $\iota'\Lambda_t\iota - \iota' B^{sys}V^{sys}(B^{sys})'\iota = \iota' D\iota = \sum_i c_i^2\sigma_{z_i}^2$, so:

$$SR_{all}^2 - SR_S^2 = \gamma_N^2 \sum_{i=1}^{N} c_i^2\sigma_{z_i}^2 + O(1/N) = O(1/N). \tag{82}$$

The gap vanishes because idiosyncratic variance $\sum_i c_i^2\sigma_{z_i}^2 = O(N)$ is scaled by $\gamma_N^2 = \gamma^2/N^2$, yielding $O(1/N)$. Under the strong factor structure assumption where $\iota'\Lambda_t\iota \sim O(N^2)$, the maximum squared Sharpe ratio $SR_{all}^2 = \gamma_N^2\iota'\Lambda_t\iota \sim O(1)$ is bounded. Dividing gives $SR_S^2/SR_{all}^2 = 1 - O(1/N) \to 1$. □

6.3 Known Means and Unknown Covariances

Proposition A.3 (Known Means, Unknown Covariances)**.** *When agents know the expected value of dividends, a, but have uncertainty about the covariance, Λ, neither CAPM nor APT hold. Further, expected pricing errors scale as $O(N/t)$ for large N. In particular:*

1. *Mean returns are given by $\mu_t = \gamma_N(\hat{\Lambda}_t - \Lambda_t)\iota + \gamma_N\Lambda_t\iota$, and hence APT fails.*

2. *CAPM fails, and the mean-variance efficient portfolio weights are given by $\frac{1}{N}\Lambda_t^{-1}(\hat{\Lambda}_t - \Lambda_t)\iota +$*

$\frac{1}{N}\iota$.

3. *Pricing errors are given by* $\alpha' \Lambda_t^{-1} \alpha = (\gamma_N)^2 \left[\iota' \hat{\Lambda}_t \Lambda_t^{-1} \hat{\Lambda}_t \iota - \frac{(\hat{\sigma}^2_{m,t})^2}{\sigma^2_{m,t}} \right]$.

4. *The expected pricing errors are given by:*

$$\mathbb{E}[\alpha_t' \Lambda^{-1} \alpha_t] = \frac{\gamma^2}{N^2} \left(\frac{1}{\nu_t - 2} \right)^2 \left[A(V_0) + t(N-1) \iota' \Lambda \iota \right] \tag{83}$$

where $\nu_t = \nu_0 + t - (S+N) + 1 > 2$ *for the expectation to exist, and*

$$A(V_0) = \iota' B_t V_0 B_t' \left(\Lambda^{-1} - \frac{\iota \iota'}{\iota' \Lambda \iota} \right) B_t V_0 B_t' \iota \geq 0$$

measures the contribution from the prior scale matrix V_0.

Moreover, since $A(V_0) \geq 0$ *with equality if and only if* $B_t V_0 B_t' \iota \propto \Lambda \iota$ *(which holds in particular when* $V_0 = V$*), expected pricing errors are minimized when the prior scale matrix* V_0 *correctly captures the aggregate contribution to market variance. In the specific case of a correct prior* $V_0 = V$ *and minimal degrees of freedom* $\nu_0 = S + N$*, the expected pricing errors reach this lower bound (provided* $\nu_t > 2$*, i.e.,* $t \geq 2$ *with minimal* ν_0*):*

$$\mathbb{E}[\alpha_t' \Lambda^{-1} \alpha_t] \big|_{V_0 = V, \nu_0 = S+N} = \frac{\gamma^2}{N^2} \left(\frac{1}{t-1} \right)^2 t(N-1)(\iota' \Lambda \iota) \tag{84}$$

For large N*, large* t*, and assuming* $\iota' \Lambda \iota = \sigma^2_{m,t} \sim O(N^2)$ *(e.g., due to strong common factor exposures), this minimum error scales as* $O\left(\frac{N}{t}\right)$.

Proof. When agents know the mean ($a_t = \hat{a}_t$), the mean estimation error $\varepsilon = 0$. From Proposition 1 (Equation 26), only the covariance estimation term remains:

$$\alpha_t' \Lambda^{-1} \alpha_t = \frac{\gamma^2}{N^2} \left(\iota' \hat{\Lambda} \Lambda^{-1} \hat{\Lambda} \iota - \frac{(\iota' \hat{\Lambda} \iota)^2}{\iota' \Lambda \iota} \right) \tag{85}$$

We drop the t subscript on Λ for notational ease throughout this proof. Recall the posterior degrees

of freedom $\nu_t = \nu_0 + t - (S+N) + 1$, where ν_0 is the prior degrees of freedom and $S+N$ is the dimension of the state vector X.

Wishart aggregation. Since $X_\tau - \pi \sim N(0, V)$ independently, the sample scatter matrix follows a Wishart distribution:

$$S_\pi := \sum_{\tau=1}^{t} (X_\tau - \pi)(X_\tau - \pi)' \sim \mathcal{W}_{S+N}(t, V) \tag{86}$$

By linear transformation, $S := BS_\pi B' \sim \mathcal{W}_N(t, \Lambda)$ where $\Lambda = BVB'$.

From the Bayesian updating equations (16)–(18), the estimated covariance is:

$$\hat{\Lambda} = \frac{1}{\nu_t - 2} B(V_0 + S_\pi)B' = \frac{1}{\nu_t - 2}(M + S) \tag{87}$$

where $M = BV_0B'$.

Wishart moment identities. For $S \sim \mathcal{W}_N(t, \Lambda)$ and any vector $u \in \mathbb{R}^N$, standard results give:

$$\mathbb{E}[u'S\Lambda^{-1}Su] = t(t + N + 1)(u'\Lambda u) \tag{88}$$

$$\mathbb{E}[(u'Su)^2] = t(t+2)(u'\Lambda u)^2 \tag{89}$$

These follow from the fourth-moment formula $\mathbb{E}[S_{ij}S_{kl}] = t(\Lambda_{ik}\Lambda_{jl} + \Lambda_{il}\Lambda_{jk}) + t^2\Lambda_{ij}\Lambda_{kl}$ (see, e.g., Muirhead (2005, Theorem 3.3.15)).

Computing the expectations. Let $\sigma_m^2 = \iota'\Lambda\iota$. Using the independence of M (deterministic) and S:

$$\mathbb{E}[\iota'(M+S)\Lambda^{-1}(M+S)\iota] = \iota'M\Lambda^{-1}M\iota + 2t(\iota'M\iota) + t(t+N+1)\sigma_m^2 \tag{90}$$

$$\mathbb{E}[(\iota'(M+S)\iota)^2] = (\iota'M\iota)^2 + 2t(\iota'M\iota)\sigma_m^2 + t(t+2)\sigma_m^4 \tag{91}$$

Substituting into (85) and simplifying:

$$\mathbb{E}[\alpha'\Lambda^{-1}\alpha] = \frac{\gamma^2}{N^2(\nu_t - 2)^2}\left[A(V_0) + t(N-1)\sigma_m^2\right] \tag{92}$$

where $A(V_0) := \iota' M \Lambda^{-1} M \iota - (\iota' M \iota)^2 / \sigma_m^2$ measures prior misspecification through $M = BV_0B'$.

Correct prior case. By Lemma A.1 with $x = M\iota$:

$$A(V_0) = (M\iota)' \Lambda^{-1} (M\iota) - \frac{(\iota' M \iota)^2}{\sigma_m^2} = \tilde{z}' P \tilde{z} \tag{93}$$

where $\tilde{z} = \Lambda^{-1/2} M \iota$ and $P = I_N - mm'/(m'm)$ with $m = \Lambda^{1/2}\iota$. When $V_0 = V$, we have $M = \Lambda$, so $\tilde{z} = \Lambda^{1/2}\iota = m$ and $P\tilde{z} = 0$. Thus $A(V) = 0$ and:

$$\mathbb{E}[\alpha' \Lambda^{-1} \alpha]\big|_{V_0 = V} = \frac{\gamma^2}{N^2 (\nu_t - 2)^2} \cdot t(N-1)\sigma_m^2 \tag{94}$$

With minimal prior $\nu_0 = S + N$ (requiring $t \geq 2$ for $\nu_t > 2$):

$$\mathbb{E}[\alpha' \Lambda^{-1} \alpha]\big|_{V_0 = V} = \frac{\gamma^2}{N^2} \cdot \frac{t(N-1)\sigma_m^2}{(t-1)^2} \tag{95}$$

Asymptotic scaling. Under standard factor structure assumptions where factor loadings and idiosyncratic variances are $O(1)$ with S fixed, the market variance satisfies $\sigma_m^2 = \iota' \Lambda \iota$, which is of order N^2. This holds when assets have $O(1)$ exposure to common factors. The expected pricing error then scales as:

$$\frac{t(N-1)N^2}{N^2(t-1)^2} \asymp \frac{N}{t}. \tag{96}$$

Incorrect prior case. Recall $A(V_0)$ measures how the prior covariance $M = BV_0B'$ differs from the true Λ. Since $P \succeq 0$, we have $A(V_0) = \tilde{z}' P \tilde{z} \geq 0$ for any V_0. Equality holds iff $\tilde{z} \in \text{span}(m)$, i.e., $M\iota \propto \Lambda\iota$.

Therefore:

$$\mathbb{E}[\alpha' \Lambda^{-1} \alpha]\big|_{V_0 \neq V} \geq \mathbb{E}[\alpha' \Lambda^{-1} \alpha]\big|_{V_0 = V} \tag{97}$$

Pricing errors are higher when agents have incorrect priors about the covariance structure. Economically, the equality condition $M\iota \propto \Lambda\iota$ means the prior is wrong about individual asset covariances but correctly captures their aggregate contribution to market variance. □

6.4 Unknown Means and Known Covariances

Proposition A.4 (Unknown Means, Known Covariances)**.** *If investors have perfect information about covariances but only the correct prior about the expectation, neither CAPM nor APT hold. In particular:*

1. *Mean returns are given by* $\mu_t = (a_t - \hat{a}_t) + \gamma_N \Lambda_t \iota$*, and hence APT fails.*
2. *CAPM fails, and the mean-variance efficient portfolio weights are given by* $\frac{1}{\gamma}\Lambda_t^{-1}(a_t - \hat{a}_t) + \frac{1}{N}\iota$*.*
3. *Pricing errors are given by* $\alpha'\Lambda_t^{-1}\alpha = \varepsilon'\Lambda_t^{-1}\varepsilon - \frac{1}{\sigma_m^2}(\iota'\varepsilon)^2$*, where* $\varepsilon = a_t - \hat{a}_t$*.*

Moreover, the average pricing errors relative to the market portfolio increase linearly with the number of assets N and decrease with the number of observations t:

$$\mathbb{E}[\alpha_t'\Lambda_t^{-1}\alpha_t] = \frac{t(N-1)}{(t+\kappa_0)^2} \sim O\left(\frac{N}{t}\right). \tag{98}$$

This bound holds exactly for a correct prior ($\pi_0 = \pi$) and serves as a lower bound when the prior is incorrect ($\pi_0 \neq \pi$), as the proof below establishes.

Proof. When agents know the covariance matrix Λ_t, the perceived covariance $\hat{\Lambda}_t = \varphi\Lambda_t$ is proportional to Λ_t (where $\varphi = (t+\kappa_0+1)/(t+\kappa_0)$ reflects parameter uncertainty about π). Corollary A.1 gives $Pv = 0$, so only the mean estimation term remains:

$$\alpha'\Lambda_t^{-1}\alpha = u'Pu \tag{99}$$

where $u = \Lambda_t^{-1/2}\varepsilon$ and $\varepsilon = a_t - \hat{a}_t$.

With the correct prior $\pi_0 = \pi$, the posterior mean is $\hat{\pi}_t = \frac{\kappa_0\pi + t\bar{X}_t}{t+\kappa_0}$, giving estimation error:

$$\varepsilon = B_t(\pi - \hat{\pi}_t) = B_t\frac{t}{t+\kappa_0}(\pi - \bar{X}_t). \tag{100}$$

Since $\pi - \bar{X}_t \sim N(0, V/t)$ and $\Lambda_t = B_t V B_t'$:

$$\text{Cov}(\varepsilon) = \frac{t^2}{(t+\kappa_0)^2} B_t \frac{V}{t} B_t' = \frac{t}{(t+\kappa_0)^2} \Lambda_t. \tag{101}$$

By Lemma A.1 Part (iv) with $c = \frac{t}{(t+\kappa_0)^2}$:

$$\mathbb{E}[\alpha' \Lambda_t^{-1} \alpha] = \frac{t(N-1)}{(t+\kappa_0)^2}. \tag{102}$$

For large t, this scales as $O(N/t)$. □

Incorrect prior case. If $\pi_0 \neq \pi$, then $\mathbb{E}[\alpha' \Lambda_t^{-1} \alpha] \geq \frac{t(N-1)}{(t+\kappa_0)^2}$.

Proof. The estimation error decomposes as $\varepsilon = \varepsilon_1 + \varepsilon_2$ where:

- $\varepsilon_1 = B_t \frac{t}{t+\kappa_0} (\pi - \bar{X}_t)$ is sampling error (mean-zero)
- $\varepsilon_2 = B_t \frac{\kappa_0}{t+\kappa_0} (\pi - \pi_0)$ is prior bias (deterministic)

Let $u_j = \Lambda_t^{-1/2} \varepsilon_j$. Since $\mathbb{E}[u_1] = 0$ and u_2 is deterministic, the cross-term vanishes: $\mathbb{E}[u_1' P u_2] = \mathbb{E}[u_1]' P u_2 = 0$. Since $P \succeq 0$, the bias term satisfies $u_2' P u_2 \geq 0$. Therefore:

$$\mathbb{E}[u' P u] = \mathbb{E}[u_1' P u_1] + u_2' P u_2 \geq \mathbb{E}[u_1' P u_1] = \frac{t(N-1)}{(t+\kappa_0)^2} \tag{103}$$

Equality holds iff $P u_2 = 0$, i.e., $\varepsilon_2 \propto \Lambda_t \iota$. This means the prior is wrong about the overall level of expected returns but correctly captures cross-sectional differences. □

6.5 Time-Varying Means: Persistent Pricing Errors

Proof of Corollary 1. With known V, only mean estimation errors matter (Corollary A.1 with $Pv = 0$). Let P_t denote the Kalman filter posterior (filtering) covariance of π_t given $\mathcal{I}_t$. The filtering error $e_t = \pi_t - \hat{\pi}_t$ satisfies $e_t \mid \mathcal{I}_t \sim N(0, P_t)$.

Step 1: Scalar reduction of the Riccati recursion. The Kalman filter prediction and update steps give

$$P_{t|t-1} = \rho^2 P_{t-1} + \Sigma_\eta, \qquad P_t = P_{t|t-1} - P_{t|t-1}(P_{t|t-1} + V)^{-1} P_{t|t-1}.$$

Under scalar proportionality ($\Sigma_\eta = \sigma_\eta^2 V$), if $P_{t-1} = \sigma_{t-1}^2 V$ then $P_{t|t-1} = (\rho^2 \sigma_{t-1}^2 + \sigma_\eta^2)V =: s_t V$. Substituting into the update step,

$$P_t = s_t V - s_t V(s_t V + V)^{-1} s_t V = s_t V - \frac{s_t^2}{s_t + 1} V = \frac{s_t}{s_t + 1} V,$$

so $P_t = \sigma_t^2 V$ with $\sigma_t^2 = s_t/(s_t + 1)$. By induction, scalar proportionality is preserved at every step.

Step 2: Steady-state convergence. Setting $\sigma_p^2 = s/(s+1)$ with $s = \rho^2 \sigma_p^2 + \sigma_\eta^2$ and rearranging yields

$$\rho^2 (\sigma_p^2)^2 + (1 - \rho^2 + \sigma_\eta^2)\sigma_p^2 - \sigma_\eta^2 = 0.$$

The quadratic has a unique positive root because the constant term $-\sigma_\eta^2 < 0$ and the leading coefficient $\rho^2 > 0$. The map $g(x) = (\rho^2 x + \sigma_\eta^2)/(\rho^2 x + \sigma_\eta^2 + 1)$ satisfies $g'(x) = \rho^2/(\rho^2 x + \sigma_\eta^2 + 1)^2 \leq \rho^2 < 1$, so it is a contraction. Since $g(\sigma_p^2) = \sigma_p^2$ and g maps $[0, \sigma_p^2]$ into itself, $\sigma_t^2 \to \sigma_p^2$ and $P_t \to P_\infty = \sigma_p^2 V \succ 0$.

Step 3: Pricing errors. Since the true conditional expectation is $\mathbb{E}[\pi_{t+1}|\mathcal{F}_t] = \rho \pi_t$ and agents forecast $\rho \hat{\pi}_t$, the pricing-relevant estimation error is $\varepsilon_t = B_t \rho e_t$ (the innovation η_{t+1} cancels as it is unknown under both information sets). Its covariance is

$$\mathrm{Cov}(\varepsilon_t) = \rho^2 B_t P_\infty B_t' = \rho^2 \sigma_p^2 \Lambda_t = \kappa_\infty \Lambda_t.$$

By Lemma A.1 Part (iv) with $c = \kappa_\infty$: $\mathbb{E}[\alpha' \Lambda^{-1} \alpha] = \kappa_\infty (N - 1)$. □

6.6 Unknown Means and Unknown Covariances

Proposition A.5 (Pricing Errors with Unknown Means and Covariances)**.** *When agents must estimate both the mean vector π and the covariance matrix V of the state variables X_{t+1}, neither*

CAPM nor APT hold. Estimating both mean and covariance generates pricing errors that scale with market complexity. In particular:

1. *Mean returns are given by the full decomposition including mean and covariance estimation errors:*

$$\mu_t = (a_t - \hat{a}_t) + \gamma_N(\hat{\Lambda}_t - \Lambda_t)\iota + \gamma_N \Lambda_t \iota.$$

   Hence, APT fails as returns depend on estimation errors.

2. *CAPM fails, and the objectively optimal mean-variance efficient portfolio weights deviate from the market portfolio due to both estimation errors:*

$$w_t^* = \frac{1}{\gamma}\Lambda_t^{-1}(a_t - \hat{a}_t) + \frac{1}{N}\Lambda_t^{-1}(\hat{\Lambda}_t - \Lambda_t)\iota + \frac{1}{N}\iota.$$

3. *Pricing errors relative to the market portfolio are given by the full decomposition including mean, covariance, and interaction terms:*

$$\alpha' \Lambda_t^{-1} \alpha = SR(w^*)^2 - SR(\frac{1}{N}\iota)^2.$$

4. *Under correctly specified priors ($\pi_0 = \pi$, $V_0 = V$) with bounded strength ($\kappa_0 \leq K$) and standard factor structure ($\sigma_m^2 = \iota' \Lambda \iota \sim O(N^2)$), the expected weighted squared pricing errors satisfy:*

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] \sim O\left(\frac{N}{t}\right), \tag{104}$$

   establishing N/t as a fundamental benchmark that cannot be improved through better prior specification within this class.

When both means and covariances are unknown, the pricing error decomposition from Proposition 1 includes mean estimation errors (Term A), covariance estimation errors (Term C), and an interaction term (Term B). With correctly specified priors ($\pi_0 = \pi$, $V_0 = V$), the interaction

term vanishes in expectation ($\mathbb{E}[B] = 0$) by sign symmetry: the mean estimation error ε is an odd function of the centered innovations $z_\tau = X_\tau - \pi$, while the covariance estimate $\hat{\Lambda}_t$ depends on z_τ only through quadratic forms and is therefore an even function. Under the sign flip $z_\tau \mapsto -z_\tau$ (which preserves the distribution when $\pi_0 = \pi$), the interaction $u'Pv$ flips sign, so its expectation is zero. In this case, the total pricing errors decompose exactly into the sum of mean and covariance estimation errors:

$$E[\alpha'\Lambda_t^{-1}\alpha] = E[\varepsilon'\Lambda_t^{-1}\varepsilon] + E[\frac{\gamma^2}{N^2}\iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota] - E[\frac{1}{\sigma_m^2}(\frac{\gamma^2}{N^2}(\hat{\sigma}_m^2)^2)] \tag{105}$$

$$= \underbrace{E[\varepsilon'\Lambda_t^{-1}\varepsilon] - E[\frac{1}{\sigma_m^2}(\iota'\varepsilon)^2]}_{\alpha \text{ from unknown means}} + \underbrace{E[\frac{\gamma^2}{N^2}\iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota] - E[\frac{\gamma^2}{N^2}\frac{(\iota'\hat{\Lambda}_t\iota)^2}{\iota'\Lambda_t\iota}]}_{\alpha \text{ from unknown covariances}}. \tag{106}$$

(107)

where the equality (rather than an approximation) holds exactly when $\pi_0 = \pi$ and $V_0 = V$. This result also serves as a lower bound for any prior specification.

Proof. We establish $\sim O(N/t)$ scaling under correctly specified prior parameters ($\pi_0 = \pi$, $V_0 = V$) with bounded strength.

Projection structure. Let $m = \Lambda_t^{1/2}\iota/\sigma_{m,t}$ (the unit vector in the direction of $\Lambda_t^{1/2}\iota$) and $P = I_N - mm'$ (the orthogonal projection onto $m^\perp$). By Corollary A.1, with $u = \Lambda_t^{-1/2}\varepsilon$ and $v = \Lambda_t^{-1/2}\hat{\Lambda}_t\iota$:

$$\alpha'\Lambda_t^{-1}\alpha = \underbrace{u'Pu}_{A} + \underbrace{2\gamma_N u'Pv}_{B} + \underbrace{\gamma_N^2 v'Pv}_{C} \tag{108}$$

Since $P \succeq 0$: $A, C \geq 0$ and $|B| \leq 2\sqrt{AC}$ by Cauchy-Schwarz (Lemma A.1).

Sign symmetry: $\mathbb{E}[B] = 0$. Assume $\pi_0 = \pi$. Let $z_\tau := X_\tau - \pi$. Under the true DGP, $(z_1, \ldots, z_t) \overset{d}{=} (-z_1, \ldots, -z_t)$ since $z_\tau \overset{iid}{\sim} N(0, V)$.

Under this sign flip:

- The posterior mean $\hat{\pi}_t = \pi + \frac{t}{t+\kappa_0}\bar{z}$ is linear in $\bar{z}$. Since $\bar{z} \mapsto -\bar{z}$, the mean estimation error $\varepsilon = a_t - \hat{a}_t = -\frac{t}{t+\kappa_0}B_t\bar{z}$ flips sign, hence $u \mapsto -u$.

- The NIW posterior scale matrix $V_0 + \sum_\tau (X_\tau - \bar{X}_t)(X_\tau - \bar{X}_t)' + \frac{\kappa_0 t}{\kappa_0+t}(\bar{X}_t - \pi_0)(\bar{X}_t - \pi_0)'$ depends on z_τ only through quadratic forms: the centered scatter matrix $\sum_\tau (z_\tau - \bar{z})(z_\tau - \bar{z})'$ and $\bar{z}\bar{z}'$, both invariant under $z_\tau \mapsto -z_\tau$. Hence $\hat{\Lambda}_t$ and v are unchanged.

Therefore $u'Pv \mapsto -u'Pv$, and by symmetry $\mathbb{E}[B] = 0$. The argument works despite the joint dependence of $\hat{\pi}_t$ and $\hat{V}_t$ in the NIW posterior because the posterior mean is an odd function of the innovations while the posterior scale matrix is an even function.

Term A (mean error). From Proposition 4:

$$\mathbb{E}[A] = \frac{t(N-1)}{(t+\kappa_0)^2} \tag{109}$$

Term C (covariance error). We compute $\mathbb{E}[C] = \gamma_N^2\,\mathbb{E}[v'Pv]$ exactly. The NIW posterior scale matrix decomposes as

$$V_0 + \underbrace{\sum_{\tau=1}^{t}(X_\tau - \bar{X}_t)(X_\tau - \bar{X}_t)'}_{\equiv S_c \sim \mathcal{W}_{S+N}(t-1,V)} + \underbrace{\frac{\kappa_0 t}{\kappa_0 + t}\bar{z}\bar{z}'}_{\equiv R} \tag{110}$$

where $\bar{z} = \bar{X}_t - \pi$. The Wishart $\mathcal{W}_{S+N}(t-1, V)$ requires $t \geq 2$; the assumption $\nu_t = \nu_0 + t - (S+N) + 1 > 2$ with $\nu_0 \geq S + N$ ensures $t \geq 2$. By Cochran's theorem, $S_c \perp R$. The estimated return covariance is $\hat{\Lambda}_t = c_t(M + \tilde{S} + \tilde{R})$ where $M = B_t V_0 B_t'$, $\tilde{S} = B_t S_c B_t' \sim \mathcal{W}_N(t-1, \Lambda_t)$, $\tilde{R} = B_t R B_t' = \frac{\kappa_0 t}{\kappa_0+t}\eta\eta'$ with $\eta = B_t\bar{z} \sim N(0, \Lambda_t/t)$, and $c_t = \frac{\kappa_0+t+1}{(\kappa_0+t)(\nu_t-2)}$, which combines the predictive variance inflation $(\kappa_0 + t + 1)/(\kappa_0 + t)$ from integrating over the posterior mean with the inverse-Wishart posterior mean divisor $\nu_t - 2 = \nu_0 + t - (S+N) - 1$ from the convention in Section **??**.

Since $v'Pv = \iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota - (\iota'\hat{\Lambda}_t\iota)^2/\sigma_m^2$, we need the fourth moments $\mathbb{E}[\iota'(M+\tilde{S}+\tilde{R})\Lambda_t^{-1}(M + \tilde{S} + \tilde{R})\iota]$ and $\mathbb{E}[(\iota'(M+\tilde{S}+\tilde{R})\iota)^2]$. Write $Q = M + \tilde{R}$ (independent of $\tilde{S}$). Expanding each fourth moment into Q-Q, Q-$\tilde{S}$, and $\tilde{S}$-$\tilde{S}$ terms:

$\tilde{S}$-$\tilde{S}$ terms. The Wishart moment identities (88)–(89) with t replaced by $t-1$ give:

$$\mathbb{E}[\iota'\tilde{S}\Lambda_t^{-1}\tilde{S}\iota] = (t-1)(t+N)\sigma_m^2, \tag{111}$$

$$\mathbb{E}[(\iota'\tilde{S}\iota)^2] = (t-1)(t+1)\sigma_m^4. \tag{112}$$

Q-$\tilde{S}$ cross-terms. Since $Q \perp \tilde{S}$ and each appears linearly, the expectation factors:

$$\mathbb{E}[\iota' Q\Lambda_t^{-1}\tilde{S}\iota] = \iota'\mathbb{E}[Q]\Lambda_t^{-1}\mathbb{E}[\tilde{S}]\iota = (t-1)\iota'\mathbb{E}[Q]\iota, \tag{113}$$

$$\mathbb{E}[(\iota' Q\iota)(\iota'\tilde{S}\iota)] = \mathbb{E}[\iota' Q\iota]\cdot(t-1)\sigma_m^2, \tag{114}$$

using $\mathbb{E}[\tilde{S}] = (t-1)\Lambda_t$. In the pricing error formula, the difference of these terms (weighted by 1 and $1/\sigma_m^2$ respectively) is:

$$2(t-1)\iota'\mathbb{E}[Q]\iota - \frac{2(t-1)\sigma_m^2\,\mathbb{E}[\iota' Q\iota]}{\sigma_m^2} = 0.$$

The Q-$\tilde{S}$ cross-terms cancel exactly.

Q-Q terms. Since M is deterministic and $\tilde{R} = \frac{\kappa_0 t}{\kappa_0+t}\eta\eta'$ with $\eta \sim N(0, \Lambda_t/t)$, we have $\mathbb{E}[\tilde{R}] = \frac{\kappa_0}{\kappa_0+t}\Lambda_t$. For the $\tilde{R}$ self-interaction, since $\tilde{R}\Lambda_t^{-1}\tilde{R} = \frac{\kappa_0^2 t^2}{(\kappa_0+t)^2}(\eta'\Lambda_t^{-1}\eta)\,\eta\eta'$, we need $\mathbb{E}[(\eta'\Lambda_t^{-1}\eta)(\iota'\eta)^2]$. Writing $\xi = \sqrt{t}\,\Lambda_t^{-1/2}\eta \sim N(0, I_N)$ and $a = \Lambda_t^{1/2}\iota$, standard normal fourth-moment calculations give $\mathbb{E}[(\xi'\xi)(a'\xi)^2] = (N+2)\|a\|^2 = (N+2)\sigma_m^2$, so:

$$\mathbb{E}[\iota'\tilde{R}\Lambda_t^{-1}\tilde{R}\iota] = \frac{\kappa_0^2(N+2)\sigma_m^2}{(\kappa_0+t)^2}. \tag{115}$$

Similarly, $\mathbb{E}[(\iota'\tilde{R}\iota)^2] = \frac{\kappa_0^2 t^2}{(\kappa_0+t)^2}\mathbb{E}[(\iota'\eta)^4] = \frac{3\kappa_0^2\sigma_m^4}{(\kappa_0+t)^2}$ using $\mathbb{E}[Z^4] = 3\sigma^4$ for $Z \sim N(0,\sigma^2)$.

Combining the M-M, M-$\tilde{R}$, and $\tilde{R}$-$\tilde{R}$ contributions, the Q-Q difference is:

$$A(V_0) + \frac{\kappa_0^2(N-1)\sigma_m^2}{(\kappa_0+t)^2} \tag{116}$$

where $A(V_0) = \iota' M \Lambda_t^{-1} M \iota - (\iota' M \iota)^2 / \sigma_m^2 \geq 0$ is the prior misspecification term from Proposition 3 (with $A(V) = 0$ when $V_0 = V$).

Exact formula. Adding the $\tilde{S}$-$\tilde{S}$ and Q-Q contributions (cross-terms vanish):

$$\mathbb{E}[C] = \frac{\gamma^2 c_t^2}{N^2} \left[A(V_0) + (t-1)(N-1)\sigma_m^2 + \frac{\kappa_0^2 (N-1)\sigma_m^2}{(\kappa_0 + t)^2} \right]. \tag{117}$$

When $V_0 = V$ (so $A(V) = 0$) and the factor structure assumption $\sigma_m^2 \sim O(N^2)$:

$$\mathbb{E}[C] = \frac{\gamma^2 (\kappa_0 + t + 1)^2 (N-1)\sigma_m^2}{N^2 (\kappa_0 + t)^2 (\nu_t - 2)^2} \left[t - 1 + \frac{\kappa_0^2}{(\kappa_0 + t)^2} \right] \sim O(\gamma^2 N / t) \tag{118}$$

since the bracketed term is $\sim O(t)$ and $c_t^2 \sigma_m^2 / N^2 \sim O(1/t^2)$ for bounded κ_0.

Bounds. Since $\mathbb{E}[B] = 0$ and $C \geq 0$:

$$\mathbb{E}[\alpha' \Lambda_t^{-1} \alpha] = \mathbb{E}[A] + \mathbb{E}[C] \geq \mathbb{E}[A] = \frac{t(N-1)}{(t + \kappa_0)^2} \geq \frac{N-1}{4t} \tag{119}$$

for bounded prior strength $\kappa_0 \leq K$ and $t \geq K$. Combined with $\mathbb{E}[C] \sim O(\gamma^2 N / t)$:

$$\mathbb{E}[\alpha' \Lambda_t^{-1} \alpha] \sim O(N/t) \tag{120}$$

Robustness to misspecification. The following regularity conditions formalize the bounded-parameter assumptions used throughout the robustness analysis.

(R.1) The factor loading matrix B_t has uniformly bounded entries: $\sup_{i,j} |[B_t]_{ij}| \leq C_B$ for a constant $C_B > 0$.

(R.2) The true covariance $\Lambda_t = B_t V B_t'$ has eigenvalues bounded away from zero: $0 < c_\lambda \leq \lambda_{\min}(\Lambda_t)$ for a constant c_λ.

(R.3) The prior parameters satisfy $\|\pi - \pi_0\| \leq C_\pi$ and $\max_{i,j} |[V_0]_{ij}| \leq C_V$ for constants C_π, C_V.

Define $d_t = B_t(\pi - \pi_0)$, $\tilde{d}_t = \Lambda_t^{-1/2} d_t$, $b_t = \iota' d_t$, and $D_t = \tilde{d}_t' P \tilde{d}_t$. Under (R.1)–(R.3), each component of d_t is $O(1)$, giving $b_t = O(N)$, and $D_t \le \|\tilde{d}_t\|^2 = d_t' \Lambda_t^{-1} d_t \le \|d_t\|^2 / c_\lambda = O(N)$.

When $V_0 \neq V$ but $\pi_0 = \pi$, sign symmetry still holds: the posterior scale matrix depends on z_τ through the centered scatter and $\bar{z}\bar{z}'$, both invariant under $z_\tau \mapsto -z_\tau$ regardless of V_0. Hence $\mathbb{E}[B] = 0$. The exact formula for $\mathbb{E}[C]$ acquires the nonnegative prior misspecification term $A(V_0) \ge 0$ from Proposition 3. In $\mathbb{E}[C]$, the $A(V_0)$ contribution is multiplied by $\gamma_N^2 c_t^2 = O(1/(N^2 t^2))$, while the $\tilde{S}$-$\tilde{S}$ contribution $(t-1)(N-1)\sigma_m^2 \sim O(tN^3)$ produces the $\sim O(N/t)$ rate. To verify negligibility, note that $A(V_0) = \iota' M \Lambda_t^{-1} M \iota - (\iota' M \iota)^2 / \sigma_m^2 \le \iota' M \Lambda_t^{-1} M \iota$. Under (R.1)–(R.3), each component of $M\iota = B_t V_0 B_t' \iota$ is $O(N)$ and $\lambda_{\max}(\Lambda_t^{-1}) \le 1/c_\lambda$, giving $A(V_0) \le \|M\iota\|^2 / c_\lambda = O(N^3)$. After the $\gamma_N^2 c_t^2 = O(1/(N^2 t^2))$ prefactor, the $A(V_0)$ contribution to $\mathbb{E}[C]$ is $O(N^3/(N^2 t^2)) = O(N/t^2)$, which is $o(N/t)$ for t sufficiently large. Hence the misspecification term is negligible and the $\sim O(N/t)$ rate persists for $t \ge t_0$.

When $\pi_0 \neq \pi$, sign symmetry breaks because $(\bar{X}_t - \pi_0)(\bar{X}_t - \pi_0)' = (\bar{z} + \pi - \pi_0)(\bar{z} + \pi - \pi_0)'$ is no longer even in $\bar{z}$, so $\mathbb{E}[B]$ need not vanish. Decompose $u = u_c + u_b$ with

$$u_c = -\frac{t}{t + \kappa_0} \Lambda_t^{-1/2} \eta, \qquad u_b = \frac{\kappa_0}{t + \kappa_0} \tilde{d}_t,$$

where $\eta = B_t \bar{z} \sim N(0, \Lambda_t / t)$. Write $v = v_M + v_S + v_R$ for the contributions from M, $\tilde{S}$, and $\tilde{R} = \frac{\kappa_0 t}{\kappa_0 + t} (\eta + d_t)(\eta + d_t)'$ respectively.

The $\tilde{S}$ interaction vanishes: $u_c \perp v_S$ by Cochran's theorem with $\mathbb{E}[u_c] = 0$, so $\mathbb{E}[u_c' P v_S] = 0$; and u_b is deterministic, so $\mathbb{E}[u_b' P v_S] = u_b' P \mathbb{E}[v_S]$, which equals zero since $P\,\mathbb{E}[v_S] = c_t(t-1)Pm = 0$. The M interaction is $\mathbb{E}[u' P v_M] = u_b' P v_M = \frac{\kappa_0 c_t}{t + \kappa_0} \tilde{d}_t' P \Lambda_t^{-1/2} M \iota$. By Cauchy-Schwarz, $|\tilde{d}_t' P \Lambda_t^{-1/2} M \iota| \le \|\tilde{d}_t\| \, \|\Lambda_t^{-1/2} M \iota\|$. Under (R.1)–(R.3), $\|\tilde{d}_t\| = O(\sqrt{N})$ and $\|\Lambda_t^{-1/2} M \iota\|^2 = \iota' M \Lambda_t^{-1} M \iota = O(N^3)$ (since each component of $M\iota$ is $O(N)$ and $\lambda_{\max}(\Lambda_t^{-1}) = O(1)$), giving $\mathbb{E}[u' P v_M] = O(N^2/t^2)$.

For the $\tilde{R}$ interaction, substitute $\xi = \sqrt{t}\, \Lambda_t^{-1/2} \eta \sim N(0, I_N)$ and set $a = P\tilde{d}_t$, so that $a'a = D_t$

and $a'm = 0$. Since $\iota'\eta = \sigma_{m,t}\, m'\xi/\sqrt{t}$ and $\tilde{R}\iota = \frac{\kappa_0 t}{\kappa_0+t}(\eta + d_t)(\iota'\eta + b_t)$:

$$u'Pv_R = \frac{c_t \kappa_0 t}{(\kappa_0 + t)^2}\Big[\kappa_0 \tilde{d}_t - \sqrt{t}\,\xi\Big]' P\Big[\tfrac{1}{\sqrt{t}}\xi + \tilde{d}_t\Big]\Big[\tfrac{\sigma_{m,t}}{\sqrt{t}} m'\xi + b_t\Big].$$

The product of the first two brackets expands as

$$[\kappa_0 \tilde{d}_t - \sqrt{t}\,\xi]' P[\tfrac{1}{\sqrt{t}}\xi + \tilde{d}_t] = \tfrac{\kappa_0}{\sqrt{t}}\, a'\xi + \kappa_0 D_t - \xi' P\xi - \sqrt{t}\, a'\xi,$$

where $a = P\tilde{d}_t$ satisfies $a'a = D_t$ and $a'm = 0$. Multiplying by $[\frac{\sigma_{m,t}}{\sqrt{t}} m'\xi + b_t]$ yields eight terms:

(i)	$\frac{\kappa_0 \sigma_{m,t}}{t}(a'\xi)(m'\xi)$	$\mathbb{E}[\cdot] = \frac{\kappa_0 \sigma_{m,t}}{t}\, a'm = 0$
(ii)	$\frac{\kappa_0 b_t}{\sqrt{t}}\, a'\xi$	$\mathbb{E}[\cdot] = 0$ (linear in ξ)
(iii)	$\frac{\kappa_0 D_t \sigma_{m,t}}{\sqrt{t}}\, m'\xi$	$\mathbb{E}[\cdot] = 0$ (linear in ξ)
(iv)	$\kappa_0 D_t\, b_t$	$\neq 0$ (deterministic)
(v)	$-\frac{\sigma_{m,t}}{\sqrt{t}}(\xi' P\xi)(m'\xi)$	$\mathbb{E}[\cdot] = 0$ (cubic in ξ)
(vi)	$-\, b_t\, \xi' P\xi$	$\mathbb{E}[\cdot] = -b_t \operatorname{tr}(P) = -b_t(N-1)$
(vii)	$-\, \sigma_{m,t}(a'\xi)(m'\xi)$	$\mathbb{E}[\cdot] = -\sigma_{m,t}\, a'm = 0$
(viii)	$-\sqrt{t}\, b_t\, a'\xi$	$\mathbb{E}[\cdot] = 0$ (linear in ξ)

Terms (i) and (vii) vanish because $a'm = 0$; terms (ii), (iii), and (viii) vanish because odd moments of $\xi \sim N(0, I_N)$ are zero; term (v) vanishes because it is cubic in ξ. The two surviving terms (iv) and (vi) give:

$$\mathbb{E}[u'Pv_R] = \frac{c_t \kappa_0 t\, b_t}{(\kappa_0 + t)^2}\Big[\kappa_0 D_t - (N-1)\Big]. \tag{121}$$

Since $c_t = O(1/t)$ and $\kappa_0 t/(\kappa_0 + t)^2 = O(1/t)$ for bounded κ_0, the prefactor is $O(1/t^2)$. Under (R.1)–(R.3), $b_t = O(N)$ and $D_t = O(N)$, so $b_t[\kappa_0 D_t - (N-1)] = O(N^2)$:

$$|\mathbb{E}[u'Pv]| \le |\mathbb{E}[u'Pv_R]| + O(N^2/t^2) = O(N^2/t^2). \tag{122}$$

The bias also contributes to $\mathbb{E}[A]$: since $\mathbb{E}[u_c] = 0$, the cross-term $\mathbb{E}[u_c' P u_b]$ vanishes, giving $\mathbb{E}[A] = \frac{t(N-1)}{(t+\kappa_0)^2} + \frac{\kappa_0^2 D_t}{(t+\kappa_0)^2} \sim O(N/t)$. The $\tilde{S}$-$\tilde{S}$ contribution to $\mathbb{E}[C]$ is unchanged by mean misspecification (since $\tilde{S} \perp \bar{z}$) and equals $\sim O(N/t)$. The Q-$\tilde{S}$ cross-terms still cancel (the identity $\iota' \mathbb{E}[Q] \iota = \mathbb{E}[\iota' Q \iota]$ holds regardless of the distribution of Q). Mean misspecification replaces $\eta\eta'$ with $(\eta + d_t)(\eta + d_t)'$ in $\tilde{R}$, adding d_t-dependent terms to the Q-Q contribution to $\mathbb{E}[v'Pv]$. Expanding $(\eta + d_t)(\eta + d_t)' = \eta\eta' + \eta d_t' + d_t \eta' + d_t d_t'$, the new terms in $\iota' \tilde{R} \Lambda_t^{-1} \tilde{R} \iota$ and $(\iota' \tilde{R} \iota)^2$ involve products of up to two factors of η with fixed vectors d_t and ι. The purely deterministic contribution is $[\kappa_0 t/(\kappa_0 + t)]^2 (b_t)^2 (d_t' \Lambda_t^{-1} d_t) = O(N^3)$ (using $b_t = O(N)$ and $d_t' \Lambda_t^{-1} d_t = O(N)$). The mixed terms (from $\eta d_t' + d_t \eta'$) contribute bilinear expressions of the form $(\iota' d_t)(d_t' \Lambda_t^{-1} \eta)(\iota' \eta)$ and similar. By Cauchy-Schwarz, each is bounded in expectation by $\sqrt{\mathbb{E}[(\cdot)^2]\, \mathbb{E}[(\cdot)^2]}$ where $\mathbb{E}[(\iota'\eta)^2] = \sigma_{m,t}^2/t = O(N^2/t)$ and $\mathbb{E}[(d_t' \Lambda_t^{-1} \eta)^2] = d_t' \Lambda_t^{-1} d_t / t = O(N/t)$, with fixed-vector factors $b_t = O(N)$ and $d_t' \Lambda_t^{-1} d_t = O(N)$. The resulting mixed contributions are $O(N^3/t)$. After the $\gamma_N^2 c_t^2 = O(1/(N^2 t^2))$ prefactor, these add $O(N/t^2)$ to $\mathbb{E}[C]$, which is $o(N/t)$ for t sufficiently large. Hence $\mathbb{E}[C] \sim O(N/t)$ for $t \geq t_0$. By Cauchy-Schwarz, $|\mathbb{E}[B]| \leq 2\sqrt{\mathbb{E}[A]\, \mathbb{E}[C]} = O(N/t)$, hence $\mathbb{E}[\alpha' \Lambda_t^{-1} \alpha] = O(N/t)$.

For the lower bound, $q(\gamma_N) \equiv \mathbb{E}[u'Pu] + 2\gamma_N \mathbb{E}[u'Pv] + \gamma_N^2 \mathbb{E}[v'Pv] \geq 0$ for all γ_N, since $\alpha' \Lambda_t^{-1} \alpha \geq 0$ a.s. The minimum of the quadratic is $(1-\rho^2)\mathbb{E}[u'Pu]$, where $\rho = \mathbb{E}[u'Pv]/\sqrt{\mathbb{E}[u'Pu]\, \mathbb{E}[v'Pv]}$ satisfies $|\rho| \leq 1$ by non-negativity of the discriminant. Writing $v = v_Q + v_S$ with $v_Q = c_t \Lambda_t^{-1/2} Q \iota$ and $v_S = c_t \Lambda_t^{-1/2} \tilde{S} \iota$, we have $\mathbb{E}[v'Pv] = \mathbb{E}[v_Q' P v_Q] + 2\mathbb{E}[v_Q' P v_S] + \mathbb{E}[v_S' P v_S]$. The cross-term $\mathbb{E}[v_Q' P v_S]$ vanishes (as shown above, the Q-$\tilde{S}$ cancellation holds regardless of Q's distribution), and $\mathbb{E}[v_Q' P v_Q] \geq 0$ since $P \succeq 0$. Hence $\mathbb{E}[v'Pv] \geq \mathbb{E}[v_S' P v_S] = c_t^2 (t-1)(N-1)\sigma_m^2 \sim O(N^3/t)$, so:

$$|\rho| = \frac{O(N^2/t^2)}{\sqrt{O(N/t) \cdot O(N^3/t)}} = \frac{O(N^2/t^2)}{O(N^2/t)} = O(1/t). \tag{123}$$

Therefore $(1 - \rho^2) \geq 1 - O(1/t^2) > 0$ for t above a fixed threshold t_0 that depends on the constants in (R.1)–(R.3) but not on N. For all $t \geq t_0$:

$$\mathbb{E}[\alpha' \Lambda_t^{-1} \alpha] \geq (1 - \rho^2)\, \mathbb{E}[u'Pu] \geq \tfrac{1}{2}\, \mathbb{E}[A] \sim O(N/t). \tag{124}$$

Combined with the upper bound $\mathbb{E}[\alpha' \Lambda_t^{-1} \alpha] = O(N/t)$ established above, the $\sim O(N/t)$ rate persists under misspecified priors satisfying (R.1)–(R.3), for $t \geq t_0$. □

6.7 Heterogeneous Equilibrium Price Derivation

Derivation of Heterogeneous Equilibrium Price. We derive the equilibrium price when investors have heterogeneous beliefs about the mean π but know the true covariance matrix V. This shows how the equilibrium price aggregates beliefs across investor types weighted by both their population mass and their effective risk tolerance.

Step 1: Market Clearing Condition Market clearing requires that aggregate demand equals the per-capita supply:

$$\frac{\iota}{N} = \sum_{i=1}^{I} \theta_i x_{i,t} \tag{125}$$

where θ_i is the population share of investor type i and $x_{i,t}$ is their individual demand.

Step 2: Substitute Individual Demands From Equation 40, each investor's demand is:

$$x_{i,t} = \frac{1}{\gamma \varphi_{i,t}} \Lambda_t^{-1} \left(\hat{a}_{i,t} - (1 + r_f) p_t \right) \tag{126}$$

Substituting into the market clearing condition:

$$\frac{\iota}{N} = \sum_{i=1}^{I} \theta_i \frac{1}{\gamma \varphi_{i,t}} \Lambda_t^{-1} (\hat{a}_{i,t} - (1 + r_f) p_t) \tag{127}$$

$$= \frac{1}{\gamma} \Lambda_t^{-1} \sum_{i=1}^{I} \vartheta_{i,t} (\hat{a}_{i,t} - (1 + r_f) p_t) \tag{128}$$

where $\vartheta_{i,t} = \theta_i / \varphi_{i,t}$ is the effective belief weight.

Step 3: Solve for Equilibrium Price Multiplying both sides by $\gamma\Lambda_t$ and rearranging:

$$\frac{\gamma}{N}\Lambda_t\iota = \sum_{i=1}^{I} \vartheta_{i,t}(\hat{a}_{i,t} - (1+r_f)p_t) \tag{129}$$

$$\frac{\gamma}{N}\Lambda_t\iota = \sum_{i=1}^{I} \vartheta_{i,t}\hat{a}_{i,t} - \left(\sum_{i=1}^{I} \vartheta_{i,t}\right)(1+r_f)p_t \tag{130}$$

Define the total effective weight $\Omega_t = \sum_{i=1}^{I} \vartheta_{i,t}$ and the aggregate expected dividend:

$$\hat{a}_{E,t} = \frac{1}{\Omega_t}\sum_{i=1}^{I} \vartheta_{i,t}\hat{a}_{i,t} \tag{131}$$

Then:

$$\frac{\gamma}{N}\Lambda_t\iota = \Omega_t\hat{a}_{E,t} - \Omega_t(1+r_f)p_t \tag{132}$$

Solving for p_t using $\gamma_N = \gamma/N$ and $\tilde{\gamma}_{N,t} = \gamma_N/\Omega_t$:

$$\Omega_t(1+r_f)p_t = \Omega_t\hat{a}_{E,t} - \gamma_N\Lambda_t\iota \tag{133}$$

$$(1+r_f)p_t = \hat{a}_{E,t} - \frac{\gamma_N}{\Omega_t}\Lambda_t\iota \tag{134}$$

$$p_t = \frac{1}{1+r_f}\left(\hat{a}_{E,t} - \tilde{\gamma}_{N,t}\Lambda_t\iota\right) \tag{135}$$

Step 4: Economic Interpretation This equilibrium price (Equation 41) maintains the same structure as the homogeneous case (Equation 20), with two modifications:

1. The expected dividend $\hat{a}_{E,t}$ is now an effective belief-weighted average across investor types, where the effective weight $\vartheta_{i,t} = \theta_i/\varphi_{i,t}$ accounts for both population share and parameter uncertainty.

2. The risk premium coefficient $\tilde{\gamma}_{N,t} = \gamma_N/\Omega_t$ reflects the aggregate effective risk tolerance of the economy. As investors become more uncertain (higher $\varphi_{i,t}$), their effective weights decline, reducing Ω_t and increasing the equilibrium risk premium.

□

6.8 Persistence of Complexity-Driven Pricing Errors with Heterogeneous Agents

Proposition A.6 (Persistence of Complexity-Driven Pricing Errors)**.** *In the heterogeneous economy under Assumption H.1, where agents know the covariance matrix V but estimate the mean π using priors $(\pi_{0,i}, \kappa_{0,i})$, the aggregate mean estimation error $\tilde{\varepsilon}_t = a_t - \hat{a}_{E,t}$ leads to pricing errors relative to the market portfolio. If there is a positive mass $\theta_u > 0$ of uninformed investors (Type u) with finite prior precision $\kappa_{0,u}$, the expected squared pricing errors satisfy a lower bound analogous to the homogeneous case (Proposition 4):*

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] \equiv \mathbb{E}[\tilde{\varepsilon}_t' \Lambda_t^{-1} \tilde{\varepsilon}_t - \frac{(\iota' \tilde{\varepsilon}_t)^2}{\iota' \Lambda_t \iota}] \geq \frac{\theta_u^2 t(N-1)}{(\kappa_{0,u} + t + 1)^2} \sim O\left(\frac{N}{t}\right), \tag{136}$$

where the constant θ_u^2 depends on the uninformed agents' population share θ_u, and the $(+1)$ in the denominator arises from the parameter uncertainty factor in the Normal-Normal conjugate posterior predictive distribution.

Proof of Proposition 6: Persistence of Complexity-Driven Pricing Errors. We prove that pricing errors persist at rate $O(N/t)$ even when some agents are fully informed, as long as a positive mass of uninformed agents exists. The proof parallels Proposition 4, leveraging the projection framework and scalar proportionality structure.

Setup Under Assumption H.1, there are I investor types with population shares $\theta_i > 0$ summing to 1. Each type knows V but estimates π using prior $(\pi_{0,i}, \kappa_{0,i})$. Type i's posterior mean is $\hat{\pi}_{i,t} = \eta_{i,t}\pi_{0,i} + (1 - \eta_{i,t})\bar{X}_t$ where $\eta_{i,t} = \kappa_{0,i}/(\kappa_{0,i} + t)$. The effective belief weight is $\vartheta_{i,t} = \theta_i/\varphi_{i,t}$ where $\varphi_{i,t} = (\kappa_{0,i} + t + 1)/(\kappa_{0,i} + t)$, and $\Omega_t = \sum_{i=1}^{I} \vartheta_{i,t} \leq 1$ (since $\varphi_{i,t} \geq 1$ for all i, with equality when $\kappa_{0,i} \to \infty$).

The aggregate estimation error is:

$$\tilde{\varepsilon}_t = a_t - \hat{a}_{E,t} = B_t \left(\frac{1}{\Omega_t} \sum_{i=1}^{I} \vartheta_{i,t} (\pi - \hat{\pi}_{i,t}) \right). \tag{137}$$

Step 1: Correctly Centered Priors When $\pi_{0,i} = \pi$ for all i, each type's estimation error is pure sampling error:

$$\pi - \hat{\pi}_{i,t} = \frac{t}{\kappa_{0,i} + t}(\pi - \bar{X}_t). \tag{138}$$

The aggregate error inherits the same structure with a scalar coefficient:

$$\tilde{\varepsilon}_t = c_t \cdot B_t(\pi - \bar{X}_t), \quad \text{where } c_t = \frac{1}{\Omega_t} \sum_{i=1}^{I} \frac{t\vartheta_{i,t}}{\kappa_{0,i} + t}. \tag{139}$$

Since $\pi - \bar{X}_t \sim N(0, V/t)$ and $\Lambda_t = B_t V B_t'$, the aggregate error has covariance:

$$\text{Cov}(\tilde{\varepsilon}_t) = \frac{c_t^2}{t}\Lambda_t. \tag{140}$$

By Lemma A.1 Part (iv) with proportionality constant c_t^2/t:

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t]_{\pi_{0,i}=\pi} = \frac{c_t^2(N-1)}{t}. \tag{141}$$

Step 2: Lower Bound from Uninformed Type If there exists an uninformed type u with $\theta_u > 0$ and finite $\kappa_{0,u}$, then c_t is bounded below by type u's contribution:

$$c_t \geq \frac{t\vartheta_{u,t}}{\Omega_t(\kappa_{0,u} + t)} = \frac{t\theta_u}{\Omega_t(\kappa_{0,u} + t + 1)} \geq \frac{t\theta_u}{\kappa_{0,u} + t + 1}, \tag{142}$$

where the last inequality uses $\Omega_t \leq 1$. Substituting into (141):

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t]_{\pi_{0,i}=\pi} \geq \frac{\theta_u^2(N-1)t}{(\kappa_{0,u} + t + 1)^2} \sim \frac{\theta_u^2(N-1)}{t} \quad \text{as } t \to \infty. \tag{143}$$

Step 3: Extension to Misspecified Priors When $\pi_{0,i} \neq \pi$, each type's error decomposes as in Proposition 4:

$$\pi - \hat{\pi}_{i,t} = \underbrace{\frac{t}{\kappa_{0,i} + t}(\pi - \bar{X}_t)}_{\text{sampling (mean-zero)}} + \underbrace{\frac{\kappa_{0,i}}{\kappa_{0,i} + t}(\pi - \pi_{0,i})}_{\text{bias}}. \tag{144}$$

Since $\mathbb{E}[\varepsilon_1] = 0$ and the bias ε_2 is deterministic, the cross-term vanishes: $\mathbb{E}[\varepsilon_1' P \varepsilon_2] = \mathbb{E}[\varepsilon_1]' P \varepsilon_2 = 0$. Since $P \succeq 0$, the bias term $\varepsilon_2' P \varepsilon_2 \geq 0$, so:

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] \geq \mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t]_{\pi_{0,i}=\pi} \geq \frac{\theta_u^2 (N-1)t}{(\kappa_{0,u} + t + 1)^2}. \tag{145}$$

The $O(N/t)$ lower bound holds universally: as long as a positive mass of uninformed investors remains, their estimation uncertainty persists in equilibrium prices, even with informed agents present.

□

6.9 Properties of Partial Information Strategies

Proposition A.7 (Properties of Partial Information Strategies)**.** *Consider a continuum of sophisticated investors who know the true covariance matrix Λ_t and observe the aggregate estimation error $\tilde{\varepsilon}_t$ on a random fraction ϕ of assets, selected by $D_{i,t}$. Assume $\tilde{\varepsilon}_t \mid \mathcal{F}_t \sim N(0, \Lambda_{a,t})$ for a positive definite error covariance $\Lambda_{a,t}$. Investors form portfolios $w_{i,t} = \frac{1}{\gamma N} \Lambda_t^{-1} D_{i,t} \tilde{\varepsilon}_t$. Define $T \equiv Tr(\Lambda_t^{-1} \Lambda_{a,t})$ and $\Delta \equiv \sum_{k=1}^{N} [\Lambda_t^{-1}]_{kk} [\Lambda_{a,t}]_{kk}$. Then, the expected properties of the dollar returns $R_{w_{i,t}}$, averaged over the distribution of $\tilde{\varepsilon}_t$ and the independent random selectors $D_{i,t}$, $D_{j,t}$, are:*

1. *Expected Return:* $\mathbb{E}[R_{w_{i,t}}] = \frac{\phi T}{\gamma N}$.
2. *Expected Variance:* $\mathbb{E}[Var_t(R_{w_{i,t}})] = \frac{\phi^2 T + \phi(1-\phi)\Delta}{(\gamma N)^2}$.
3. *Squared Sharpe Ratio:* $SR_{i,t}^2 = \frac{\phi T^2}{\phi T + (1-\phi)\Delta}$.
4. *Ex-ante Correlation:* $Corr^{ex\text{-}ante}(R_{w_{i,t}}, R_{w_{j,t}}) = \frac{\phi}{\phi + (1-\phi)\Delta/T}$ *between two distinct investors i and j, where the ex-ante correlation is* $\mathbb{E}[Cov]/\sqrt{\mathbb{E}[Var_i] \cdot \mathbb{E}[Var_j]}$.

Under the model's Bayesian learning structure (Proposition 6), $\Lambda_{a,t} \propto \Lambda_t$, so $\Delta/T = \chi_t \geq 1$ and $Corr^{ex\text{-}ante} \leq \phi$. The upper bound ϕ is achieved when $\chi_t = 1$ (diagonal Λ_t); there is no positive universal lower bound.

Proof of Proposition 7: Properties of Partial Information Strategies. We prove the four properties of partial information strategies for sophisticated investors who observe the aggregate estimation error $\tilde{\varepsilon}_t$ on a random fraction ϕ of assets.

Setup and Assumptions Consider a continuum of sophisticated investors indexed by i. Under Assumption S.1:

- Each investor knows the true covariance matrix Λ_t.
- Each investor i receives a signal $D_{i,t}\tilde{\varepsilon}_t$, where $D_{i,t} = \text{diag}(d_{i1,t}, \ldots, d_{iN,t})$ is a diagonal selector matrix with $d_{ij,t} \sim \text{Bernoulli}(\phi)$ independently.
- The aggregate estimation error follows $\tilde{\varepsilon}_t \mid \mathcal{F}_t \sim N(0, \Lambda_{a,t})$ for a positive definite error covariance $\Lambda_{a,t}$.
- Investors form portfolios $w_{i,t} = \frac{1}{\gamma N}\Lambda_t^{-1} D_{i,t}\tilde{\varepsilon}_t$.

Note that $D_{i,t}^2 = D_{i,t}$ (idempotence) and $\mathbb{E}[\text{Tr}(D_{i,t})] = N\phi$.

Identity for Quadratic Forms with Random Selectors For deterministic matrices A, B and diagonal selector $D = \text{diag}(d_1, \ldots, d_N)$ with i.i.d. $d_j \sim \text{Bernoulli}(\phi)$:

$$\mathbb{E}_D[\text{Tr}(DADB)] = \phi^2 \text{Tr}(AB) + \phi(1-\phi)\sum_{j=1}^{N} A_{jj}B_{jj}. \tag{146}$$

This follows from expanding the trace as $\sum_{j,k} d_j d_k A_{jk} B_{kj}$ and using $\mathbb{E}[d_j d_k] = \phi^2$ for $j \neq k$ and $\mathbb{E}[d_j^2] = \phi$ for $j = k$.

Define the trace and diagonal terms:

$$T \equiv \text{Tr}(\Lambda_t^{-1}\Lambda_{a,t}), \qquad \Delta \equiv \sum_{k=1}^{N} [\Lambda_t^{-1}]_{kk}[\Lambda_{a,t}]_{kk}. \tag{147}$$

Applying the identity with $A = \Lambda_t^{-1}$ and $B = \Lambda_{a,t}$:

$$\mathbb{E}_D[\mathrm{Tr}(D\Lambda_t^{-1}D\Lambda_{a,t})] = \phi^2 T + \phi(1-\phi)\Delta. \tag{148}$$

When $\Lambda_{a,t} \propto \Lambda_t$ (as holds under Proposition 6's Bayesian structure), the ratio Δ/T reduces to the diagonal overlap coefficient $\chi_t \equiv \frac{1}{N}\sum_{j=1}^{N}(\Lambda_t^{-1})_{jj}(\Lambda_t)_{jj}$, which satisfies $\chi_t \geq 1$ since $(\Lambda_t^{-1})_{jj}(\Lambda_t)_{jj} \geq 1$ for each j by Cauchy-Schwarz: $1 = (e_j'\Lambda_t^{1/2}\Lambda_t^{-1/2}e_j)^2 \leq (e_j'\Lambda_t e_j)(e_j'\Lambda_t^{-1}e_j) = (\Lambda_t)_{jj}(\Lambda_t^{-1})_{jj}$.

Part 1: Expected Return The realized dollar payoff is $R_{w_{i,t}} = w_{i,t}'r_{t+1}$. Since $r_{t+1} = \tilde{\varepsilon}_t + (\hat{a}_{E,t} - (1+r_f)p_t) + (d_{t+1} - a_t)$, the component related to the signal is:

$$w_{i,t}'\tilde{\varepsilon}_t = \frac{1}{\gamma N}(\Lambda_t^{-1}D_{i,t}\tilde{\varepsilon}_t)'\tilde{\varepsilon}_t = \frac{1}{\gamma N}\tilde{\varepsilon}_t'D_{i,t}\Lambda_t^{-1}\tilde{\varepsilon}_t. \tag{149}$$

The expectation over the distribution of $\tilde{\varepsilon}_t$ (conditional on $\mathcal{F}_t$) is computed using the standard result for quadratic forms: for $x \sim N(\mu, \Sigma)$, $\mathbb{E}[x'Ax] = \mathrm{Tr}(A\Sigma) + \mu'A\mu$. Here $\mu = 0$, $x = \tilde{\varepsilon}_t$, $A = D_{i,t}\Lambda_t^{-1}$, and $\Sigma = \Lambda_{a,t}$:

$$\mathbb{E}[w_{i,t}'\tilde{\varepsilon}_t \mid \mathcal{F}_t, D_{i,t}] = \frac{1}{\gamma N}\mathrm{Tr}(D_{i,t}\Lambda_t^{-1}\Lambda_{a,t}). \tag{150}$$

Since $D_{i,t}$ appears on only one side, $\mathrm{Tr}(D_{i,t}\Lambda_t^{-1}\Lambda_{a,t}) = \sum_j d_{ij,t}[\Lambda_t^{-1}\Lambda_{a,t}]_{jj}$. Taking the expectation over $D_{i,t}$:

$$\mathbb{E}[R_{w_{i,t}}] = \frac{\phi}{\gamma N}\mathrm{Tr}(\Lambda_t^{-1}\Lambda_{a,t}) = \frac{\phi T}{\gamma N}. \quad \blacksquare \tag{151}$$

Part 2: Expected Variance The conditional variance of the portfolio return, given $\mathcal{F}_t$ and the signal realization $\tilde{\varepsilon}_t$, is:

$$\mathrm{Var}_t(R_{w_{i,t}}) \mid \mathcal{F}_t, \tilde{\varepsilon}_t = w_{i,t}'\Lambda_t w_{i,t}. \tag{152}$$

Substituting $w_{i,t} = \frac{1}{\gamma N}\Lambda_t^{-1} D_{i,t}\tilde{\varepsilon}_t$:

$$w_{i,t}'\Lambda_t w_{i,t} = \frac{1}{(\gamma N)^2}\tilde{\varepsilon}_t' D_{i,t}\Lambda_t^{-1}\Lambda_t\Lambda_t^{-1} D_{i,t}\tilde{\varepsilon}_t \tag{153}$$

$$= \frac{1}{(\gamma N)^2}\tilde{\varepsilon}_t' D_{i,t}\Lambda_t^{-1} D_{i,t}\tilde{\varepsilon}_t. \tag{154}$$

Taking the expectation over $\tilde{\varepsilon}_t$ using the standard quadratic form result:

$$\mathbb{E}[\mathrm{Var}_t(R_{w_{i,t}}) \mid \mathcal{F}_t, D_{i,t}] = \frac{1}{(\gamma N)^2}\mathrm{Tr}(D_{i,t}\Lambda_t^{-1} D_{i,t}\Lambda_{a,t}). \tag{155}$$

Unlike Part 1, here $D_{i,t}$ appears on both sides of Λ_t^{-1}. Since $D_{i,t}$ does not commute with Λ_t^{-1} when Λ_t has off-diagonal elements, we apply the trace identity (148):

$$\mathbb{E}_D[\mathrm{Tr}(D_{i,t}\Lambda_t^{-1} D_{i,t}\Lambda_{a,t})] = \phi^2 T + \phi(1-\phi)\Delta. \tag{156}$$

Therefore:

$$\mathbb{E}[\mathrm{Var}_t(R_{w_{i,t}})] = \frac{\phi^2 T + \phi(1-\phi)\Delta}{(\gamma N)^2}. \quad \blacksquare \tag{157}$$

Part 3: Squared Sharpe Ratio From Parts 1 and 2:

$$SR_{i,t}^2 = \frac{(\mathbb{E}[R_{w_{i,t}}])^2}{\mathbb{E}[\mathrm{Var}_t(R_{w_{i,t}})]} = \frac{\phi^2 T^2/(\gamma N)^2}{[\phi^2 T + \phi(1-\phi)\Delta]/(\gamma N)^2} = \frac{\phi T^2}{\phi T + (1-\phi)\Delta}. \quad \blacksquare \tag{158}$$

Part 4: Ex-ante Correlation The conditional covariance between returns of two different investors i and j is:

$$\mathrm{Cov}_t(R_{w_{i,t}}, R_{w_{j,t}}) \mid \mathcal{F}_t, \tilde{\varepsilon}_t = w_{i,t}'\Lambda_t w_{j,t}. \tag{159}$$

Substituting the portfolio weights:

$$w'_{i,t}\Lambda_t w_{j,t} = \frac{1}{(\gamma N)^2}\tilde{\varepsilon}'_t D_{i,t}\Lambda_t^{-1}\Lambda_t\Lambda_t^{-1}D_{j,t}\tilde{\varepsilon}_t \tag{160}$$

$$= \frac{1}{(\gamma N)^2}\tilde{\varepsilon}'_t D_{i,t}\Lambda_t^{-1}D_{j,t}\tilde{\varepsilon}_t. \tag{161}$$

Taking the expectation over $\tilde{\varepsilon}_t$:

$$\mathbb{E}[\mathrm{Cov}_t(R_{w_{i,t}}, R_{w_{j,t}}) \mid \mathcal{F}_t, D_{i,t}, D_{j,t}] = \frac{1}{(\gamma N)^2}\mathrm{Tr}(D_{i,t}\Lambda_t^{-1}D_{j,t}\Lambda_{a,t}). \tag{162}$$

For $i \neq j$, the selector matrices $D_{i,t}$ and $D_{j,t}$ are drawn independently. Since the elements $d_{ik,t}$ and $d_{jl,t}$ are independent Bernoulli(ϕ) random variables for all k, l:

$$\mathbb{E}[d_{ik,t}d_{jl,t}] = \mathbb{E}[d_{ik,t}]\mathbb{E}[d_{jl,t}] = \phi^2 \quad \text{for all } k, l. \tag{163}$$

Therefore $\mathbb{E}_{D_i,D_j}[D_{i,t}AD_{j,t}] = \phi^2 A$ for any matrix A when $i \neq j$. This is the main distinction from Part 2: because the two selectors are independent, we do not get the diagonal correction term. Thus:

$$\mathbb{E}[\mathrm{Tr}(D_{i,t}\Lambda_t^{-1}D_{j,t}\Lambda_{a,t})] = \phi^2\mathrm{Tr}(\Lambda_t^{-1}\Lambda_{a,t}) = \phi^2 T. \tag{164}$$

Therefore:

$$\mathbb{E}[\mathrm{Cov}_t(R_{w_{i,t}}, R_{w_{j,t}})] = \frac{\phi^2 T}{(\gamma N)^2}. \tag{165}$$

Define the ex-ante correlation as $\mathrm{Corr}^{ex\text{-}ante} \equiv \mathbb{E}[\mathrm{Cov}]/\sqrt{\mathbb{E}[\mathrm{Var}_i]\cdot\mathbb{E}[\mathrm{Var}_j]}$, where expectations are over signal distributions. By symmetry ($\mathbb{E}[\mathrm{Var}_i] = \mathbb{E}[\mathrm{Var}_j]$):

$$\mathrm{Corr}^{ex\text{-}ante}(R_{w_{i,t}}, R_{w_{j,t}}) = \frac{\phi^2 T}{\phi^2 T + \phi(1-\phi)\Delta} = \frac{\phi}{\phi + (1-\phi)\Delta/T}. \tag{166}$$

Under the model's proportionality ($\Lambda_{a,t} \propto \Lambda_t$, a consequence of Proposition 6), $\Delta/T = \chi_t \geq 1$, giving $\mathrm{Corr}^{ex\text{-}ante} = \phi/[\phi + (1-\phi)\chi_t] \leq \phi$. ■

□

6.10 Robustness: Sophisticated Investors with Positive Mass

The results in Proposition 7 (Properties of Partial Information Strategies) extend to a rational expectations equilibrium where sophisticated investors have positive total mass $\theta_s > 0$ and participate in price formation. The findings are:

1. Pricing errors persist: Mispricing is attenuated by factor $\rho^* = \frac{\Omega_t^{(L)}}{\Omega_t^{(L)}+\theta_s\phi} < 1$ but maintains $O(N/t)$ scaling:

$$\mathbb{E}[\alpha_t'\Lambda_t^{-1}\alpha_t]^{(new)} = (\rho^*)^2 \cdot \mathbb{E}[\alpha_t'\Lambda_t^{-1}\alpha_t]^{(L)} \geq \frac{(\rho^*)^2\theta_2^2\, t(N-1)}{(\kappa_{0,2}+t+1)^2} \sim O\left(\frac{N}{t}\right) \tag{167}$$

2. Ex-ante correlation bounded by information overlap: The ex-ante correlation between sophisticated investors' strategies satisfies:

$$\text{Corr}^{ex\text{-}ante}(R_{w_{i,t}}, R_{w_{j,t}}) = \frac{\phi}{\phi + (1-\phi)\Delta/T} \tag{168}$$

   independent of θ_s, N, t, or ρ^*. Under the model's proportionality, $\Delta/T = \chi_t \geq 1$, so $\text{Corr}^{ex\text{-}ante} \leq \phi$.

3. Sharpe ratios scale with complexity: The squared Sharpe ratio is $SR_{i,t}^2 = \phi T^2/[\phi T + (1-\phi)\Delta]$ where T, Δ are defined from $\Lambda_{a,t}^{(new)} = (\rho^*)^2\Lambda_{a,t}^{(L)}$, maintaining $O(N/t)$ scaling. Under proportionality:

$$SR_{i,t} = \rho^*\sqrt{\frac{N}{t+\kappa_E}} \cdot \frac{\sqrt{\phi}}{\sqrt{\phi + (1-\phi)\chi_t}} \sim O\left(\sqrt{\frac{N}{t+\kappa_E}}\right) \tag{169}$$

   The attenuation factor ρ^* appears once in the Sharpe ratio.

The proof establishes a rational expectations fixed-point equilibrium where sophisticated investors infer the aggregate belief $\hat{a}_{E,t}^{(new)}$ from prices and use it for unobserved assets, eliminating

any common signal component across investors. The attenuation factor ρ^* depends on effective informed capital $\theta_s\phi$ (total sophisticated mass times information coverage), capturing the competitive impact of sophisticated capital on mispricing and returns.

Equilibrium with Positive Mass of Sophisticated Investors. We show that when sophisticated investors have positive mass $\theta_s > 0$ and participate in a rational expectations equilibrium, the main results persist: (i) pricing errors remain $O(N/t)$ (though attenuated), (ii) the ex-ante correlation between their active strategies is bounded above by ϕ, and (iii) Sharpe ratios scale as $O(\sqrt{N/t})$. These results require Assumption RE (partial revelation), which prevents sophisticated investors from inverting prices to recover the true mean a_t on unobserved assets, the standard Grossman–Stiglitz wedge needed for mispricing to persist. The ex-ante correlation and Sharpe ratio admit closed-form expressions in full generality (Equations 204 and 206). Under the model's proportionality (a consequence of Proposition 6), $\text{Corr}^{ex\text{-}ante} \leq \phi$.

Setup: Three Groups of Investors The economy contains three groups with total measure 1. Write $a_t \equiv \pi$ for the true (time-invariant) expected dividend, so that learning agents estimate π while sophisticated investors observe $a_t = \pi$ on subsets of assets.

1. Informed learning agents (mass θ_1): Know $a_t = \pi$ (or equivalently, $\pi_{0,1} = \pi$ with $\kappa_{0,1} \to \infty$). Have $\varphi_{1,t} = 1$ and effective weight $\vartheta_{1,t} = \theta_1$.

2. Uninformed learning agents (mass θ_2): Estimate π from data with prior $(\pi_{0,2}, \kappa_{0,2})$. Have $\varphi_{2,t} = \frac{\kappa_{0,2}+t+1}{\kappa_{0,2}+t}$ and effective weight $\vartheta_{2,t} = \theta_2/\varphi_{2,t}$.

3. Sophisticated investors (SIs) (total mass θ_s): A continuum indexed by $i \in [0, \theta_s]$. Each knows the true covariance Λ_t and observes a_t on a random subset of assets selected by diagonal matrix $D_{i,t} = \text{diag}(d_{i1,t}, \ldots, d_{iN,t})$ where $d_{ij,t} \overset{\text{iid}}{\sim} \text{Bernoulli}(\phi)$ across assets j, investors i, and time t.

Total population constraint: $\theta_1 + \theta_2 + \theta_s = 1$.

Let $\Omega_t^{(L)} = \vartheta_{1,t} + \vartheta_{2,t} = \theta_1 + \frac{\theta_2}{\varphi_{2,t}}$ be the total effective weight of learning agents, and $\hat{a}_{E,t}^{(L)} = \frac{\vartheta_{1,t}\hat{a}_{1,t} + \vartheta_{2,t}\hat{a}_{2,t}}{\Omega_t^{(L)}}$ be their aggregate belief (as in Proposition 6).

Assumption RE (Partial Revelation) Sophisticated investors observe the equilibrium price p_t and know the pricing map

$$(1 + r_f)p_t = \hat{a}_{E,t}^{(new)} - \tilde{\gamma}_{N,t}^{(total)}\Lambda_t \iota. \tag{170}$$

We assume that $\tilde{\gamma}_{N,t}^{(total)} = \gamma_N / \Omega_t^{(total)}$ is common knowledge: all agents know the aggregate risk-aversion coefficient γ_N, the covariance matrix Λ_t, population masses $(\theta_1, \theta_2, \theta_s)$, and the prior precision $\kappa_{0,2}$ (hence $\varphi_{2,t}$ and $\Omega_t^{(total)} = \theta_1 + \theta_2/\varphi_{2,t} + \theta_s$). Given this, SIs can infer $\hat{a}_{E,t}^{(new)}$ from prices by solving $\hat{a}_{E,t}^{(new)} = (1 + r_f)p_t + \tilde{\gamma}_{N,t}^{(total)}\Lambda_t \iota$.

However, SIs cannot recover the underlying decomposition $\left(\Omega_t^{(L)}, \hat{a}_{E,t}^{(L)}\right)$ because learning agents' posterior means depend on their private signals: specifically, the uninformed agents' prior mean $\pi_{0,2}$ and their accumulated data, which are not publicly observable. SIs nonetheless form rational expectations: their beliefs about unobserved assets are consistent with the $\hat{a}_{E,t}^{(new)}$ that will obtain in equilibrium.

For the analysis below (Steps 1–6), we assume that SIs use $\hat{a}_{E,t}^{(\text{new})}$ directly for unobserved assets.

We maintain the belief-updating rules of learning agents as in Proposition 6; only SIs form rational expectations about the equilibrium mapping from information to prices.

Remark (Why partial revelation is needed) If $\left(\Omega_t^{(L)}, \hat{a}_{E,t}^{(L)}\right)$ were common knowledge, then combining the inferred $\hat{a}_{E,t}^{(new)}$ with Equation 180 would allow SIs to back out a_t even on unobserved assets, eliminating the estimation-error component of mispricing. Without some wedge preventing full price revelation, private information gets impounded and arbitrage opportunities disappear (the standard Grossman–Stiglitz requirement).

Step 1: Rational Beliefs and the Fixed-Point Problem By Assumption RE, SIs can infer $\hat{a}_{E,t}^{(new)}$ from p_t but cannot invert equilibrium objects to recover a_t for assets they do not directly

observe. Hence, for unobserved assets their rational expectation coincides with the aggregate belief embedded in prices.

In equilibrium, the price p_t and the aggregate belief $\hat{a}_{E,t}^{(new)}$ are mutually consistent:

$$p_t = \frac{1}{1+r_f}\left(\hat{a}_{E,t}^{(new)} - \tilde{\gamma}_{N,t}^{(total)}\Lambda_t \iota\right) \tag{171}$$

where $\Omega_t^{(total)} = \Omega_t^{(L)} + \theta_s$ is the total risk-bearing capacity, and $\tilde{\gamma}_{N,t}^{(total)} = \gamma_N/\Omega_t^{(total)}$ is the aggregate risk-aversion coefficient.

A sophisticated investor i forms their expectation for all N assets. For the fraction ϕ of assets they observe (where $D_{i,t}$ has ones), they know the true mean is a_t. For the fraction $(1-\phi)$ of assets they do not observe (where $D_{i,t}$ has zeros), their rational expectation (given Assumption RE) is the aggregate belief $\hat{a}_{E,t}^{(new)}$.

Therefore, the rational expectation for SI i is:

$$E_i[d_{t+1}] = D_{i,t}a_t + (I - D_{i,t})\hat{a}_{E,t}^{(new)} \tag{172}$$

This is the main difference from a naive analysis: sophisticated investors use the equilibrium aggregate belief $\hat{a}_{E,t}^{(new)}$ (which they can infer from prices) rather than the learning agents' belief $\hat{a}_{E,t}^{(L)}$ that would obtain without their participation. Assumption RE prevents them from using this inference to back out a_t on unobserved assets.

Step 2: Solving for the Equilibrium Aggregate Belief The aggregate belief $\hat{a}_{E,t}^{(new)}$ is the belief-weighted average of all agents' expected dividends. We solve for the fixed point where the beliefs that form the price are the same as the beliefs formed by the price.

The market clearing condition, expressed in terms of beliefs, is:

$$\Omega_t^{(total)}\hat{a}_{E,t}^{(new)} = \underbrace{\Omega_t^{(L)}\hat{a}_{E,t}^{(L)}}_{\text{Learning Agents' Belief Mass}} + \underbrace{\int_0^{\theta_s} E_i[d_{t+1}]\, di}_{\text{SIs' Belief Mass}} \tag{173}$$

We aggregate the SIs' beliefs. By the Law of Large Numbers over the continuum of independent $D_{i,t}$ draws, $\int_0^{\theta_s} D_{i,t}\, di = \theta_s \phi I_N$:

$$\int_0^{\theta_s} E_i[d_{t+1}]\, di = \int_0^{\theta_s} \left(D_{i,t} a_t + (I - D_{i,t}) \hat{a}_{E,t}^{(new)} \right) di \tag{174}$$

$$= \theta_s \phi a_t + \theta_s (1 - \phi) \hat{a}_{E,t}^{(new)} \tag{175}$$

Substitute this back into the market clearing condition:

$$\Omega_t^{(total)} \hat{a}_{E,t}^{(new)} = \Omega_t^{(L)} \hat{a}_{E,t}^{(L)} + \theta_s \phi a_t + \theta_s (1 - \phi) \hat{a}_{E,t}^{(new)} \tag{176}$$

Now, solve for $\hat{a}_{E,t}^{(new)}$:

$$(\Omega_t^{(total)} - \theta_s (1 - \phi)) \hat{a}_{E,t}^{(new)} = \Omega_t^{(L)} \hat{a}_{E,t}^{(L)} + \theta_s \phi a_t \tag{177}$$

$$(\Omega_t^{(L)} + \theta_s - \theta_s + \theta_s \phi) \hat{a}_{E,t}^{(new)} = \Omega_t^{(L)} \hat{a}_{E,t}^{(L)} + \theta_s \phi a_t \tag{178}$$

$$(\Omega_t^{(L)} + \theta_s \phi) \hat{a}_{E,t}^{(new)} = \Omega_t^{(L)} \hat{a}_{E,t}^{(L)} + \theta_s \phi a_t \tag{179}$$

This yields the correct aggregate belief in the rational expectations equilibrium:

$$\hat{a}_{E,t}^{(new)} = \frac{\Omega_t^{(L)} \hat{a}_{E,t}^{(L)} + \theta_s \phi a_t}{\Omega_t^{(L)} + \theta_s \phi} \tag{180}$$

Step 3: Modified Aggregate Estimation Error The new aggregate estimation error, $\tilde{\varepsilon}_t^{(new)}$, is $a_t - \hat{a}_{E,t}^{(new)}$:

$$\tilde{\varepsilon}_t^{(new)} = a_t - \frac{\Omega_t^{(L)} \hat{a}_{E,t}^{(L)} + \theta_s \phi a_t}{\Omega_t^{(L)} + \theta_s \phi} \tag{181}$$

$$= \frac{(\Omega_t^{(L)} + \theta_s \phi) a_t - \Omega_t^{(L)} \hat{a}_{E,t}^{(L)} - \theta_s \phi a_t}{\Omega_t^{(L)} + \theta_s \phi} \tag{182}$$

$$= \left(\frac{\Omega_t^{(L)}}{\Omega_t^{(L)} + \theta_s \phi} \right) (a_t - \hat{a}_{E,t}^{(L)}) \tag{183}$$

$$\equiv \rho^* \cdot \tilde{\varepsilon}_t^{(L)} \tag{184}$$

where

$$\rho^* = \frac{\Omega_t^{(L)}}{\Omega_t^{(L)} + \theta_s \phi} \in (0, 1) \tag{185}$$

is the attenuation factor. Note that $\rho^* < 1$ (mispricing is reduced) and $\rho^* \to 1$ as $\theta_s \to 0$ (recovering Proposition 6 in the limit). The attenuation depends on effective informed capital $\theta_s \phi$, not just total sophisticated capital θ_s.

Result (i): Scaling of Pricing Errors The total mispricing $\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t]^{(new)}$ is proportional to $\mathbb{E}[(\tilde{\varepsilon}_t^{(new)})' \Lambda_t^{-1} \tilde{\varepsilon}_t^{(new)}]$, which scales as $(\rho^*)^2$:

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t]^{(new)} = (\rho^*)^2 \cdot \mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t]^{(L)} \geq \frac{(\rho^*)^2 \theta_2^2 \, t(N-1)}{(\kappa_{0,2} + t + 1)^2} \sim O\left(\frac{N}{t}\right) \tag{186}$$

Mispricing is attenuated by the presence of SIs (since $\rho^* < 1$), but it maintains the $O(N/t)$ scaling. As long as $\Omega_t^{(L)} > 0$ (learning agents have positive effective mass), pricing errors persist. The attenuation factor $\rho^* \to 0$ only as $\theta_s \phi \to \infty$, which is impossible given finite population masses.

Step 4: The SI's Optimal Portfolio Now we derive the optimal portfolio for SI i. Their demand is based on their perceived excess return, $E_i[r_{t+1}]$:

$$E_i[r_{t+1}] = E_i[d_{t+1}] - (1 + r_f)p_t \tag{187}$$

$$= E_i[d_{t+1}] - (\hat{a}_{E,t}^{(new)} - \tilde{\gamma}_{N,t}^{(total)}\Lambda_t\iota) \tag{188}$$

$$= [D_{i,t}a_t + (I - D_{i,t})\hat{a}_{E,t}^{(new)}] - \hat{a}_{E,t}^{(new)} + \tilde{\gamma}_{N,t}^{(total)}\Lambda_t\iota \tag{189}$$

$$= D_{i,t}a_t - D_{i,t}\hat{a}_{E,t}^{(new)} + \tilde{\gamma}_{N,t}^{(total)}\Lambda_t\iota \tag{190}$$

$$= D_{i,t}(a_t - \hat{a}_{E,t}^{(new)}) + \tilde{\gamma}_{N,t}^{(total)}\Lambda_t\iota \tag{191}$$

$$= D_{i,t}\tilde{\varepsilon}_t^{(new)} + \tilde{\gamma}_{N,t}^{(total)}\Lambda_t\iota \tag{192}$$

The optimal CARA demand $x_{s,i,t} = \frac{1}{\gamma}\Lambda_t^{-1}E_i[r_{t+1}]$ is:

$$x_{s,i,t} = \underbrace{\frac{1}{\gamma}\Lambda_t^{-1}D_{i,t}\tilde{\varepsilon}_t^{(new)}}_{\text{Speculative Component}} + \underbrace{\frac{\tilde{\gamma}_{N,t}^{(total)}}{\gamma}\iota}_{\text{Market Component}} \tag{193}$$

The speculative signal $E_i[r_{t+1}] - \tilde{\gamma}_{N,t}^{(total)}\Lambda_t\iota = D_{i,t}\tilde{\varepsilon}_t^{(new)}$ contains no common component on unobserved assets: sophisticated investors rationally use the equilibrium aggregate belief $\hat{a}_{E,t}^{(new)}$ for assets they do not observe, eliminating what would otherwise be a common signal component $[\hat{a}_{E,t}^{(L)} - \hat{a}_{E,t}^{(new)}]$. Any cross-SI comovement arises solely from overlap in the assets they directly observe (i.e., overlap in $D_{i,t}$ and $D_{j,t}$).

Step 5: Variance and Covariance Calculations We now analyze the properties of the speculative component:

$$w_{i,t} = \frac{1}{\gamma N}\Lambda_t^{-1}D_{i,t}\tilde{\varepsilon}_t^{(new)} \tag{194}$$

where we include the $1/N$ scaling factor to keep typical dollar holdings stable as N grows, consistent with Proposition 7.

From Proposition 6's Bayesian structure (under correctly specified priors), the learning agents'

aggregate estimation error has covariance proportional to Λ_t:

$$\Lambda_{a,t}^{(L)} = \text{Var}(\tilde{\varepsilon}_t^{(L)}) = \frac{1}{t + \kappa_E}\Lambda_t \tag{195}$$

where $\kappa_E \geq 0$ is the effective aggregate prior precision. Since $\tilde{\varepsilon}_t^{(new)} = \rho^* \tilde{\varepsilon}_t^{(L)}$ (Step 3), proportionality is preserved:

$$\Lambda_{a,t}^{(new)} = \text{Var}(\tilde{\varepsilon}_t^{(new)}) = (\rho^*)^2 \frac{1}{t + \kappa_E}\Lambda_t \tag{196}$$

Define the trace and diagonal terms for the new error covariance:

$$T \equiv \text{Tr}(\Lambda_t^{-1}\Lambda_{a,t}^{(new)}), \qquad \Delta \equiv \sum_{k=1}^{N} [\Lambda_t^{-1}]_{kk}[\Lambda_{a,t}^{(new)}]_{kk}. \tag{197}$$

Under proportionality, $T = (\rho^*)^2 N/(t + \kappa_E)$ and $\Delta/T = \chi_t \equiv \frac{1}{N}\sum_k [\Lambda_t^{-1}]_{kk}[\Lambda_t]_{kk} \geq 1$. The inequality $\chi_t \geq 1$ follows from Cauchy-Schwarz (Proposition 7 proof). Hence $\Delta \geq T$.

Expected Covariance of Strategies i and j ($i \neq j$):

For the covariance between two different sophisticated investors, the expectation over independent $D_{i,t}$ and $D_{j,t}$ yields (using independence and $\mathbb{E}[D_{i,t}] = \phi I_N$):

$$\mathbb{E}[\text{Cov}_t(R_{w_{i,t}}, R_{w_{j,t}})] = \frac{1}{\gamma^2 N^2}\text{Tr}(\mathbb{E}[D_{i,t}]\Lambda_t^{-1}\mathbb{E}[D_{j,t}]\Lambda_{a,t}^{(new)}) \tag{198}$$

$$= \frac{\phi^2}{\gamma^2 N^2}\text{Tr}(\Lambda_t^{-1}\Lambda_{a,t}^{(new)}) = \frac{\phi^2}{\gamma^2 N^2}T \tag{199}$$

This covariance formula holds in full generality (no diagonality required), since independence allows us to factor expectations.

Step 6: General Formulas for Correlation and Sharpe Ratio We now state general formulas in terms of T and Δ (defined in Step 5), then show how the model's proportionality yields the χ_t special cases.

General Expected Moments:

The expected return and variance of the speculative component are:

$$\mathbb{E}[R_{w_{i,t}}] = \frac{\phi}{\gamma N} T \tag{200}$$

$$\mathbb{E}[\mathrm{Var}_t(R_{w_{i,t}})] = \frac{1}{\gamma^2 N^2} \left[\phi^2 T + \phi(1-\phi)\Delta\right] \tag{201}$$

For the covariance between two different sophisticated investors ($i \neq j$):

$$\mathbb{E}[\mathrm{Cov}_t(R_{w_{i,t}}, R_{w_{j,t}})] = \frac{\phi^2}{\gamma^2 N^2} T \tag{202}$$

Result (ii): Ex-Ante Correlation (General Formula) Define the ex-ante correlation as

$$\mathrm{Corr}^{ex\text{-}ante}(R_{w_{i,t}}, R_{w_{j,t}}) \equiv \frac{\mathbb{E}[\mathrm{Cov}_t(R_{w_{i,t}}, R_{w_{j,t}})]}{\sqrt{\mathbb{E}[\mathrm{Var}_t(R_{w_{i,t}})] \cdot \mathbb{E}[\mathrm{Var}_t(R_{w_{j,t}})]}} \tag{203}$$

Substituting our expressions:

$$\mathrm{Corr}^{ex\text{-}ante}(R_{w_{i,t}}, R_{w_{j,t}}) = \frac{\frac{\phi^2}{\gamma^2 N^2} T}{\frac{1}{\gamma^2 N^2}\left[\phi^2 T + \phi(1-\phi)\Delta\right]} = \frac{\phi^2 T}{\phi^2 T + \phi(1-\phi)\Delta} = \frac{\phi}{\phi + (1-\phi)\Delta/T} \tag{204}$$

This formula holds in full generality. The ratio Δ/T determines the ex-ante correlation, which always lies in $(0, 1)$. Under the model's proportionality (Proposition 6), $T = cN$ where $c = (\rho^*)^2/(t + \kappa_E)$, and $\Delta = cN\chi_t$ where $\chi_t \equiv \frac{1}{N}\sum_k [\Lambda_t^{-1}]_{kk}[\Lambda_t]_{kk} \geq 1$ is the diagonal overlap coefficient. Hence $\Delta/T = \chi_t$ and the correlation becomes $\phi/[\phi + (1-\phi)\chi_t] \leq \phi$, bounded above by ϕ since $\chi_t \geq 1$ by Cauchy-Schwarz.

The squared Sharpe ratio is therefore:

$$SR_{i,t}^2 = \frac{(\mathbb{E}[R])^2}{\mathbb{E}[\mathrm{Var}]} = \frac{\left(\frac{\phi}{\gamma N} T\right)^2}{\frac{1}{\gamma^2 N^2}\left[\phi^2 T + \phi(1-\phi)\Delta\right]} = \frac{\phi^2 T^2}{\phi^2 T + \phi(1-\phi)\Delta} \tag{205}$$

Dividing numerator and denominator by ϕ:

$$SR_{i,t}^2 = \frac{\phi T^2}{\phi T + (1-\phi)\Delta} \tag{206}$$

This is the general formula, valid for any positive definite $\Lambda_{a,t}$.

Result (iii): Properties of the General Sharpe Ratio This formula has several important properties:

1. Monotonicity in Information Coverage: $SR_{i,t}^2$ is strictly increasing in ϕ. To see this, write $SR^2(\phi) = \frac{\phi T^2}{\Delta + \phi(T-\Delta)}$. Then

$$\frac{d}{d\phi} SR^2(\phi) = \frac{T^2 \Delta}{(\Delta + \phi(T-\Delta))^2} > 0 \tag{207}$$

   for $T > 0$, $\Delta > 0$. The limits are economically sensible:

   - As $\phi \to 1$ (full information): $SR^2 \to \frac{T^2}{T} = T$
   - As $\phi \to 0$ (no information): $SR^2 \to 0$

2. Scaling from Proposition 6: From Proposition 6's pricing error bound:

$$T = \mathrm{Tr}(\Lambda_t^{-1} \Lambda_{a,t}^{(new)}) \sim O(N/t) \tag{208}$$

   For the diagonal term Δ, we state a concrete sufficient condition:

   Assumption (Bounded Eigenvalues and Diagonal Scaling): There exist constants $0 < \underline{\lambda} \leq \bar{\lambda} < \infty$ such that $\underline{\lambda} \leq \lambda_{\min}(\Lambda_t) \leq \lambda_{\max}(\Lambda_t) \leq \bar{\lambda}$ uniformly in N and t, and the average diagonal element of the belief covariance satisfies $\frac{1}{N} \sum_{k=1}^{N} [\Lambda_{a,t}^{(new)}]_{kk} = O(1/t)$.

   Under this assumption:

   - Bounded eigenvalues imply $[\Lambda_t^{-1}]_{kk} \in [\bar{\lambda}^{-1}, \underline{\lambda}^{-1}]$, so diagonal elements of Λ_t^{-1} are $O(1)$

- The diagonal scaling condition ensures $\frac{1}{N}\sum_{k=1}^{N}[\Lambda_{a,t}^{(new)}]_{kk} = O(1/t)$
- Therefore $\Delta = \sum_{k=1}^{N}[\Lambda_t^{-1}]_{kk}[\Lambda_{a,t}^{(new)}]_{kk} \leq \underline{\lambda}^{-1}\sum_{k=1}^{N}[\Lambda_{a,t}^{(new)}]_{kk} = O(N/t)$

Since both T and Δ scale as $O(N/t)$:

$$SR^2 = \frac{\phi \cdot O(N^2/t^2)}{\phi \cdot O(N/t) + (1-\phi)\cdot O(N/t)} = \frac{\phi \cdot O(N^2/t^2)}{O(N/t)} = \phi \cdot O(N/t) \tag{209}$$

Therefore:

$$SR_{i,t} \sim O\left(\sqrt{\phi \cdot \frac{N}{t}}\right) \tag{210}$$

This scaling relies on Proposition 6's pricing error bound and standard scaling assumptions for covariance matrices.

Economic Interpretation The economic insights from Propositions 6 and 7 are robust to a rational expectations equilibrium with positive mass of sophisticated investors.

1. Pricing Errors Persist: Mispricing scales as $O(N/t)$ from Proposition 6. Sophisticated capital with effective coverage $\theta_s\phi$ attenuates the error by factor $\rho^* = \frac{\Omega_t^{(L)}}{\Omega_t^{(L)}+\theta_s\phi} \in (0,1)$, but does not eliminate the fundamental scaling. As long as learning agents have positive effective mass ($\Omega_t^{(L)} > 0$), pricing errors remain.

2. Sharpe Ratios Scale with Complexity and Information: The general formula (Equation 206) establishes: (a) $SR \sim O(\sqrt{\phi N/t})$ scaling from Proposition 6's pricing error bound, (b) monotonicity in ϕ, and (c) attenuation by ρ^* as informed capital enters.

3. Information Dispersion Limits Arbitrage: Independent information sets ($D_{i,t}$ drawn independently) prevent sophisticated investors from collectively eliminating mispricing. Each investor exploits different dimensions of $\tilde{\varepsilon}_t^{(new)}$, maintaining profitability.

4. Low Ex-Ante Correlations: Under the model's proportionality, $\text{Corr}^{ex\text{-}ante} = \phi/[\phi + (1-\phi)\chi_t] \leq \phi$ (Equation 204). Sophisticated investors rationally use the equilibrium belief

$\hat{a}_{E,t}^{(new)}$ for unobserved assets, eliminating any common signal component.

As markets become more complex (N increases), opportunities for sophisticated investors scale as $\sqrt{N/t}$, encouraging entry and specialization. The proliferation of profitable, specialized strategies with low cross-correlations is a robust prediction of the general T/Δ framework.

The limiting case $\theta_s \to 0$ recovers Proposition 7 exactly, with $\rho^* \to 1$, $\hat{a}_{E,t}^{(new)} \to \hat{a}_{E,t}^{(L)}$, and $\tilde{\varepsilon}_t^{(new)} \to \tilde{\varepsilon}_t^{(L)}$.

□

6.11 Undiversifiable Estimation Uncertainty

Overview

This appendix establishes a lower bound on pricing errors relative to the best implementable mean-variance portfolio. Expected unspanned squared Sharpe ratio is bounded below by terms proportional to N/t and N, where $N \to \infty$ with $K = o(N)$ and t possibly growing.[21] The first component is learnable and shrinks with sample size. The second is persistent, reflecting time-varying fundamentals that learning does not eliminate. Holding t fixed, the lower bound is linear in N; as $t \to \infty$, only the N/t component vanishes.

The result characterizes irreducible mispricing for any estimator based on public information. Among $\mathcal{G}_t$-measurable portfolios, $w_t = \Lambda_t^{-1}\hat{\mu}_t$ maximizes expected Sharpe ratio given the information available to agents. The $\Omega(N)$ lower bound on unspanned squared Sharpe ratio applies to every implementable portfolio (Corollary A.2), whether unconstrained or restricted to a factor subspace.

Sharper bounds under factor structure of Λ_t appear in Proposition A.8. We first establish the result assuming known Λ_t; the Unknown Covariance section shows the bound holds when agents estimate Λ_t.

21. We use standard asymptotic notation: $f(N) = O(g(N))$ means f is bounded above by a constant times g; $f(N) = \Omega(g(N))$ means f is bounded below by a constant times g; $f(N) = o(g(N))$ means $f/g \to 0$.

Setup

Notation. Let $r_{t+1} \in \mathbb{R}^N$ denote one-period excess returns with conditional mean $\mu_t \equiv \mathbb{E}_t[r_{t+1}]$ and covariance $\Lambda_t \equiv \mathrm{Cov}_t(r_{t+1}) \succ 0$, where $\mathbb{E}_t$ conditions on nature's information set $\mathcal{F}_t \supseteq \mathcal{G}_t$, which determines the true state of the world at time t. Let $x_{i,t} \in \mathbb{R}^L$ denote observable characteristics for asset i. The public information set is $\mathcal{G}_t = \sigma(\{p_s, r_s, x_s\}_{s \le t})$, comprising prices, returns, and characteristics through time t. Throughout, t denotes the length of return history in $\mathcal{G}_t$; calendar time and sample size coincide. Learning rates $O(1/t)$ refer to precision accumulating over this history. An estimator $\tilde{\mu}_t$ is any $\mathcal{G}_t$-measurable random vector. The best public-information approximation to μ_t is $\hat{\mu}_t \equiv \mathbb{E}[\mu_t \mid \mathcal{G}_t]$, and the residual uncertainty about μ_t is $V_t \equiv \mathbb{V}\mathrm{ar}(\mu_t \mid \mathcal{G}_t)$. Until the Unknown Covariance section, Λ_t is assumed $\mathcal{G}_t$-measurable.

Define $u_t \equiv \mu_t - \hat{\mu}_t$, the gap between nature's expected return and what agents can infer from observed data; by construction, $\mathbb{E}[u_t \mid \mathcal{G}_t] = 0$. Returns then decompose as

$$r_{t+1} = \hat{\mu}_t + u_t + \epsilon_{t+1},$$

where ϵ_{t+1} is mean-zero conditional on $\mathcal{F}_t$.

Pricing errors. Let $w_t = \Lambda_t^{-1}\hat{\mu}_t$ denote the (unnormalized) implementable tangency portfolio.[22] Define the beta of asset i with respect to the benchmark:

$$\beta_{i,t} \equiv \frac{\mathrm{Cov}_t(r_{i,t+1}, w_t' r_{t+1})}{\mathbb{V}\mathrm{ar}_t(w_t' r_{t+1})} = \frac{(\Lambda_t w_t)_i}{w_t' \Lambda_t w_t} = \frac{\hat{\mu}_{i,t}}{\hat{\mu}_t' \Lambda_t^{-1} \hat{\mu}_t}. \tag{211}$$

The pricing error (alpha) of asset i is expected return minus beta times the benchmark expected return:

$$\alpha_{i,t} \equiv \mu_{i,t} - \beta_{i,t} \cdot \mathbb{E}_t[w_t' r_{t+1}] = \mu_{i,t} - \frac{\hat{\mu}_{i,t} \cdot (\hat{\mu}_t' \Lambda_t^{-1} \mu_t)}{\hat{\mu}_t' \Lambda_t^{-1} \hat{\mu}_t}. \tag{212}$$

In vector form, $\alpha_t = \mu_t - \hat{\mu}_t (\hat{\mu}_t' \Lambda_t^{-1} \mu_t)/(\hat{\mu}_t' \Lambda_t^{-1} \hat{\mu}_t)$.

22. Since r_{t+1} denotes excess returns, $\hat{\mu}_t = 0$ would imply zero expected risk premia across all assets, which we rule out.

Geometric representation. The precision-weighted sum of squared alphas has a geometric interpretation. In whitened space, define

$$b_t \equiv \frac{\Lambda_t^{-1/2}\hat{\mu}_t}{\|\Lambda_t^{-1/2}\hat{\mu}_t\|}, \qquad P_t \equiv b_t b_t', \tag{213}$$

so P_t projects onto the direction of the implementable tangency portfolio. Then

$$\begin{aligned}\alpha_t'\Lambda_t^{-1}\alpha_t &= \mu_t'\Lambda_t^{-1}\mu_t - 2\frac{(\hat{\mu}_t'\Lambda_t^{-1}\mu_t)^2}{\hat{\mu}_t'\Lambda_t^{-1}\hat{\mu}_t} + \frac{(\hat{\mu}_t'\Lambda_t^{-1}\mu_t)^2}{\hat{\mu}_t'\Lambda_t^{-1}\hat{\mu}_t} \\ &= \|\Lambda_t^{-1/2}\mu_t\|^2 - (b_t'\Lambda_t^{-1/2}\mu_t)^2 = \|(I - P_t)\Lambda_t^{-1/2}\mu_t\|^2 \equiv \delta_{\perp,t}^2.\end{aligned} \tag{214}$$

Since $(I - P_t)\Lambda_t^{-1/2}\hat{\mu}_t = 0$, the unspanned component depends only on estimation error:

$$\delta_{\perp,t}^2 = \|(I - P_t)\Lambda_t^{-1/2}(\mu_t - \hat{\mu}_t)\|^2. \tag{215}$$

The oracle squared Sharpe ratio decomposes as

$$\|\Lambda_t^{-1/2}\mu_t\|^2 = (b_t'\Lambda_t^{-1/2}\mu_t)^2 + \delta_{\perp,t}^2. \tag{216}$$

The first term is spanned by the best implementable portfolio. The second is the residual that no public-information strategy can capture: the squared Sharpe ratio left on the table due to estimation uncertainty.

Assumptions

The assumptions are stated in terms of the risk-adjusted mean uncertainty matrix $A_t = \Lambda_t^{-1/2}V_t\Lambda_t^{-1/2}$, which measures mean uncertainty in units of return risk.

Assumption 3 (Risk-Adjusted Uncertainty)**.** *There exist constants $c > 0$, $\sigma^2 > 0$, and $t_0 < \infty$,*

independent of N, such that for all N and all $t \geq t_0$,

$$\frac{1}{N}\mathbb{E}[tr(A_t)] \geq \frac{c}{t} + \sigma^2. \tag{217}$$

The trace $\text{tr}(A_t) = \text{tr}(\Lambda_t^{-1} V_t)$ measures total mean uncertainty relative to return variance. The c/t term reflects learning from finite noisy signals: precision accumulates linearly, so estimation error decays as $O(1/t)$. The σ^2 term reflects the gap between $\mathcal{F}_t$ and $\mathcal{G}_t$: some information in nature's state has not yet been revealed through observable returns. Even with infinite data, time-varying fundamentals generate irreducible uncertainty. The condition is an average across assets and states.

Assumption 4 (Dispersed Risk-Adjusted Uncertainty)**.**

$$\mathbb{E}[\lambda_{\max}(A_t)] = o(N) \quad as\ N \to \infty. \tag{218}$$

No single portfolio direction absorbs a non-vanishing fraction of risk-adjusted mean uncertainty. A sufficient condition is uniformly bounded eigenvalues:

$$\sup_t \mathbb{E}[\lambda_{\max}(A_t)] < \infty. \tag{219}$$

Bounded diagonal entries $(A_t)_{ii} = O(1)$ and bounded absolute row sums $\sum_j |(A_t)_{ij}| = O(1)$ imply $\lambda_{\max}(A_t) = O(1)$ by Gershgorin's theorem. Economically, assets with high mean uncertainty also tend to have high return variance: the firm-specific events that make expected returns difficult to forecast also make realized returns volatile. The ratio $(A_t)_{ii}$ therefore remains bounded. The condition would fail for an asset whose expected return is highly uncertain but whose realized returns are very stable.

Remark 2 (Why Dispersion Drives the Bound)**.** *Assumption 4 is the central condition behind the $\Omega(N)$ scaling. The bound takes the form $\mathbb{E}[tr(A_t) - \lambda_{\max}(A_t)]$: total risk-adjusted uncertainty minus the largest eigenvalue. When $\mathbb{E}[\lambda_{\max}(A_t)] = o(N)$ while $\mathbb{E}[tr(A_t)] = \Omega(N)$, the gap scales*

as $\Omega(N) - o(N) = \Omega(N)$*. Equivalently,* $\mathbb{E}[\lambda_{\max}(A_t)]/\mathbb{E}[tr(A_t)] \to 0$*: no single direction explains a nontrivial fraction of total uncertainty. Any implementable portfolio spans a single direction in* N*-dimensional return space. If risk-adjusted uncertainty were concentrated in that direction, the portfolio could absorb it. Dispersion ensures that uncertainty spreads across many directions, which no single portfolio can span.*

Preliminary Results

The following results establish that the bound applies to any estimator based on public information.

Lemma A.2 (Dominance)**.** *For any* $\mathcal{G}_t$*-measurable estimator* $\tilde{\mu}_t$*, let* $\tilde{u}_t \equiv \mu_t - \tilde{\mu}_t$ *denote the estimation error. Then:*

$$\mathbb{E}[\tilde{u}_t \tilde{u}_t' \mid \mathcal{G}_t] = V_t + (\hat{\mu}_t - \tilde{\mu}_t)(\hat{\mu}_t - \tilde{\mu}_t)' \succeq V_t. \tag{220}$$

Suboptimal estimators can only increase error covariance.

Proof. Write $\tilde{u}_t = (\mu_t - \hat{\mu}_t) + (\hat{\mu}_t - \tilde{\mu}_t) = u_t + (\hat{\mu}_t - \tilde{\mu}_t)$. Since $\hat{\mu}_t - \tilde{\mu}_t$ is $\mathcal{G}_t$-measurable and $\mathbb{E}[u_t \mid \mathcal{G}_t] = 0$:

$$\mathbb{E}[\tilde{u}_t \tilde{u}_t' \mid \mathcal{G}_t] = \mathbb{V}\mathrm{ar}(u_t \mid \mathcal{G}_t) + (\hat{\mu}_t - \tilde{\mu}_t)(\hat{\mu}_t - \tilde{\mu}_t)' = V_t + (\hat{\mu}_t - \tilde{\mu}_t)(\hat{\mu}_t - \tilde{\mu}_t)'.$$

□

This lemma makes the result estimator-agnostic: the bound depends on V_t, which is an objective property of the DGP and information structure, not on any agent's choice of estimator.

Main Result

We now lower-bound the residual squared Sharpe ratio by the total risk-adjusted mean uncertainty outside the best one-dimensional subspace.

Lemma A.3 (Geometric Inequality)**.** *For any $\mathcal{G}_t$-measurable estimator $\tilde{\mu}_t$ and $\mathcal{G}_t$-measurable unit vector b_t, the unspanned risk-adjusted error satisfies*

$$\mathbb{E}\Big[\|(I - b_t b_t')\Lambda_t^{-1/2}(\mu_t - \tilde{\mu}_t)\|^2 \,\Big|\, \mathcal{G}_t\Big] \geq tr(A_t) - \lambda_{\max}(A_t), \tag{221}$$

where $A_t = \Lambda_t^{-1/2} V_t \Lambda_t^{-1/2}$. With eigenvalues ordered $\lambda_1(A_t) \geq \cdots \geq \lambda_N(A_t)$, the bound equals $\sum_{j=2}^{N} \lambda_j(A_t)$: variance outside the best rank-1 subspace.

Proof. Let $\tilde{u}_t = \mu_t - \tilde{\mu}_t$ and $\bar{A}_t = \Lambda_t^{-1/2}\mathbb{E}[\tilde{u}_t\tilde{u}_t' \mid \mathcal{G}_t]\Lambda_t^{-1/2}$. Then

$$\mathbb{E}[\|(I - b_t b_t')\Lambda_t^{-1/2}\tilde{u}_t\|^2 \mid \mathcal{G}_t] = \mathrm{tr}(\bar{A}_t) - b_t'\bar{A}_t b_t.$$

By Lemma A.2, $\mathbb{E}[\tilde{u}_t\tilde{u}_t' \mid \mathcal{G}_t] \succeq V_t$, so $\bar{A}_t \succeq A_t$. Let $\Delta = \bar{A}_t - A_t \succeq 0$. Then

$$[\mathrm{tr}(\bar{A}_t) - b_t'\bar{A}_t b_t] - [\mathrm{tr}(A_t) - b_t' A_t b_t] = \mathrm{tr}(\Delta) - b_t'\Delta b_t \geq 0$$

since $b_t'\Delta b_t \leq \mathrm{tr}(\Delta)$ for any unit vector and positive semidefinite matrix. Since $b_t' A_t b_t \leq \lambda_{\max}(A_t)$, the result follows. The first inequality is tight for the optimal estimator $\tilde{\mu}_t = \hat{\mu}_t$; the second is tight when b_t aligns with the top eigenvector of A_t. Since b_t is determined by $\hat{\mu}_t$, not by A_t, the overall bound $\mathrm{tr}(A_t) - \lambda_{\max}(A_t)$ is generally not achieved exactly. □

Theorem 1 (Undiversifiable Uncertainty)**.** *Under Assumptions 3–4, for N sufficiently large and $t \geq t_0$,*

$$\mathbb{E}[\delta_{\perp,t}^2] = \Omega(N/t) + \Omega(N). \tag{222}$$

The bound has two components:

- Learnable: $\Omega(N/t)$ from uncertainty about permanent expected returns. With infinite data, this component vanishes.
- Persistent: $\Omega(N)$ from time-varying expected returns. New shocks each period prevent learning from eliminating this component.

Even with infinite data and optimal estimation, the persistent component survives. Any implementable portfolio spans a single ray in N-dimensional return space, while estimation uncertainty disperses across all N coordinates. At most one dimension of uncertainty can be absorbed by the optimal direction; the remaining $N - 1$ dimensions contribute to the lower bound on pricing error.

Proof. By Lemma A.3, $\mathbb{E}[\delta_{\perp,t}^2 \mid \mathcal{G}_t] \geq \text{tr}(A_t) - \lambda_{\max}(A_t)$.

Taking expectations and applying Assumptions 3–4:

$$\mathbb{E}[\delta_{\perp,t}^2] \geq \mathbb{E}[\text{tr}(A_t)] - \mathbb{E}[\lambda_{\max}(A_t)] \tag{223}$$

$$\geq N\left(\frac{c}{t} + \sigma^2\right) - o(N), \tag{224}$$

where c and σ^2 are the constants from Assumption 3. Hence $\mathbb{E}[\delta_{\perp,t}^2] = \Omega(N/t) + \Omega(N)$: the first component from learning, the second from persistent uncertainty. □

Corollary A.2 (Bound for Arbitrary Estimators)**.** *Let $\tilde{\mu}_t$ be any $\mathcal{G}_t$-measurable estimator of μ_t, and define $\tilde{\delta}_{\perp,t}^2 \equiv \|(I - \tilde{P}_t)\Lambda_t^{-1/2}\mu_t\|^2$ where $\tilde{P}_t$ projects onto the direction of $\Lambda_t^{-1/2}\tilde{\mu}_t$. Under Assumptions 3–4, for N sufficiently large and $t \geq t_0$:*

$$\mathbb{E}[\tilde{\delta}_{\perp,t}^2] = \Omega(N/t) + \Omega(N).$$

The bound depends on $A_t = \Lambda_t^{-1/2} V_t \Lambda_t^{-1/2}$, not on the estimator's optimality.

Proof. Since $\tilde{P}_t$ projects onto $\Lambda_t^{-1/2}\tilde{\mu}_t$, we have $(I - \tilde{P}_t)\Lambda_t^{-1/2}\tilde{\mu}_t = 0$, so $\tilde{\delta}_{\perp,t}^2 = \|(I - \tilde{P}_t)\Lambda_t^{-1/2}(\mu_t - \tilde{\mu}_t)\|^2$. Lemma A.3 with $b_t = \Lambda_t^{-1/2}\tilde{\mu}_t / \|\Lambda_t^{-1/2}\tilde{\mu}_t\|$ gives $\mathbb{E}[\tilde{\delta}_{\perp,t}^2 \mid \mathcal{G}_t] \geq \text{tr}(A_t) - \lambda_{\max}(A_t)$. Taking expectations and applying Assumptions 3–4 as in Theorem 1 yields the result. □

Sufficient Condition: Approximate Factor Structure

When Λ_t has approximate factor structure, the bound takes a sharper form. This section shows the general result nests factor-model specifications common in empirical work.

Assumption 5 (Factor Structure)**.** *Returns satisfy* $r_{t+1} = B_t f_{t+1} + e_{t+1}$*, where* $f_{t+1} \in \mathbb{R}^K$ *are common factors and* e_{t+1} *is idiosyncratic. The covariance is* $\Lambda_t = B_t \Sigma_{f,t} B_t' + \Psi_t$*, where* $B_t \in \mathbb{R}^{N \times K}$ *are loadings,* $\Sigma_{f,t}$ *is the factor covariance, and* Ψ_t *is diagonal with* $(\Psi_t)_{ii} \in [\underline{d}, \bar{d}]$*. The number of factors satisfies* $K = o(N)$*.*

Factor premium uncertainty and idiosyncratic mean uncertainty are orthogonal:

$$\mathbb{E}[(B_t \bar{f}_t - \mathbb{E}[B_t \bar{f}_t \mid \mathcal{G}_t])(\theta_t - \mathbb{E}[\theta_t \mid \mathcal{G}_t])' \mid \mathcal{G}_t] = 0,$$

where $\bar{f}_t = \mathbb{E}_t[f_{t+1}]$ *and* $\theta_t = \mathbb{E}_t[e_{t+1}]$*. Factor premia reflect aggregate risk compensation; idiosyncratic expected returns reflect firm-specific fundamentals. Factor premia and idiosyncratic expected returns arise from distinct information sources.*

Role of orthogonality. Under Assumption 5, expected returns decompose as $\mu_t = B_t \bar{f}_t + \theta_t$. The orthogonality condition yields

$$V_t \equiv \mathbb{V}\text{ar}(\mu_t \mid \mathcal{G}_t) = B_t \mathbb{V}\text{ar}(\bar{f}_t \mid \mathcal{G}_t) B_t' + V_t^\theta \succeq V_t^\theta. \tag{225}$$

Theorem 1 applies to any DGP satisfying Assumptions 3–4, using full V_t. Orthogonality is not required for the bound to hold. What orthogonality provides is the decomposition $V_t \succeq V_t^\theta$: a weaker bound using only idiosyncratic uncertainty V_t^θ would also be valid. Without orthogonality, negative covariance between factor premium uncertainty and idiosyncratic uncertainty could violate $V_t \succeq V_t^\theta$, precluding the "idiosyncratic only" interpretation, but the general bound remains in force.

Under Assumption 5, define

$$S_t = \Psi_t^{-1/2} V_t \Psi_t^{-1/2},$$

measuring mean uncertainty in idiosyncratic-risk units.

Proposition A.8 (Factor-Structure Bound)**.** *Under Assumptions 3–4 and 5,*

$$\mathbb{E}[\delta_{\perp,t}^2] \geq \mathbb{E}\big[tr(S_t) - (K+1)\lambda_{\max}(S_t)\big]. \tag{226}$$

Proof. By the Woodbury identity, $\Lambda_t^{-1} = \Psi_t^{-1} - R_t$ where

$$R_t = \Psi_t^{-1} B_t (\Sigma_{f,t}^{-1} + B_t' \Psi_t^{-1} B_t)^{-1} B_t' \Psi_t^{-1} \succeq 0$$

and $\text{rank}(R_t) \leq K$. Define $A_t = \Lambda_t^{-1/2} V_t \Lambda_t^{-1/2}$.

Trace bound. We have

$$\text{tr}(A_t) = \text{tr}(\Lambda_t^{-1} V_t) = \text{tr}(\Psi_t^{-1} V_t) - \text{tr}(R_t V_t) = \text{tr}(S_t) - \text{tr}(R_t V_t). \tag{227}$$

Let $M_t = \Psi_t^{1/2} R_t \Psi_t^{1/2}$. Since $\Lambda_t^{-1} = \Psi_t^{-1} - R_t \succeq 0$, we have $\Psi_t^{-1} \succeq R_t$, hence $I \succeq \Psi_t^{1/2} R_t \Psi_t^{1/2} = M_t \succeq 0$. Also $\text{rank}(M_t) \leq K$, so $\text{tr}(M_t) \leq K$. Then

$$\text{tr}(R_t V_t) = \text{tr}(M_t S_t) \leq \text{tr}(M_t) \lambda_{\max}(S_t) \leq K \lambda_{\max}(S_t).$$

Hence $\text{tr}(A_t) \geq \text{tr}(S_t) - K \lambda_{\max}(S_t)$.

Eigenvalue bound. Since $\Lambda_t = B_t \Sigma_{f,t} B_t' + \Psi_t \succeq \Psi_t$, by the generalized Rayleigh quotient:

$$\lambda_{\max}(A_t) = \max_{y \neq 0} \frac{y' V_t y}{y' \Lambda_t y} \leq \max_{y \neq 0} \frac{y' V_t y}{y' \Psi_t y} = \lambda_{\max}(S_t). \tag{228}$$

Combine. $\text{tr}(A_t) - \lambda_{\max}(A_t) \geq \text{tr}(S_t) - (K+1) \lambda_{\max}(S_t)$. □

The $(K+1)$ factor arises from two sources: K directions absorbed by the factor structure in the trace bound, plus one direction absorbed by the best implementable portfolio in the eigenvalue bound. Since $\delta_{\perp,t}^2 \geq 0$, the bound is equivalently $\mathbb{E}[\max\{0, \text{tr}(S_t) - (K+1)\lambda_{\max}(S_t)\}]$; under dispersion the argument of the max is positive and scales as $\Omega(N)$.

Remark 3 (Dispersion under Factor Structure)**.** *Under factor structure, the Proposition A.8 bound scales as $\Omega(N)$ when $K \cdot \mathbb{E}[\lambda_{\max}(S_t)] = o(N)$. Combined with $\mathbb{E}[tr(S_t)] \geq \mathbb{E}[tr(A_t)] = \Omega(N)$ from Assumption 3, this gives $\mathbb{E}[tr(S_t) - (K+1)\lambda_{\max}(S_t)] = \Omega(N) - o(N) = \Omega(N)$. Bounded per-firm idiosyncratic mean uncertainty gives $\mathbb{E}[\lambda_{\max}(S_t)] = O(1)$, so the condition reduces to $K = o(N)$.*

Leading Example: Permanent and Transitory Components

The following specification satisfies Assumptions 3–5 and provides economic content for the bound's two components.

Expected idiosyncratic returns have permanent and time-varying parts:

$$\theta_{i,t} = \phi_i + \psi_{i,t}, \tag{229}$$

where $\phi_i \overset{\text{iid}}{\sim} (0, \sigma_\phi^2)$ across assets is a time-invariant firm effect, and $\psi_{i,t}$ follows

$$\psi_{i,t} = \rho\psi_{i,t-1} + \eta_{i,t}, \qquad |\rho| < 1, \qquad \eta_{i,t} \overset{\text{iid}}{\sim} (0, \sigma_\eta^2) \tag{230}$$

with $\eta_{i,t}$ independent across assets and time.

Observation structure. The idiosyncratic return at time t is $e_{i,t} = \phi_i + \psi_{i,t-1} + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is mean-zero noise. Because $e_{i,t}$ loads on $\psi_{i,t-1}$ rather than $\psi_{i,t}$, returns observed through time t depend only on innovations $\eta_{i,1}, \ldots, \eta_{i,t-1}$. The current innovation $\eta_{i,t}$ is independent of all observed data: $\eta_{i,t} \perp \mathcal{G}_t$.

Lemma A.4 (Learning Rates)**.** *Under the specification above, with $e_{i,s} = \phi_i + \psi_{i,s-1} + \varepsilon_{i,s}$ observed for $s = 1, \ldots, t$:*

1. $\mathbb{V}ar(\phi_i \mid \mathcal{G}_t) = O(1/t)$.
2. $\mathbb{V}ar(\psi_{i,t-1} \mid \mathcal{G}_t) \to P_{22}^*$ *as* $t \to \infty$, *where* P_{22}^* *depends on* ρ, σ_η^2, *and* σ_ε^2 *but not on* t.
3. $\mathbb{V}ar(\phi_i + \rho\psi_{i,t-1} \mid \mathcal{G}_t) = O(1/t) + O(1)$.

Proof. The state $\xi_{i,t} = (\phi_i, \psi_{i,t})'$ evolves as $\xi_{i,t} = F\xi_{i,t-1} + (0, \eta_{i,t})'$ with $F = \text{diag}(1, \rho)$, observed through $e_{i,s} = H\xi_{i,s-1} + \varepsilon_{i,s}$ where $H = (1, 1)$. The Kalman filter posterior covariance P_t satisfies the Riccati recursion with prediction step $P_{t|t-1} = FP_{t-1|t-1}F' + Q$, $Q = \text{diag}(0, \sigma_\eta^2)$, and update step $P_{t|t} = P_{t|t-1} - P_{t|t-1}H'(HP_{t|t-1}H' + \sigma_\varepsilon^2)^{-1}HP_{t|t-1}$. Because $|\rho| < 1$ and ψ receives fresh

noise each period, the $(2,2)$ block converges to a steady-state value P_{22}^*, giving part (2). For ϕ_i: each observation is $e_{i,s} = \phi_i + \nu_{i,s}$ where $\nu_{i,s} = \psi_{i,s-1} + \varepsilon_{i,s}$ is stationary with absolutely summable autocovariances (since $|\rho| < 1$). The posterior variance $(P_t)_{11}$ is bounded above by the MSE of the sample mean $\bar{e}_{i,t}$ (the initial condition is irrelevant since $|\rho| < 1$), which equals $t^{-2}\sum_{s,s'}\mathrm{Cov}(\nu_{i,s},\nu_{i,s'}) = O(1/t)$, giving part (1). For part (3), the cross term satisfies $|\mathrm{Cov}(\phi_i, \psi_{i,t-1} \mid \mathcal{G}_t)| \le \sqrt{O(1/t)\cdot O(1)} = o(1)$. □

Verification of assumptions. Since $\eta_{i,t}$ is independent of $(\phi_i, \psi_{i,t-1}, \mathcal{G}_t)$, the conditional variance decomposes as

$$\mathbb{V}\mathrm{ar}(\theta_{i,t} \mid \mathcal{G}_t) = \underbrace{\mathbb{V}\mathrm{ar}(\phi_i + \rho\psi_{i,t-1} \mid \mathcal{G}_t)}_{\text{learnable}} + \underbrace{\sigma_\eta^2}_{\text{irreducible}}. \tag{231}$$

The second term equals σ_η^2 exactly because $\eta_{i,t} \perp \mathcal{G}_t$. For the first term, Lemma A.4 gives $\mathbb{V}\mathrm{ar}(\phi_i + \rho\psi_{i,t-1} \mid \mathcal{G}_t) = O(1/t) + O(1)$. The learnable component is in fact $\Theta(1/t)$: the noise $\nu_{i,s} = \psi_{i,s-1} + \varepsilon_{i,s}$ is stationary with spectral density $f(0) = \sigma_\varepsilon^2 + \sigma_\eta^2/(1-\rho)^2 > 0$ at frequency zero, so posterior precision for ϕ_i grows at rate $t/f(0)$, giving $\mathbb{V}\mathrm{ar}(\phi_i \mid \mathcal{G}_t) \ge f(0)/t \cdot (1 + o(1))$. Only $\sigma^2 > 0$ is needed for the $\Omega(N)$ persistent component. As $t \to \infty$, the floor is $\rho^2 P_{22}^* + \sigma_\eta^2 > 0$.

Factor premia $\bar{f}_t$ are treated as known in this example, so $V_t = V_t^\theta$; including factor premium uncertainty would only strengthen the bound by Proposition A.9. Since ϕ_i, $\psi_{i,t}$, and $\eta_{i,t}$ are independent across assets in the DGP and each firm's returns reveal only its own parameters, V_t^θ is diagonal.[23] By the eigenvalue bound in the proof of Proposition A.8, $\lambda_{\max}(A_t) \le \lambda_{\max}(S_t)$ where $S_t = \Psi_t^{-1/2} V_t^\theta \Psi_t^{-1/2}$. Since V_t^θ and Ψ_t are both diagonal with per-firm entries of order $O(1)$, $\lambda_{\max}(S_t) = O(1)$, satisfying Assumption 4. For Assumption 3, each diagonal entry of S_t satisfies $(S_t)_{ii} = \mathbb{V}\mathrm{ar}(\theta_{i,t} \mid \mathcal{G}_t)/d_i \ge (\rho^2 P_{22}^* + \sigma_\eta^2)/\bar{d}$ for large t, so

$$\frac{1}{N}\mathrm{tr}(S_t) \ge \frac{\rho^2 P_{22}^* + \sigma_\eta^2}{\bar{d}} > 0.$$

23. Factor premia are treated as known, so the information set relevant to firm i's idiosyncratic state $(\phi_i, \psi_{i,t})$ consists of firm i's own idiosyncratic returns. Because idiosyncratic innovations are independent across firms, the posterior distributions of $\theta_{i,t}$ and $\theta_{j,t}$ are conditionally independent given $\mathcal{G}_t$ for $i \ne j$, so $\mathrm{Cov}(\theta_{i,t}, \theta_{j,t} \mid \mathcal{G}_t) = 0$.

The trace bound $\text{tr}(A_t) \geq \text{tr}(S_t) - K\lambda_{\max}(S_t)$ with $K = o(N)$ then gives $\frac{1}{N}\mathbb{E}[\text{tr}(A_t)] \geq c/t + \sigma^2$ for $\sigma^2 = (\rho^2 P_{22}^* + \sigma_\eta^2)/\bar{d}$ and large N.

Unknown Factor Premia

The preceding results use full V_t, which includes both idiosyncratic and factor premium uncertainty. A simple monotonicity argument shows that a weaker bound using only V_t^θ would also be valid: including factor premium uncertainty cannot reduce the bound.

Define $f(A) = \text{tr}(A) - \lambda_{\max}(A)$ for symmetric A. By Lemma A.3, $\mathbb{E}[\delta_{\perp,t}^2 \mid \mathcal{G}_t] \geq f(A_t)$ where $A_t = \Lambda_t^{-1/2} V_t \Lambda_t^{-1/2}$.

Proposition A.9 (Monotonicity)**.** *For PSD matrices A_0, Δ with $A_1 = A_0 + \Delta$,*

$$f(A_1) \geq f(A_0). \tag{232}$$

Proof. We have

$$f(A_1) - f(A_0) = \text{tr}(\Delta) - [\lambda_{\max}(A_1) - \lambda_{\max}(A_0)].$$

By Weyl's inequality, $\lambda_{\max}(A_1) \leq \lambda_{\max}(A_0) + \lambda_{\max}(\Delta)$. For any PSD matrix, $\lambda_{\max}(\Delta) \leq \text{tr}(\Delta)$. Hence $f(A_1) - f(A_0) \geq \text{tr}(\Delta) - \lambda_{\max}(\Delta) \geq 0$. □

Corollary A.3 (Bound with Unknown Factor Premia)**.** *Under Assumptions 3–4 and 5, including uncertainty in factor risk premia $\bar{f}_t$ cannot reduce the bound on $\mathbb{E}[\delta_{\perp,t}^2]$. A bound using only idiosyncratic uncertainty V_t^θ would be weaker but still valid.*

Proof. With $\text{Cov}(B_t \bar{f}_t, \theta_t \mid \mathcal{G}_t) = 0$, we have $V_t = B_t \Sigma_{\bar{f},t} B_t' + V_t^\theta \succeq V_t^\theta$. Define $A_t = \Lambda_t^{-1/2} V_t \Lambda_t^{-1/2}$ and $A_t^\theta = \Lambda_t^{-1/2} V_t^\theta \Lambda_t^{-1/2}$. Then $A_t = A_t^\theta + \Delta$ for PSD Δ. By Proposition A.9, $f(A_t) \geq f(A_t^\theta)$. A bound stated in terms of A_t^θ alone would therefore remain valid, and Proposition A.8 (which uses full A_t) is at least as strong. □

Unknown Covariance

The preceding results assume Λ_t is $\mathcal{G}_t$-measurable. We now relax this assumption: agents may form any $\mathcal{G}_t$-measurable estimator $\tilde{\Lambda}_t \succ 0$ and choose any $\mathcal{G}_t$-measurable portfolio w_t (e.g., $w_t = \tilde{\Lambda}_t^{-1}\tilde{\mu}_t$), but pricing errors are evaluated using true conditional second moments. Unknown covariance can only increase pricing errors, so the bound applies a fortiori.

Assumption 6 (Covariance-Mean Independence)**.** $u_t \perp\!\!\!\perp \Lambda_t \mid \mathcal{G}_t$.

Observing Λ_t provides no incremental information about mean estimation errors beyond public history. The condition holds when mean innovations and covariance dynamics arise from distinct information sources: mean innovations reflect firm-specific fundamentals, while covariance dynamics reflect aggregate risk and volatility.

Remark 4 (Assumptions under Unknown Covariance)**.** *The proof of Theorem 2 conditions on $\sigma(\mathcal{G}_t, \Lambda_t)$ and uses $\tilde{V}_t = \mathbb{V}ar(u_t \mid \mathcal{G}_t, \Lambda_t)$. Under Assumption 6, the conditional distribution of u_t given $(\mathcal{G}_t, \Lambda_t)$ equals its distribution given $\mathcal{G}_t$ alone, so $\tilde{V}_t = V_t$, $\tilde{A}_t = A_t$, and Assumptions 3–4 as originally stated are exactly the conditions needed. Agents are not required to observe Λ_t; the assumptions are properties of the DGP.*

Lemma A.5 (Alphas for an Arbitrary Benchmark)**.** *Fix any $\mathcal{G}_t$-measurable portfolio w_t. Define*

$$\beta_t(w) = \frac{Cov_t(r_{t+1}, w_t' r_{t+1})}{\mathbb{V}ar_t(w_t' r_{t+1})} = \frac{\Lambda_t w_t}{w_t' \Lambda_t w_t}, \qquad \alpha_t(w) = \mu_t - \beta_t(w) \cdot (w_t' \mu_t).$$

Then

$$\alpha_t(w)' \Lambda_t^{-1} \alpha_t(w) = \mu_t' \Lambda_t^{-1} \mu_t - \frac{(w_t' \mu_t)^2}{w_t' \Lambda_t w_t}.$$

Equivalently, letting $b_t(w) = \Lambda_t^{1/2} w_t / \|\Lambda_t^{1/2} w_t\|$ and $P_t(w) = b_t(w) b_t(w)'$,

$$\alpha_t(w)' \Lambda_t^{-1} \alpha_t(w) = \|(I - P_t(w)) \Lambda_t^{-1/2} \mu_t\|^2.$$

Proof. Write $\alpha_t(w) = \mu_t - \frac{\Lambda_t w_t}{w_t'\Lambda_t w_t}(w_t'\mu_t)$. Then

$$\begin{aligned}\alpha_t(w)'\Lambda_t^{-1}\alpha_t(w) &= \mu_t'\Lambda_t^{-1}\mu_t - 2\frac{(w_t'\mu_t)^2}{w_t'\Lambda_t w_t} + \frac{(w_t'\mu_t)^2}{w_t'\Lambda_t w_t} \\ &= \mu_t'\Lambda_t^{-1}\mu_t - \frac{(w_t'\mu_t)^2}{w_t'\Lambda_t w_t}.\end{aligned}$$

The second representation follows from

$$b_t(w)'\Lambda_t^{-1/2}\mu_t = \frac{w_t'\mu_t}{\|\Lambda_t^{1/2}w_t\|}, \qquad \|(I - P_t(w))x\|^2 = \|x\|^2 - (b_t(w)'x)^2.$$

□

Theorem 2 (Undiversifiable Uncertainty with Unknown Covariance)**.** *Under Assumptions 3–4 and 6, with Λ_t not required to be $\mathcal{G}_t$-measurable, for any $\mathcal{G}_t$-measurable portfolio w_t, N sufficiently large, and $t \geq t_0$,*

$$\mathbb{E}[\alpha_t(w)'\Lambda_t^{-1}\alpha_t(w)] = \Omega(N/t) + \Omega(N).$$

Covariance estimation error can only increase pricing errors.

Proof. Using the decomposition $\mu_t = \hat{\mu}_t + u_t$, Lemma A.5 gives

$$\alpha_t(w)'\Lambda_t^{-1}\alpha_t(w) = \|(I - P_t(w))\Lambda_t^{-1/2}\mu_t\|^2.$$

Expanding $\mu_t = \hat{\mu}_t + u_t$:

$$\begin{aligned}\|(I - P_t(w))\Lambda_t^{-1/2}\mu_t\|^2 = \|(I - P_t(w))\Lambda_t^{-1/2}\hat{\mu}_t\|^2 + \|(I - P_t(w))\Lambda_t^{-1/2}u_t\|^2 \\ + 2\langle(I - P_t(w))\Lambda_t^{-1/2}\hat{\mu}_t, (I - P_t(w))\Lambda_t^{-1/2}u_t\rangle.\end{aligned}$$

The cross term vanishes in expectation by Assumption 6: conditional independence implies $\mathbb{E}[u_t \mid \mathcal{G}_t, \Lambda_t] = \mathbb{E}[u_t \mid \mathcal{G}_t] = 0$. Let $M_t = \Lambda_t^{-1/2}(I - P_t(w))\Lambda_t^{-1/2}$. Since w_t is $\mathcal{G}_t$-measurable and $P_t(w) = b_t(w)b_t(w)'$ depends on w_t and Λ_t through the whitening $b_t(w) = \Lambda_t^{1/2}w_t/\|\Lambda_t^{1/2}w_t\|$, the

matrix M_t is $\sigma(\mathcal{G}_t, \Lambda_t)$-measurable. Since $\hat{\mu}_t$ is $\mathcal{G}_t$-measurable, the tower property gives

$$\mathbb{E}[\hat{\mu}_t' M_t u_t \mid \mathcal{G}_t] = \mathbb{E}\big[\mathbb{E}[\hat{\mu}_t' M_t u_t \mid \mathcal{G}_t, \Lambda_t] \,\big|\, \mathcal{G}_t\big] = \mathbb{E}\big[\hat{\mu}_t' M_t \mathbb{E}[u_t \mid \mathcal{G}_t, \Lambda_t] \,\big|\, \mathcal{G}_t\big] = 0.$$

Dropping the first (nonnegative) term,

$$\mathbb{E}[\alpha_t(w)' \Lambda_t^{-1} \alpha_t(w)] \geq \mathbb{E}[\|(I - P_t(w))\Lambda_t^{-1/2} u_t\|^2].$$

Let $b_t(w) = \Lambda_t^{1/2} w_t / \|\Lambda_t^{1/2} w_t\|$ so that $P_t(w) = b_t(w) b_t(w)'$. Define $\tilde{V}_t = \mathbb{V}\mathrm{ar}(u_t \mid \mathcal{G}_t, \Lambda_t)$ and $\tilde{A}_t = \Lambda_t^{-1/2} \tilde{V}_t \Lambda_t^{-1/2}$. Conditioning on $(\mathcal{G}_t, \Lambda_t)$, where $b_t(w)$ is determined,

$$\mathbb{E}[\|(I - P_t(w))\Lambda_t^{-1/2} u_t\|^2 \mid \mathcal{G}_t, \Lambda_t] = \mathrm{tr}(\tilde{A}_t) - b_t(w)' \tilde{A}_t b_t(w) \geq \mathrm{tr}(\tilde{A}_t) - \lambda_{\max}(\tilde{A}_t).$$

Taking expectations yields

$$\mathbb{E}[\|(I - P_t(w))\Lambda_t^{-1/2} u_t\|^2] \geq \mathbb{E}[\mathrm{tr}(\tilde{A}_t) - \lambda_{\max}(\tilde{A}_t)].$$

By Assumption 6, $\tilde{V}_t = V_t$ and $\tilde{A}_t = A_t$ (Remark 4), so Assumptions 3–4 apply directly, giving the stated bound. □

Discussion

Components of the bound. The $\Omega(N/t)$ component reflects signal extraction with finite-variance noise: precision accumulates linearly in sample size. The $\Omega(N)$ component reflects time-varying expected returns. Each period, new innovations shift expected returns, but the returns driven by those innovations have not yet been realized and are absent from observed data. No estimator can reduce uncertainty about innovations that have not yet appeared in any return.

Cross-sectional learning. A larger cross-section means more coordinates to forecast, not better forecasting of any given coordinate. Any systematic correlation across firms constitutes a common

factor and should be absorbed into the factor structure. The residual after proper factor extraction is dispersed by construction (Assumption 4).

Implications

Why APT fails. Classical APT predicts that pricing errors vanish as markets grow: idiosyncratic return risk diversifies, so arbitrage portfolios can be constructed with vanishing variance. But the relevant object for pricing is estimation uncertainty in conditional means, not return risk. Each additional asset adds a new dimension to the unknown mean vector μ_t. Any implementable portfolio spans a single ray; the remaining $N - 1$ dimensions of estimation error are unspanned. Pricing errors grow as $\Omega(N)$ even though idiosyncratic return risk diversifies to zero.

Out-of-sample failure. The tangency portfolio $w^* \propto \Lambda_t^{-1}\hat{\mu}_t$, optimal given available information, leaves $\Omega(N)$ squared Sharpe ratio unspanned. Specifying K traded factors does not help: the best mean-variance portfolio in that class is still a single tangency direction, which is weakly dominated by w^*. Any factor model researchers propose therefore leaves at least $\Omega(N)$ unspanned. The oracle portfolio $\Lambda_t^{-1}\mu_t$ would span all expected returns, but estimating μ_t incurs irreducible $\Omega(N)$ error.

Anomaly abundance. The $\Omega(N)$ unspanned squared Sharpe ratio spreads across firm-specific coordinates. Because expected-return shocks are dispersed rather than factor-driven, no finite set of portfolios can span them. Each characteristic-sorted portfolio captures a different linear combination of unspanned expected return, consistent with the empirical proliferation of anomalies.

Oracle, not arbitrage. The theorem demonstrates existence of unspanned expected return. Constructing the oracle portfolio requires knowing μ_t directly, not estimating from data. The bound holds whenever prices are $\mathcal{G}_t$-measurable: the irreducible uncertainty arises from time-varying fundamentals and noisy signals, not from market frictions. Agents with private signals correlated with $\mu_t - \hat{\mu}_t$ can partially align with the oracle and earn positive expected alpha, but unless such signals aggregate into public information, the residual component remains unspanned. The result

characterizes fundamental limits of learning, not implementable arbitrage.