When Complexity Outpaces Factor Models *

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

[1] *Indiana University*

[2] *University of Florida*

This Version: March 2025

Abstract

We propose a solution to several asset pricing puzzles by investigating how complexity—the dimensionality of the parameter space relative to available sample size—affects pricing errors when rational, risk-averse agents have imperfect knowledge of the data-generating process. Our theory yields three implications as complexity increases: (1) equilibrium pricing errors grow systematically, (2) the optimal portfolio increasingly exploits estimation error components rather than fundamental risk, and (3) multiple strategies can achieve higher Sharpe ratios while maintaining low cross-correlations. Our model explains the limited pricing power of parsimonious factor models, the weak relationship between betas and average returns, and the proliferation of anomalies. Empirically, we document substantial complexity in return predictability and covariance structures. Consistent with our model, we find systematic hedge funds' active positions exhibit near-zero correlations (average R^2 below 1%), suggesting they can exploit distinct dimensions of estimation errors.

*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 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

The quest for parsimonious models linking expected returns to systematic risk has long defined asset pricing theory. Yet, despite the elegance of the capital asset pricing model (CAPM) and its multi-factor successors, these frameworks consistently struggle in empirical tests. Meanwhile, machine-learning strategies deliver high Sharpe ratios without explaining each other's returns, and hundreds of predictive signals defy ties to traditional risk factors. This gap between theoretical parsimony and empirical complexity poses a core puzzle: Why do factor models struggle to explain asset prices, and what drives the surge of diverse, profitable anomalies?

We suggest that complexity—defined as the dimensionality of the parameter space relative to available sample size—fundamentally underlies this puzzle. Our equilibrium model investigates how rational, risk-averse agents operate in high-dimensional markets when faced with uncertainty about expected returns and covariances. Without invoking market inefficiencies or behavioral biases, we show that estimation uncertainty alone can explain the empirical patterns observed in asset markets. As complexity increases through either more assets or predictive signals, pricing errors systematically rise, and optimal portfolios rely more on estimation errors than fundamental risk. This framework provides a unified explanation for factor model failures, the weak relationship between systematic risk and return, and the emergence of diverse, high-Sharpe-ratio strategies—all as natural consequences of rational optimization in complex market environments.

Our theoretical framework yields three main implications. First, we show that pricing errors of factor models increase linearly with complexity. This arises from a core tension: as complexity increases, the Sharpe ratio gap widens between the optimal portfolio under the true data-generating process, and any feasible portfolio investors can construct in equilibrium. The Gibbons, Ross, and Shanken (1989) framework links this divergence to larger pricing errors, as even factors perfectly capturing risk premia in simpler settings fail to span the growing array of estimation errors. In equilibrium, the market portfolio's Sharpe ratio, set by risk aversion and market risk, remains stable despite complexity, exposing opportunities for optimization beyond traditional risk premia.

Second, we show that optimal portfolio weights capitalize on these opportunities, increasingly deriving profitability from estimation errors rather than fundamental risk premia. Under uncertainty,

the optimal portfolio can be decomposed into three components: (1) the market portfolio, scaled by risk aversion, with a Sharpe ratio largely independent of complexity; (2) weights exploiting errors in expected payoffs; and (3) weights leveraging covariance estimation errors. As complexity grows, the Sharpe ratios of the latter two rise relative to the market's, fueled by abundant estimation uncertainty, while fundamental risk compensation stays relatively flat.

Third, we introduce a method to construct multiple trading strategies that achieve high Sharpe ratios while maintaining low correlation as market complexity rises. In high-dimensional markets, the increasing number of estimation errors creates multiple independent dimensions for exploitation. Strategies focusing on distinct error components naturally show low correlations and high Sharpe ratios, remaining unspanned by one another. As complexity rises, more strategies can attain a given Sharpe ratio while staying weakly correlated, and a fixed set of low-correlation strategies can achieve higher Sharpe ratios. Intuitively, rising complexity allows more ways to split the increasing number of estimation errors, letting strategies exploit distinct opportunities absent in simpler markets without overlap.

Together, our model proposes a unified solution to several asset pricing puzzles. The limited power of parsimonious factor models arises as complexity increases pricing errors, outpacing what simple frameworks can span. The weak link between betas and average returns follows from these errors, which grow with market dimensionality and obscure systematic risk's role (Daniel and Titman (1997), Lopez-Lira and Roussanov (2022)). Likewise, the surge of unrelated anomalies reflects abundant estimation errors that optimal portfolios exploit over fundamental risk. Our framework suggests that sophisticated quantitative strategies achieve high Sharpe ratios while remaining largely unspanned by one another, as they target orthogonal error dimensions (Baba Yara, Boyer, and Davis (2021)). Far from indicating inefficiency, these patterns may reflect rational responses to high dimensionality, questioning the view of anomalies as mispricing.

Moreover, we document pervasive complexity in the stock market, suggesting we operate in a high-dimensional market. We use the 17,517 characteristic-sorted portfolios from Chen, Lopez-Lira, and Zimmermann (2023). 347 portfolios (2%) exhibit absolute t-statistics above 3 over our

sample period, with 197 retaining such significance even after 2000, despite evidence of declining return predictability.[1] The covariance structure further reveals this dimensionality: the first ten principal components account for just 20% of return variance, while 75 factors are needed to capture 75%. This high-dimensional environment complicates estimating expected returns and covariances, supporting our model's assumptions.

Beyond supporting our model's high-dimensional assumptions, we also provide empirical evidence for its implications, particularly the emergence of diverse trading strategies in complex markets. Quantitative strategies show little overlap in their return structures: a diverse set of proposed factor models from the literature yields an average pairwise R^2 of just 7%, while the active components of systematic hedge fund portfolios, based on 13F holdings data, show an even lower pairwise average R^2 below 1%. This weak explanatory power among sophisticated investors' positions confirms our model's prediction that high-dimensional markets allow strategies to exploit distinct estimation error dimensions.

Related Literature

Our paper contributes to the literature on economies with high-dimensional properties, building on models that explore complexity in asset pricing. The closest work is Da, Nagel, and Xiu (2022), which adopts a partial equilibrium approach to examine the statistical limits of arbitrage. In their framework, alphas are assumed exogenous, focusing on characterizing attainable Sharpe ratios for investors under uncertainty. By contrast, we develop a general equilibrium model where Sharpe ratios, asset prices, and pricing errors emerge endogenously from the interplay of arbitrageurs' trading strategies and other investors' beliefs. Our equilibrium reveals that risk compensation remains largely insensitive to the economy's complexity while pricing errors increase markedly.

Martin and Nagel (2022) model an economy with numerous potential predictors, where risk-neutral Bayesian investors learn about expected returns. We advance their framework by incorporat-

1. See Harvey, Liu, and Zhu (2016), McLean and Pontiff (2016), Yan and Zheng (2017), Chen (2021), Chen and Zimmermann (2022).

ing risk-averse agents and uncertainty in the covariance matrix, enabling us to analyze risk pricing and the relationships among trading strategies. Similarly, Kozak and Nagel (2024) demonstrate that characteristics-based factors can span the stochastic discount factor (SDF) under covariance conditions more readily met with many characteristics. We build on their theoretical spanning results by illustrating why, in practice, such spanning becomes increasingly elusive as complexity rises, driven by estimation errors.

Andrei, Cujean, and Wilson (2023) illustrate how informational disparities between investors and empiricists can bias beta estimates, flattening the Security Market Line. We extend their framework by demonstrating that market complexity alone generates pricing errors, even without dispersed information, as the estimation challenge impacts all participants. Similarly, Merton (1987) employs incomplete information in asset pricing, emphasizing heterogeneous investor recognition as a driver of market outcomes. We build on this line of inquiry by exploring how market complexity alone can sustain persistent pricing anomalies, even among rational, homogeneous investors with shared information. In such high-dimensional settings, anomalies endure despite uniform rationality and access to data.

Finally, we advance the literature on high-dimensional asset pricing by offering an equilibrium-based explanation for the efficacy of complex factor models, such as those in Kelly, Malamud, and Zhou (2022), Bryzgalova, Pelger, and Zhu (2020a), and Didisheim et al. (2024). Our model rests on the assumption of return predictability, supported by a robust body of research, including Fama and MacBeth (1973), Jacobs and Levy (1988), Lewellen (2015), Gu, Kelly, and Xiu (2020), and Freyberger, Neuhierl, and Weber (2020a), among others.[2]

With this foundation in the literature on high-dimensional asset pricing, we now introduce our general equilibrium model to explore how rising complexity shapes market outcomes. Featuring rational agents under uncertainty, the model begins with a homogeneous agent economy to analyze pricing errors and optimal portfolio behavior, then extends to heterogeneous information settings to explain the emergence of diverse, uncorrelated trading strategies.

2. See Alti and Titman (2019) for an alternative model justifying return predictability.

1 Model

We work with economies with an increasing number of assets (N). 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]

Dividends serve as the primary driver of asset payoffs in our model. This design choice ensures that the covariance matrix of returns, denoted as $\Lambda_t = \mathbb{V}\text{ar}_t[d_{t+1}] = \mathbb{V}\text{ar}_t[r_{t+1}]$, remains independent of endogenous prices in the subsequent period, enhancing the model's traceability. We define the expected dividends under the true 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.

Initially, we assume a unit measure of identical agents in the economy, each endowed with constant absolute risk aversion (CARA) utility with risk aversion γ facing a per-asset supply of $\frac{1}{N}$ consistent with Kozak, Nagel, and Santosh (2018). Additionally, let ι denote the N-dimensional column vector of ones.[4] Later in this section, we extend the framework to accommodate heterogeneous agents.

1.1 Statistical Assumptions

Assume 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

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 remains well-behaved as N increases. Specifically, with a unit asset supply, $SR(w^*)^2 = \mu'\Lambda_t^{-1}\mu = \gamma^2\iota'\Lambda_t\Lambda_t^{-1}\Lambda_t\iota = \gamma^2\iota'\Lambda_t\iota \in O(N^2)$. For uncorrelated assets, $SR(w^*)^2 \in O(N)$.

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 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 $c_{i,t}$. This idiosyncratic component prevents the covariance structure of dividends from collapsing into a perfect factor structure, ensuring that estimation uncertainty persists as the number of assets N increases.

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 $x_t = (x_{1,t}, ..., x_{S,t}, z_{1,t}, ..., z_{N,t})'$ and B_t is the $N \times (S + N)$ matrix defined as

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

We assume that investors perfectly observe B_t. Additionally, conditional on the information set at time t, denoted $\mathcal{F}_t$, the state variables follow:

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

where π is the mean vector and V is the covariance matrix of X_{t+1}. Consequently, dividends follow a multivariate normal distribution with mean $B_t\pi$ and covariance B_tVB_t.

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

1.2 Agents' Estimation

In this economy, agents operate as Bayesian learners, facing uncertainty about both the mean and covariance matrix of the state variables. Specifically, we assume they hold normal-inverse-Wishart (NIW) priors over these parameters:

$$(\pi, V) \sim NIW(\pi_0, \kappa_0, V_0, \nu), \tag{5}$$

where π is the expected value of X_t and V its covariance matrix and π_0, κ_0, V_0, and ν denote the prior mean, prior precision scalar, prior covariance scale matrix, and degrees of freedom, respectively.

Given their perfect knowledge of B_t, agents can, at each time t, recover the realized vector of state variables X_{t+1} by projecting the observed dividends d_{t+1} onto B_t.

Let $\pi_t = (\kappa_0\pi_0 + t\bar{X}_t)/(t + \kappa_0)$ be the posterior mean. Let

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

be the posterior variance. Finally, let $\nu_t = \nu_0 + t - (S + N) + 1$. This implies that the posterior distribution is given by

$$X_{t+1} \sim t_\nu \left(\pi_t, \hat{V}_t \right), \tag{6}$$

where this notation indicates a multivariate t distribution. We assume that multivariate t distributions are intractable for investors, so they approximate the posterior using a multivariate normal distribution. In particular, investors believe:

$$X_{t+1} \sim N \left(\pi_t, V_t \right) \tag{7}$$

where $V_t = \frac{\nu_t}{\nu_t - 2} \hat{V}_t$ because of the normal approximation.

This implies investors believe that

$$r_{t+1} \sim N\left(\hat{u}_t, \hat{\Lambda}_t\right) \tag{8}$$

where $\hat{\mu}_t = \hat{a}_t - (1 + r_{f,t})p_t$, $a_t = B_t\mu_t$, and $\hat{\Lambda}_t = B_t V_t B_t'$.

1.3 Equilibrium

Because we are focusing on CARA investors with absolute risk aversion γ who model the world as normal, their demand has the form:

$$x_t = \frac{1}{\gamma}\hat{\Lambda}_t^{-1}(\hat{a}_t - (1 + r_f)p_t). \tag{9}$$

Where $\hat{\Lambda}_t$, and $\hat{a}_t$ are agents' estimates of the covariance matrix and the expected dividends. Agents hold the feasible mean-variance portfolios, which usually differ from the optimal portfolio under the true data-generating process portfolios due to imperfect information about Λ_t and a_t. In the extreme perfect knowledge case, for example, agents will have $\hat{\Lambda}_t = \Lambda_t$ and $\hat{a}_t = a_t$.

Plugging in the demand function and solving for the price leads to equilibrium prices:

$$p_t = \frac{1}{1 + r_f}\left(\hat{a} - \frac{\gamma}{N}\hat{\Lambda}_t\iota\right). \tag{10}$$

For notational simplicity, let $\gamma_N = \frac{\gamma}{N}$. The (data-generating process) expected excess returns are given by:

$$\text{Expected Returns} = \mu_t \equiv \mathbb{E}_t[d_{t+1}] - (1 + r_f)p_t \tag{11}$$

$$= \underbrace{a_t - \hat{a}_t}_{\text{means' estimation error}} + \underbrace{\gamma_N(\hat{\Lambda}_t - \Lambda_t)\iota}_{\text{covariance's estimation error}} + \underbrace{\gamma_N\Lambda_t\iota}_{\text{risk compensation}} \tag{12}$$

In contrast, subjective expected returns are given by $\gamma_N\Lambda_t\iota$. To have a better intuition on the risk

compensation, consider the case of uncorrelated normal factors (w.l.o.g., since there is a rotation to make the factor orthogonal). In this setting, the risk compensation term for an asset i simplifies to:

$$\gamma_N(\hat{\Lambda}_t\iota)_i \approx \gamma \sum_{k=1}^{S} b_{i,t,k}\bar{b}_{t,k}\sigma_{t,k}, \tag{13}$$

where the contribution of each factor k to the risk compensation is:

$$\gamma\bar{b}_{t,k}b_{i,t,k}\sigma_{t,k}, \tag{14}$$

with γ as the absolute risk aversion, $\bar{b}_{t,k} = \frac{1}{N}\sum_{j=1}^{N} b_{j,t,k}$ as the average loading (or beta) on factor k, and $b_{i,t,k}$ as asset i's loading (or beta) on factor k. For example, without estimation errors, a portfolio with a beta of 1 on factor m and zero on all other factors has an expected excess return of:

$$\mu_t^w = \gamma\bar{b}_{t,m}\sigma_{t,k}, \tag{15}$$

where $\gamma\bar{b}_{t,m}\sigma_{t,k}$ is the risk premium, often termed the *price of risk* for factor m, reflecting the compensation per unit of exposure to factor m based on the market's average loading and scaled by risk aversion.

A factor's exposure is only priced if its average beta across the market is non-zero. Specifically, the risk premium for a portfolio with unit exposure to factor m, given by $\mu_t^w = \gamma\bar{b}_{t,m}$, depends critically on $\bar{b}_{t,m} = \frac{1}{N}\sum_{j=1}^{N} b_{j,t,m}$. 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. In contrast, a non-zero $\bar{b}_{t,m}$ indicates a pervasive directional exposure, in which risk-averse agents demand compensation for bearing proportional to their collective risk aversion γ.

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 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 also partially explains why in Lopez-Lira and Roussanov (2022), it is optimal to hedge any factors (except for the first one). If the factors betas are net zero, then the Sharpe ratio is increasing in the number of hedged factors.

1.4 Objective Optimal Portfolio

The subjective optimal portfolio in equilibrium is given by the market portfolio. That is, prices adjust so that agents are satisfied with holding the market portfolio. In general, the objective optimal portfolio, which is the optimal portfolio under the data-generating process, will differ from the market portfolio.

Let $\sigma^2_{m,t} = \iota'\Lambda_t\iota$ denote the market variance and $\hat{\sigma}^2_{m,t} = \iota'\hat{\Lambda}_t\iota$ the perceived variance. The market variance does not depend on agents' expectations since prices adjust so that investors hold the same share of each asset, and the variance depends only on the dividends.

Hence, the vector of weights of the optimal (data-generating process) mean-variance portfolio is given by

$$w^*_t = \Lambda_t^{-1}\mu_t = \underbrace{\Lambda_t^{-1}(a_t - \hat{a}_t)}_{\text{errors in means}} + \underbrace{\gamma_N\Lambda_t^{-1}(\hat{\Lambda}_t - \Lambda_t)\,\iota}_{\text{errors in covariances}} + \underbrace{\gamma_N\iota}_{\text{market portfolio}}. \tag{16}$$

Intuitively, given a vector of prices formed with imperfect information beliefs, the optimal strategy is to form a portfolio that exploits these estimation errors arising from both the covariance and expected returns while also having a market portfolio component exploiting the positive risk premium.

1.5 Pricing Errors

Moreover, the weighted sum of squared pricing errors in the economy, relative to the market portfolio, is given by

$$\begin{aligned}\alpha'\Lambda_t^{-1}\alpha &= SR(w^*)^2 - SR(\frac{1}{N}\iota)^2 \\ &= (\varepsilon + \gamma_N\hat{\Lambda}_t\iota)'\Lambda_t^{-1}(\varepsilon + \gamma_N\hat{\Lambda}_t\iota) - \frac{1}{\sigma_m^2}((\iota'\varepsilon)^2 + 2\gamma_N(\iota'\varepsilon)\hat{\sigma}_m^2 + \frac{\gamma^2}{N^2}(\hat{\sigma}_m^2)^2).\end{aligned} \tag{17}$$

Equation 17 completely characterizes the pricing errors given the statistical properties of the dividends and agents' estimates about them. In what follows, we first consider specific cases when either the means or covariances are known, and then we use the statistical assumptions to get analytical results.

1.5.1 Known Means and Unknown Covariances

To get some intuition, consider the case of $\varepsilon = 0$, where agents know the true expected value of the fundamental. In that case, equation 17 specializes to:

$$\alpha'\Lambda_t^{-1}\alpha = \frac{\gamma^2}{N^2}\iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota - \frac{\gamma^2}{N^2}\frac{(\hat{\sigma}_m^2)^2}{\sigma_m^2}. \tag{18}$$

To illustrate the implications of this result, we provide a simple calibration. Figure 1 depicts the components of the weighted squared pricing errors for a simulated economy calibrated to target an equilibrium Sharpe ratio of 0.5 when there is perfect information. We assume that asset payoffs are driven by a multi-factor model with 5 factors. The first factor has a dominant influence, with a variance of 0.04, while the remaining factors each contribute a smaller total variance of 0.01. The loadings (Betas) for the first factor are centered around 1, so most assets have significant exposure to the primary market factor, while the other factors have Betas centered around 0. The total systematic variance from all factors is 0.05, and we adjust the idiosyncratic variance to 0.16

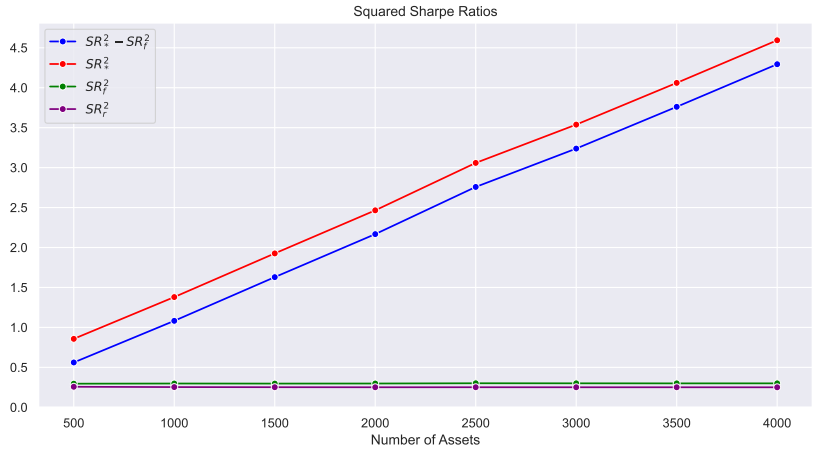
to achieve a target total variance of 0.25 for each stock.

Agents in the economy know the expected dividend payoffs and have to estimate the covariance Λ_t from sample realizations. SR_*^2, SR_f^2, SR_r^2 denote the squared Sharpe ratios implied by the optimal portfolio under the true data-generating process, the feasible optimal weights, and the weights in a rational expectation equilibrium. In figure 1a, agents regularize their estimation using one factor extracted from PCA, while in figure 1b, agents shrink their estimate towards the sample diagonal matrix.

Neither the feasible nor the rational equilibrium squared Sharpe ratios depend on N. However, the optimal portfolio under the true data-generating process Sharpe ratio is, on expectation, a linear function of N. Intuitively, as the number of assets grows, the estimation gets more challenging, and the number of mispriced assets increases. Hence, there are better optimal combinations. The squared pricing errors increase because the optimal portfolio under the true data-generating process Sharpe ratio increases but not the feasible one.

For the specific case of known means and unknown variances, we have the following theorem.

Figure 1: Squared Sharpe Ratios as a Function of the Number of Assets



(a) Agents Using Sample Covariance Matrix



(b) Agents Using Regularized Sample Covariance Matrix

The figure shows the Sharpe ratios of a simulated economy with different numbers of assets. The economy has a one-factor model structure in assets' payoffs with $\beta \sim N(1, 0.25^2)$, variance of aggregate factor = 0.2^2, idiosyncratic variance = 0.45^2, $\gamma = 2.5$, and $t = 252$. The blue line denotes the difference in squared Sharpe ratios between the optimal portfolio under the true data-generating process and feasible squared Sharpe ratios. The red line denotes the optimal portfolio under the true data-generating process squared Sharpe ratio. The green line denotes the feasible squared Sharpe ratio. The purple line denotes the perfect information squared Sharpe ratio.

Theorem 1. *If investors have perfect information about the mean but only the correct prior about the covariance, the average pricing errors relative to the feasible optimal portfolio increase in the number of assets and states.*

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] = \frac{\gamma^2}{N^2} \left(\frac{1}{t+1} \right)^2 t(N-1)(S+N)^2 \bar{\varrho} \sim O(\frac{N}{t} + \frac{S^2}{NT} + \frac{S}{t}) \tag{19}$$

Proof. See Appendix □

1.5.2 Unknown Means and Known Covariances

If covariances are known, then the squared pricing errors are given by

$$\alpha' \Lambda_t^{-1} \alpha = \varepsilon' \Lambda_t^{-1} \varepsilon + 2\gamma_N \iota' \Lambda_t \Lambda_t^{-1} \varepsilon + \frac{\gamma^2}{N^2} \iota' \Lambda_t \Lambda_t^{-1} \Lambda_t \iota - \frac{1}{\sigma_m^2} ((\iota' \varepsilon)^2 + 2\gamma (\iota' \varepsilon) \sigma_m^2 + \frac{\gamma^2}{N^2} (\sigma_m^2)^2) \tag{20}$$

$$= \varepsilon' \Lambda_t^{-1} \varepsilon - \frac{1}{\sigma_m^2} (\iota' \varepsilon)^2. \tag{21}$$

Similarly to the previous case, the pricing errors are generally increasing in the number of assets or states. To get some analytical insights, we'll focus on the case where investors' uncertainty about the mean is proportional to the true covariance matrix and the correct prior about the expected value of the state variables.

Theorem 2. *If investors have perfect information about covariances but only the correct prior about the expectation, the average pricing errors relative to the feasible optimal portfolio increase in the number of assets and states at a rate of* $\frac{N+S}{t}$.

$$\mathbb{E}[\alpha_t' \Lambda_t^{-1} \alpha_t] = \frac{t}{(t+\kappa_0)^2} (N+S) - \frac{t}{(t+\kappa_0)^2} = \frac{t}{(t+\kappa_0)^2} (N+S-1) \sim O(\frac{N+S}{t}) \tag{22}$$

Proof. See Appendix □

1.5.3 Unknown Means and Covariances

The last case of interest is considering unknown means and covariances. For tractability, we consider the case where the estimation error is approximately zero on average, $E[\iota'\varepsilon] \approx 0$, and a slightly stronger approximation imposing that $E[\iota' A\varepsilon] \approx 0$ if A is non-random and positive definite. In that case, the total pricing errors can be approximated by the Effectively, sum of both estimation errors about expected values and covariances.

$$E[\alpha'\Lambda_t^{-1}\alpha] \approx E[\varepsilon'\Lambda_t^{-1}\varepsilon] + E[\frac{\gamma^2}{N^2}\iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota] - E[\frac{1}{\sigma_m^2}(\frac{\gamma^2}{N^2}(\hat{\sigma}_m^2)^2)] \tag{23}$$

$$= \underbrace{E[\varepsilon'\Lambda_t^{-1}\varepsilon] - E[\frac{1}{\sigma_m^2}(\iota'\varepsilon)^2]}_{\alpha \text{ from unknown means}} + \underbrace{E[\frac{\gamma^2}{N^2}(\iota'\hat{\Lambda}_t\Lambda_t^{-1}\hat{\Lambda}_t\iota] - E[\frac{(\iota'\hat{\Lambda}_t\iota)^2}{\iota'\Lambda_t\iota})]}_{\alpha \text{ from unknown means covariances}}. \tag{24}$$

$$\tag{25}$$

1.6 Bayesian Learning About Means with Heterogeneous Beliefs

The homogeneous agent setting can be interpreted as a special case of our framework where only an atomistic (measure zero) set of agents might have correct beliefs about the data-generating process. This helps explain why mispricing increases with complexity but cannot tell us what happens when some investors with positive mass have correct beliefs. To understand how mispricing persists in equilibrium even when a significant fraction of investors knows the true data-generating process, we extend the model to include heterogeneous beliefs. Throughout this section, we maintain standard assumptions that agents agree to disagree, are price takers, and do not learn from prices. While the model could be extended to relax these assumptions, we leave such extensions for future research.

First, we consider investors who learn about the mean of X_t and know covariances. This setup makes the problem more intuitive, but it is straightforward to adapt it to unknown means and unknown covariances at the cost of more algebra. Investors use the t observations to learn about

X_t. We consider I different types of investors with different priors, with θ_i mass for each investor type i.

Each investor i has a prior belief about π:

$$\pi \sim N(\pi_i, V_i) \tag{26}$$

where $V_i = k_i V$, and $k_i > 0$ is a scalar representing investor i's confidence in their prior. A larger k_i indicates more uncertainty about the prior mean.

After observing t periods of data $\{X_t\}_{t=1}^{t}$, investor i updates their belief to:

$$\pi_{\text{post},i} = \alpha_i \pi_i + (1 - \alpha_i)\bar{X} \tag{27}$$

where

$$\alpha_i = \frac{1}{1 + k_i t}, \quad \bar{X} = \frac{1}{t}\sum_{\tau=1}^{t} X_\tau \tag{28}$$

and the posterior variance of the mean is:

$$V_{\text{post},i} = (1 - \alpha_i)V \tag{29}$$

In the Appendix, we show that the variance-covariance matrix perceived by investor i combines fundamental uncertainty with parameter uncertainty:

$$\Lambda_i = B_t V \left(\frac{1 + k_i(t+1)}{1 + k_i t} \right) B_t^\top = \Lambda_t \varphi_i \tag{30}$$

where $\Lambda_t = B_t V B_t^\top$ is the variance of dividends and $\varphi_i = \frac{1+k_i(t+1)}{1+k_i t} > 1$ adjusts for uncertainty about means. This additional uncertainty leads investors to take more conservative positions.

Given their beliefs, each investor's demand follows a mean-variance optimization:

$$x_{i,t} = \frac{1}{\gamma} \Lambda_t^{-1} \varphi_i^{-1} \left(B_t \pi_{\text{post},i} - (1 + r_f) p_t \right). \tag{31}$$

This demand combines the standard mean-variance portfolio Λ_t^{-1} (expected excess returns) with an adjustment factor φ_i^{-1} that reduces positions based on parameter uncertainty.

Market clearing requires that the aggregate demand equals the per capita supply of risky assets:

$$\frac{\iota}{N} = \sum_{i=1}^{I} \theta_i x_{i,t}. \tag{32}$$

Solving for equilibrium prices yields:

$$p_t = \frac{1}{1+r_f}\left(\hat{a}_E - \frac{\gamma}{NW}\Lambda_t \iota\right), \tag{33}$$

where $\hat{a}_E = B_t\left(\frac{1}{W}\sum i = 1^I w_i \pi_{\text{post},i}\right)$ represents the economy's weighted average expectation. The weights $w_i = \theta_i \varphi_i^{-1}$ and total weight $W = \sum_{i=1}^{I} w_i$ $(0 < W < 1)$ reflect both the mass of each investor type and their uncertainty-adjusted demand.

The resulting vector of true expected returns mirrors the homogeneous case but with aggregate beliefs:

$$\mu_t \equiv \mathbb{E}_t[d_{t+1}] - (1+r_f)p_t = a_t - \hat{a}_E + \frac{\gamma}{NW}\Lambda_t \iota. \tag{34}$$

The expected returns compensate for both fundamental risk ($\frac{\gamma}{NW}\Lambda_t \iota$) and loads on the estimation errors ($a_t - \hat{a}_E$). Each investor type perceives different expected returns based on their beliefs:

$$\hat{\mu}_{i,t} \equiv \mathbb{E}_{i,t}[d_{t+1}] - (1+r_f)p_t = a_i - \hat{a}_E + \frac{\gamma}{NW}\Lambda_t \iota. \tag{35}$$

This heterogeneity in beliefs leads each agent to perceive mispricing relative to their own valuation. Defining $\varepsilon_i = a_i - \hat{a}_E$ as the perceived mispricing, each investor's optimal portfolio combines a speculative component trading on perceived mispricing with a market component:

$$x_i = \frac{1}{\gamma \varphi_i}\Lambda_t^{-1}\varepsilon_i + \frac{1}{N\varphi_i W}\iota. \tag{36}$$

The pricing error properties from the homogeneous case persist in a straightforward way through the weighted average expectation $\hat{a}_E$. Since $\hat{a}_E$ is a weighted average of each investor type's beliefs, and each investor faces growing estimation challenges as complexity increases (higher N or S), the estimation errors of $\hat{a}_E$ itself grow with market complexity. In other words, the market's aggregate expectation inherits the same fundamental relationship between complexity and estimation error that we found in the homogeneous case. Even informed investors do not eliminate these pricing errors due to risk aversion limiting the size of their positions.

1.6.1 Two Types of Investors

To better understand how mispricing persists in equilibrium even with informed investors, consider a special case with two types of investors. Type 1 investors are fully informed about the true mean μ and have complete confidence in their knowledge ($k_1 = 0$). Type 2 investors are less informed, with a prior belief $\mu_2 \neq \mu$ and confidence level $k_2 > 0$.

The posterior beliefs for Type 1 (informed) investors are straightforward: $\pi_{\text{post},1} = \mu$, $V_{\text{post},1} = 0$, and $\varphi_1 = 1$.

For Type 2 (uninformed) investors, the posterior beliefs reflect both their prior and the data:

$$\pi_{\text{post},2} = \alpha_2 \mu_2 + (1 - \alpha_2)\bar{X} \tag{37}$$

$$V_{\text{post},2} = \frac{k_2 V}{1 + k_2 T} \tag{38}$$

$$\varphi_2 = \frac{1 + k_2(T+1)}{1 + k_2 T} > 1 \tag{39}$$

where $\alpha_2 = \frac{1}{1+k_2 T}$.

The weighted expectations of the economy combine these beliefs according to:

$$\hat{a}_E = B_t \left(\frac{\theta_1 \mu + \theta_2 \varphi_2^{-1} (\alpha_2 \mu_2 + (1 - \alpha_2)\bar{X})}{\theta_1 + \theta_2 \varphi_2^{-1}} \right) \tag{40}$$

where $w_1 = \theta_1$ is the weight of informed investors, $w_2 = \theta_2 \varphi_2^{-1}$ is the (effective) weight of unformed

investors, and $W = w_1 + w_2$ is the total weight.

The estimation error in this two-type model reveals why mispricing persists:

$$\varepsilon = a - \hat{a}_E \tag{41}$$

$$= B_t\mu - B_t\left(\frac{1}{W}(w_1\mu + w_2\pi_{\text{post},2})\right) \tag{42}$$

$$= B_t\left(\frac{w_2}{W}\left[\alpha_2(\mu - \mu_2) + (1-\alpha_2)(\mu - \bar{X})\right]\right) \tag{43}$$

This decomposition shows that mispricing persists even with informed investors because of two key factors. First, risk aversion limits the size of positions that informed investors can take, preventing them from completely eliminating mispricing. Second, the presence of less informed investors with different beliefs creates a weighted average expectation that differs from the true mean, leading to systematic deviations in prices. The magnitude of these deviations depends on both the relative mass of informed versus uninformed investors (θ_1 vs θ_2) and the confidence level of the uninformed investors (k_2). As before, the estimation errors increases with complexity.

1.7 Sophisticated Investors with Partial Information

The previous sections demonstrated how market complexity impacts pricing errors and the limitations of parsimonious factor models. We now extend this analysis by examining how sophisticated investors with partial information can generate profitable trading strategies that maintain low correlations with each other, even as market complexity increases. This finding helps explain the empirical observation that different quantitative strategies can simultaneously achieve high Sharpe ratios while remaining largely uncorrelated.

Consider sophisticated investors who have perfect information about a subset of the available assets. We model this partial information structure by assuming each investor i has access to a random subset of assets, represented by a diagonal selection matrix S_i where each diagonal element equals 1 with probability q (indicating the asset is known/accessible to the investor) and 0 otherwise. This captures the realistic constraint that even sophisticated investors cannot perfectly analyze and

trade all available securities.

We consider the case where atomistic informed investors trade in an otherwise homogeneous agents economy as in the first section. For tractability, all agents have full information about the covariance matrix Λ_t, and we assume the homogeneous agents have the correct prior about expected payoffs.

For tractability, we assume the payoffs' covariance matrix follows a one-factor structure:

$$\Lambda = \beta\beta'\sigma_m^2 + \sigma_\epsilon^2 I \tag{44}$$

where β captures aggregate factor exposures, σ_m^2 is the aggregate variance, and σ_ϵ^2 represents idiosyncratic risk. While this simplified structure might appear restrictive, it captures the essential feature needed for our analysis: the tension between diversifiable idiosyncratic risk and undiversifiable systematic risk. The key insights about correlation scaling emerge from the interplay between estimation errors and risk structure, not the specific number of factors.

Given their partial information set, investors form portfolios with weights:

$$w_i = \frac{1}{N}\Lambda^{-1}S_i(a - \hat{a}_E) \tag{45}$$

where Λ is the return covariance matrix, a represents true expected payoffs, and $\hat{a}_E$ captures the market's aggregate estimate of expected payoffs. The estimation error $(a - \hat{a}_E)$ follows a normal distribution with mean zero and covariance matrix Λ_a that reflects both fundamental uncertainty and parameter estimation error:

$$(a - \hat{a}_E) \sim N(0, \Lambda_a) \tag{46}$$

where

$$\Lambda_a = \left(\frac{1}{\kappa} + t\right)^{-1}\left(\beta\beta'\sigma_m^2 + \sigma_\epsilon^2 I\right) \tag{47}$$

The expected return of this strategy scales with the fraction of assets known as $O(q)$.

$$\mathrm{E}(R_w) = \frac{1}{N}(\Sigma^{-1} S_i(a - \hat{a}_E))'(a - \hat{a}_E) + \gamma_N \Sigma\beta) \sim O(q) \tag{48}$$

The quadratic form $(a - \hat{a}_e)'\Sigma^{-1} S_i(a - \hat{a}_e)/N$, which corresponds to the expected return component arising from estimation errors, has expected value q because S_i selects q fraction of a chi-squared random variable and the $1/N$ scaling in portfolio weights gives us final expectation q. The component corresponding to the risk premium, $a(a - \hat{a}_e)' S_i \beta$, has zero expected value because the portfolio is zero-cost and estimation errors are centered at zero. Economically, this means strategies' expected returns depend only on the fraction of assets an investor understands deeply (q), not market size (N), as any benefits from more trading opportunities are exactly offset by the need to reduce position sizes to manage estimation risk.

Using the Sherman-Morrison formula, we can decompose the inverse of the assets variance as:

$$\Lambda^{-1} = \frac{1}{\sigma_\epsilon^2} I - \frac{\sigma_m^2}{\sigma_\epsilon^2(\sigma_\epsilon^2 + \sigma_m^2 \beta'\beta)} \beta\beta' \tag{49}$$

The variance of each investor's portfolio contains two components. The first comes from idiosyncratic risk and scales as $O(q/N)$, declining with the number of assets but increasing with the fraction of assets known. The second component reflects systematic market risk and scales as $O(q^2/N)$. Intuitively, as investors have access to more assets (higher q), they can better diversify idiosyncratic risk but remain exposed to systematic risk that cannot be diversified away. Thus, the total variance of the portfolio scales as:

$$\mathrm{Var}(R_w) = \frac{1}{N^2}(a - \hat{a}_E)' S_i \Lambda^{-1} S_i (a - \hat{a}_E) \sim O\left(\frac{q}{N}\right).$$

The Sharpe ratio of the strategy scales as $O(q\sqrt{N})$ because expected returns scale with q while the standard deviation scales as $O(\sqrt{q/N})$. This scaling has a powerful economic interpretation: as markets become more complex (higher N), investors who understand a fixed fraction of assets

(constant q) can achieve higher risk-adjusted returns by taking smaller positions in more assets, benefiting from improved diversification without sacrificing the profitability that comes from their understanding of individual securities.

The covariance between two strategies scales as $O(q^2/N)$ and each strategy's variance scales as $O(q/N)$, leading to correlations that scale as $O(q)$ independent of market size. The economic intuition is compelling: while both covariance and variance benefit from diversification (the $1/N$ terms), the correlation depends only on the overlap in known assets (q). This explains why even sophisticated investors pursuing similar strategies can maintain low correlations in complex markets - as long as they focus on different subsets of assets, the benefits of a larger investment universe (N) reduce both risk and comovement proportionally, leaving the correlation dependent only on the fraction of shared expertise (q).

$$\text{Cov}(R_{w_i}, R_{w_j}) = \frac{1}{N^2}(a - \hat{a}_E)' S_i \Lambda^{-1} S_j (a - \hat{a}_E) \sim O\left(\frac{q^2}{N}\right) \tag{50}$$

Hence, the correlation between different investors' portfolios scales as $O(q)$:

$$\text{Corr}(R_{w_i}, R_{w_j}) = \frac{\text{Cov}(R_{w_i}, R_{w_j})}{\sqrt{\text{Var}(R_{w_i})\text{Var}(R_{w_j})}} \sim O(q) \tag{51}$$

This means that as markets become more complex (larger N), investors who each analyze different subsets of assets will naturally form portfolios with low correlations, even when pursuing similar statistical arbitrage strategies. In high-dimensional markets, there are many different ways to construct portfolios with attractive risk-return tradeoffs by focusing on different subsets of assets.

These theoretical predictions align with our empirical findings about the low correlations between systematic hedge fund strategies and the challenges faced by parsimonious factor models. They 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 q) or spreading resources more thinly (declining q). Our model predicts that successful funds will maintain stable q by scaling

research capabilities with N, explaining why large quantitative funds invest heavily in research infrastructure. Funds that fail to scale research appropriately should see declining Sharpe ratios relative to competitors.

The model also makes specific predictions about cross-fund correlation patterns. Correlations between funds should depend primarily on research overlap (shared q) rather than strategy similarity. Funds using similar techniques but focusing on different market segments should maintain low correlations. This explains why even "copycat" strategies can succeed without eroding the profits of the original strategy. Furthermore, increasing market complexity should lead to more, not fewer, profitable trading strategies. However, these strategies should become increasingly specialized and research-intensive. The "factor zoo" should continue to expand with market complexity, but new factors should show less overlap with existing ones.

These predictions help rationalize why we observe a proliferation of seemingly unrelated anomalies and suggest that this pattern is a natural outcome of rational investors optimally responding to the opportunities and constraints created by market complexity. Rather than indicating market inefficiency, the existence of numerous successful but uncorrelated strategies reflects the inherent challenge of spanning high-dimensional markets with limited information processing capacity.

2 Empirical Study

We first look for empirical backing for the assumption that we are in a high-dimensional setting. We provide evidence of the high dimensionality of stock returns. Then, we look for empirical support for the model's predictions

2.1 Data

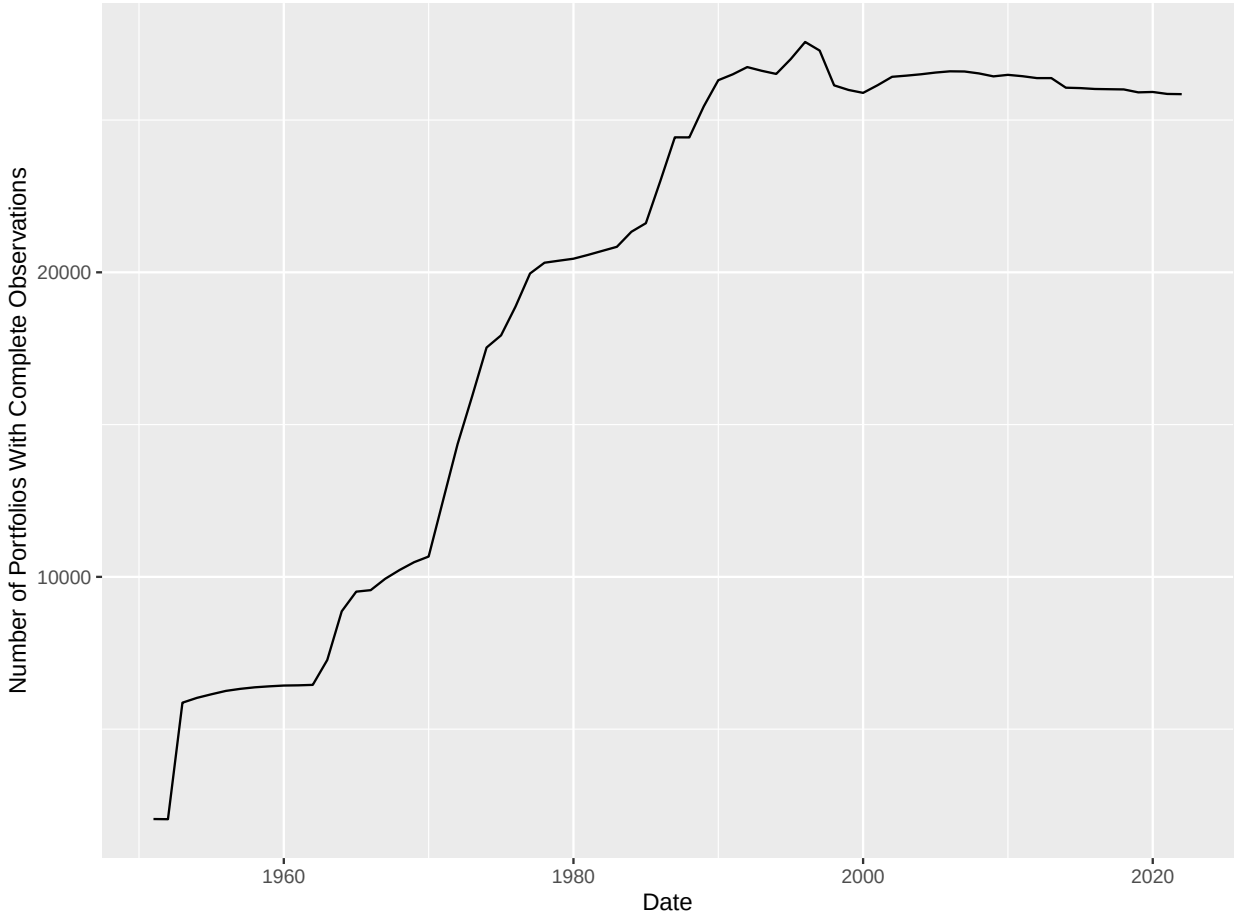
We use the data from Chen, Lopez-Lira, and Zimmermann (2023). They generate 29,315 firm-level signals as follows. Let X be one of 242 Compustat accounting variables + CRSP market equity and Y be one of the 65 variables observed and positive for $> 25\%$ of firms in 1963 with

matched CRSP data. They form signals by combining all combinations of ratios (X/Y) and scaled first differences $(\Delta X/\text{lag}(Y))$. This procedure would lead to $242 \times 65 \times 2 = 31,460$ signals, but there are 2,145 signals redundant in the "unsigned" portfolio sorts. They lag each signal by six months and then form long-short decile strategies by sorting stocks on the lagged signals each June. Delisting returns and other data handling methods follow Chen and Zimmermann (2022) to ensure that the published and data-mined strategies are comparable.

We consider the in-sample period to be $[1980, 2000]$ and the out-of-sample period to be $[2001, 2021]$. We restrict our analysis to the 17517 long-short portfolios that, due to data availability, have complete observations for the sample of 1980-2021. We refer to these signals simply as the 'portfolios.' There are no firm survivorship bias issues since the delisting returns of individual firms are handled appropriately. Rather, the data for some firm characteristics was unavailable until later in the sample.

These portfolios map nicely into our theoretical framework since each is constructed using one signal at a time, and hence, they proxy for the different characteristics that map to states. Moreover, being diversified portfolios implies little if any, idiosyncratic variation left, and we can focus on the systematic components.

Figure 2: Number of Portfolios with Complete Observations by Year



2.2 High-Dimensionality

We first motivate the high-dimensional nature of stock returns by exploring these portfolios' cross-sectional and time-series variations. First, we focus on the time-series variation. Even though these portfolios are very diversified, they retain a high-dimensional structure as depicted in Figure 3. About 20% of the overall variance of one-year daily return variance is explained by the first ten greatest principal components (in-sample). To account for 50% of the variation, we need about 50 factors. We need 75 factors to explain roughly 75% of the variation. The prior result is still obtained even though PCA is known to yield the (in-sample) solution for which the (time-series) variance explained is the highest for a given number of factors.

Then, we explore the cross-sectional variation and show that consistent with the factor zoo literature, there are too many signals with out-of-sample predictability. First, we collect the top 100 portfolios with the highest t-stats (we flip the portfolio returns if the t-stat is negative). Then, we look at the out-of-sample distribution of t-statistics. We find that there are way too many signals to be due to chance. Figure 4 shows the distribution of out-of-sample t-statistics and a fitted normal distribution. Panel 4a looks at the long-short returns while Panel 4b looks at the CAPM alphas. In both cases, we find many signals retain the out-of-sample predictability. The result holds despite the well-known fact that most of the predictability coming from characteristics declines very fast after the year 2000 (Lopez-Lira and Roussanov (2022)).

Figure 3: Cumulative Percentage of Variance Explained by the Principal Components

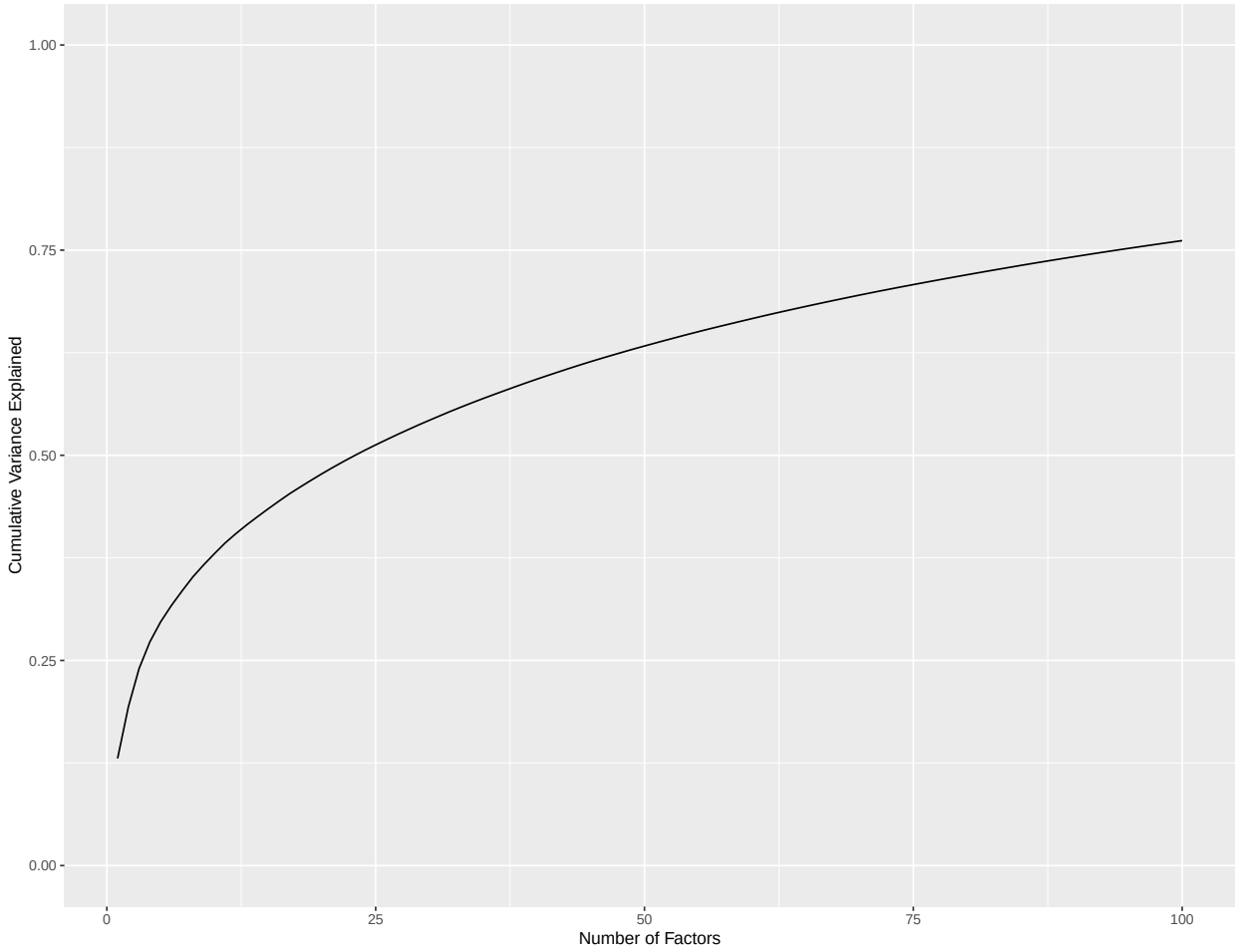
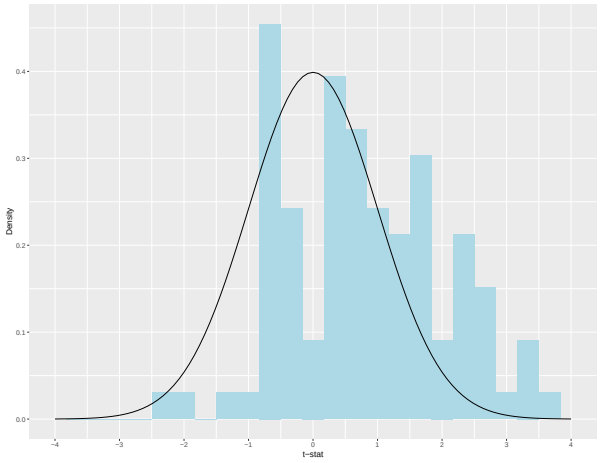
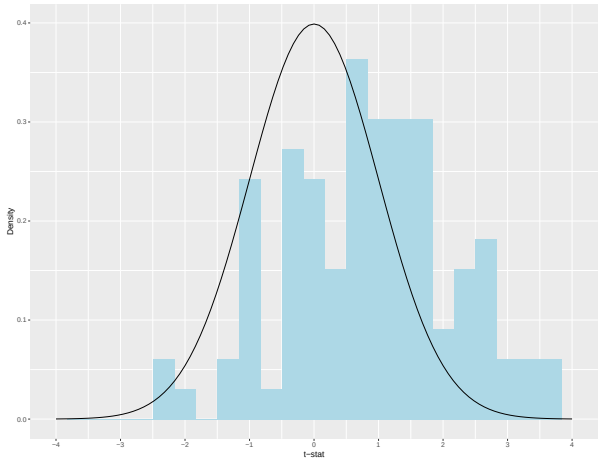


Figure 4: Comparative distributions of Out-of-Sample t-statistics



(a) Distribution of Out-of-Sample t-statistics from the top-100 In-Sample Portfolios



(b) Distribution of Out-of-Sample CAPM Alpha t-statistics from the top-100 In-Sample Portfolios

2.3 Low Correlations Between Strategies

An important implication from the model is that in a high-dimensional world, portfolios with high Sharpe ratios can have low correlations. We explore if this is the case empirically. We first consider portfolio correlations of investment strategies taken from the literature.

It is well known that a factor model prices the cross-section if and only if the factor model can be collapsed down to a mean-variance efficient portfolio (Huberman and Kandel 1987). Table 1 shows the variety of models from the literature that are used here, and Davis 2024 has more details about the implementation of these models. The Freyberger, Neuhierl, and Weber 2020b dataset of 62 asset characteristics is used. We follow Kozak, Nagel, and Santosh 2020 and Kelly, Pruitt, and Su 2019 by converting characteristics to cross-sectional percentiles (between 0 and 1) and then subtracting 0.5 to create long-short characteristic-managed portfolios. The data is monthly. The data from July 1962 up until the end of 1989 is used to train the model, and the parameters of each model are refitted every year. The results come from January 1990 to December of 2023, which are trained using strictly ex-ante data.

Panel (a) of Figure 6 shows the distribution of pair-wise correlations across all strategies. The average pair-wise correlation is about 0.26 which corresponds to an R^2 of a little bit more than 5%.

While these results already show that different methods to produce expected return estimates result in very low correlation, we further consider how is the correlation of actual strategies that are traded in the market using 13F holdings data. We follow the KY methodology to correct investors types in the 13F data, and define a systematic hedge fund as an "investment advisor" that holds at least 500 stocks. There are many hedge funds that hold only 10 or 20 stocks, instead of investing in systematic strategies, and these do not correspond to the sophisticated and diversified strategies we consider. With this filter, we capture some well-known hedge funds that use quantitative models to pursue systematic equities trading strategies, including AQR, Renaissance, Tower Research, Citadel, Two Sigma, S.A.C Capital, D.E. Shaw, Roll and Ross Asset Management, Paradigm Asset Management, Quantitative Systematic Strategies, Bridgewater Associates, Cubist Systematic Strategies, Martingale Asset Management, and Quantitative Systematic Strategies.

We consider only the active part of the portfolio weights, meaning we take the portfolio weights of the systematic hedge funds, and subtract off the fund's market beta times the market portfolio weights. In particular, if the weights of fund i are given by w_i, and its beta given by β_i, we define the market-neutral weights as

$$w_{i|M} = w_i - \beta_i w_M. \tag{52}$$

This result in the correlation of the active part of funds i and j given by:

$$\text{corr}_{i,j} = \frac{(w_{i|M})'\Sigma(w_{j|M})}{\sqrt{(w_{i|M})'\Sigma(w_{i|M})}\sqrt{(w_{j|M})'\Sigma(w_{j|M})}} \tag{53}$$

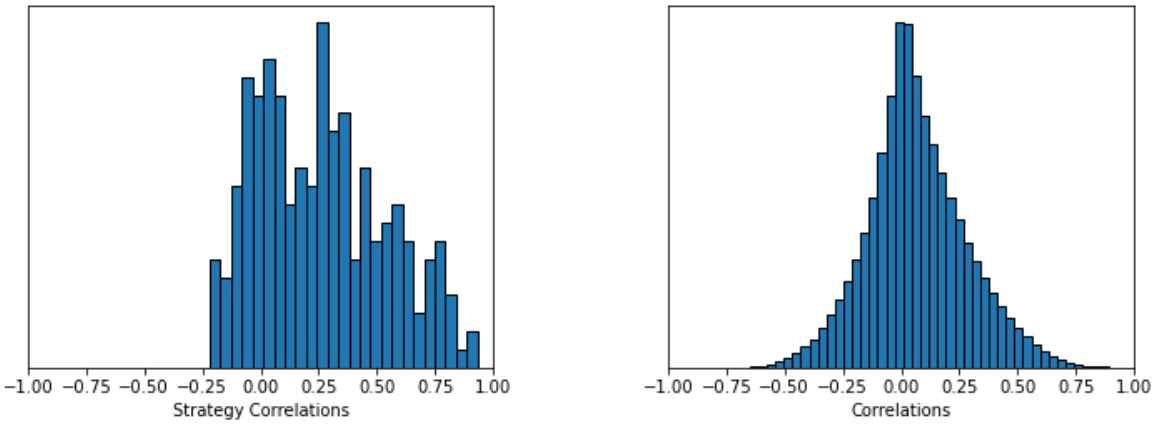
Panel (b) of Figure 6 shows the pair-wise correlations of these strategies. The average pair-wise correlation is about 0.07. This is in many ways similar to Koijen and Yogo 2019, who show that much of demand is not explained by classic characteristics. We go one step further by showing that much of the active component of hedge fund portfolios is not even explained by even other hedge fund's active portfolios.

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 2020b
KNS	Kozak, Nagel, and Santosh 2020
BSV	Brandt and Santa-Clara 2006
DGU	DeMiguel, Garlappi, and Uppal 2007

Table 1: **Statistical Arbitrageur Demand Models** This table lists the initialisms and sources for the investment strategy models.

Figure 5: Correlations of Portfolios



(a) Trading Strategy Correlations

(b) Systematic Hedge Fund Cross Correlations

Figure 6: This figure shows the correlations across trading strategies in Panel (a), and correlations of the active component of Systematic Hedge Funds from the 13F data in Panel (b).

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: **Weber CAPM Regressions.** This table presents the CAPM regression results for the Weber dataset, divided into two panels. Panel A contains results from the first part of the data, and Panel B contains results from the second part. The table includes the relevant coefficients and standard errors, calculated over the specified period. Detailed model descriptions are available in Table 1.

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

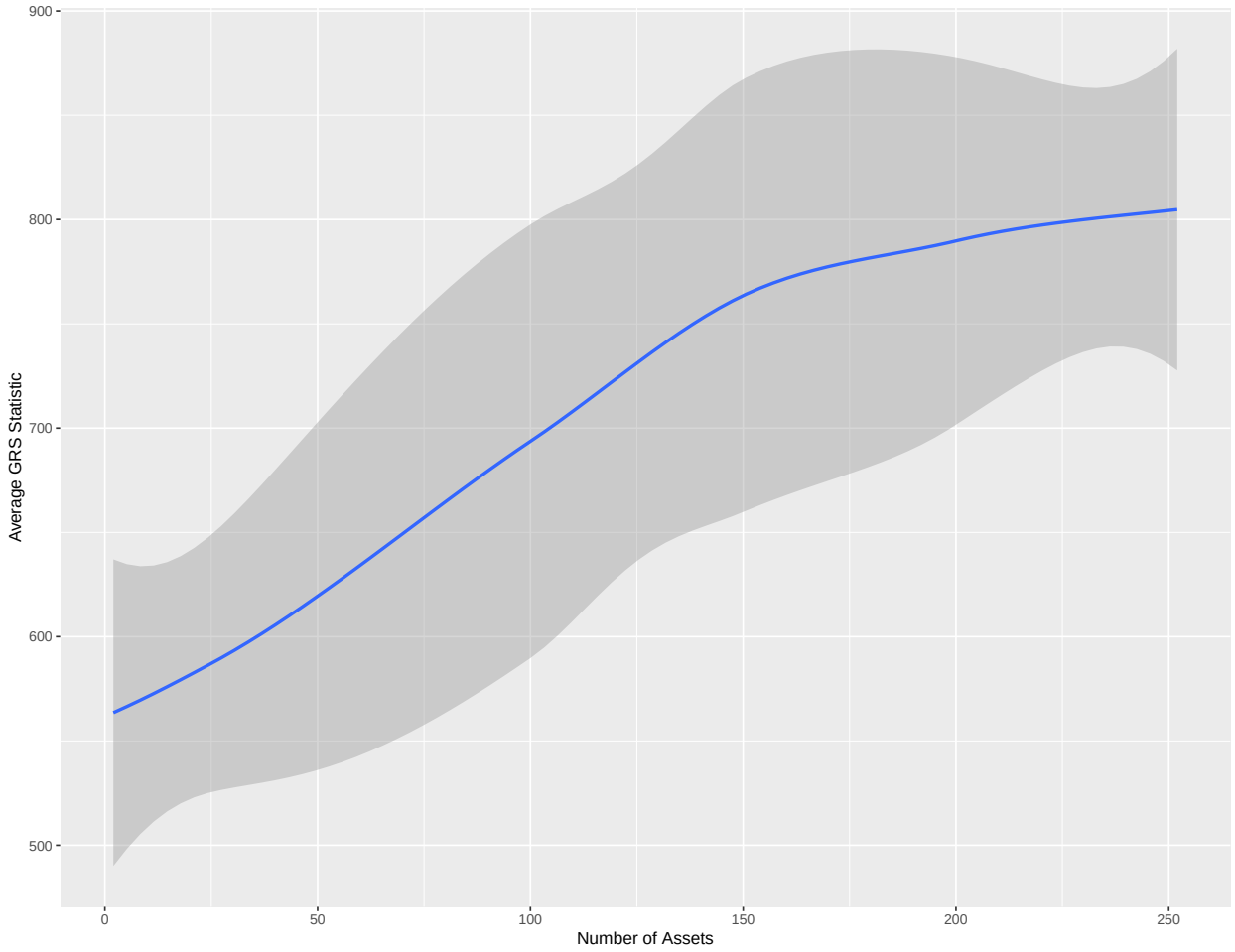
Panel B: Weber Cross Regressions Beta Part 2

	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 3: **Weber Unconditional Cross Regressions (Betas).** Panel B presents the regression results for the Weber dataset, focusing on beta coefficients from the second part of the data.

Figure 7: Average GRS Statistic Between the Best Portfolios and Number of Assets



3 Conclusion

Our framework resolves several enduring asset pricing mysteries by reframing them as natural outcomes of market complexity. In high-dimensional environments, where the parameter space outstrips available data, rational optimization under uncertainty generates systematic pricing errors, dilutes the explanatory power of risk premiums, and generates diverse anomalies. Far from being robust universal solutions, traditional factor models like the CAPM and its multi-factor successors falter—not due to market inefficiencies but because their parsimony cannot keep pace with the estimation challenges posed by modern markets.

This perspective recasts the proliferation of anomalies and the breakdown of parsimonious models as evidence of rational adaptation rather than market failure. Sophisticated investors can exploit distinct estimation errors to construct high-Sharpe-ratio strategies that remain little correlated. Our theory reveals this divergence as a feature of high-dimensional markets, where abundant error dimensions allow multiple strategies to thrive without overlapping.

Empirical evidence reinforces these insights. Analyzing thousands of characteristic-sorted portfolios, we uncover a stark high-dimensional reality: the first ten principal components explain just 20% of the return variance of diversified portfolios, and even 75 factors leave 25% unaccounted for. Similarly, the active positions of systematic hedge funds show an average R^2 below 1%, highlighting their ability to target orthogonal facets of market complexity.

References

Alti, Aydogan, and Sheridan Titman. 2019. "A Dynamic Model of Characteristic-Based Return Predictability." *Journal of Finance* 74, no. 6 (December): 3187–3216. ISSN: 15406261. 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): 2703–2762. ISSN: 0034-6527. https://doi.org/10.1093/RESTUD/RDAD013.

Baba Yara, Fahiz, Brian H. Boyer, and Carter Davis. 2021. "Messy Asset Pricing: Can AI Models Lead to a Consensus?" *SSRN Electronic Journal* (November). https://doi.org/10.2139/SSRN.3967588.

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, Markus Pelger, and Jason Zhu. 2020a. "Forest Through the Trees: Building Cross-Sections of Stock Returns." *working paper* (September). 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.

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

Chen, Andrew Y., Alejandro Lopez-Lira, and Tom Zimmermann. 2023. "Peer-Reviewed Theory Does Not Help Predict the Cross-section of Stock Returns." *SSRN Electronic Journal* (March). https://doi.org/10.2139/SSRN.4308069.

Chen, Andrew Y., and Tom Zimmermann. 2022. "Open Source Cross-Sectional Asset Pricing." *Critical Finance Review* 11, no. 2 (May): 207–264. ISSN: 2164-5744. https://doi.org/10.1561/104.00000112.

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

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

Daniel, Kent, and Sheridan Titman. 1997. "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns." *The Journal of Finance* 52 (1): 1–33. https://doi.org/https://doi.org/10.1111/j.1540-6261.1997.tb03806.x.

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

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). 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. 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. 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.

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). 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). 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). 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): 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): 483–510. ISSN: 1540-6261. https://doi.org/10.1111/J.1540-6261.1987.TB04565.X.

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): 1382–1423. ISSN: 0893-9454. https://doi.org/10.1093/RFS/HHX001.