# Schrödinger's Sparsity in the Cross Section of Stock Returns *

Doron Avramov Guanhao Feng Jingyu He Shuhua Xiao

First draft: Jan. 2025; this draft: Dec. 2025

**Abstract**

This paper develops a hierarchical Bayesian IPCA framework in which the degree of characteristic-driven sparsity is learned from the data rather than imposed *ex ante*. The model delivers state-dependent interior sparsity: the characteristic mapping for factor loadings is neither strictly sparse nor fully dense, and posterior sparse probabilities vary across assets and over the business cycle. Allowing for characteristic-driven mispricing yields a clear division of labor, as mispricing coefficients are markedly sparser than factor-loading coefficients, stabilizing beta estimates and sharpening the decomposition of expected returns. Relative to dense and fixed-sparsity benchmarks, the approach improves out-of-sample forecasting and portfolio performance.

**Keywords:** Illusion of Sparsity, Bayesian IPCA, Model Uncertainty, Conditional Factor Model, Return Decomposition

**JEL Classification:** C12, C52, G11, G12

*We thank Ben Charoenwong, Bong-Geun Choi, Liyuan Cui, Phillip Gharghori (discussant), Ai He (discussant), Bryan Kelly, Yuan Liao, Semyon Malamud, Stefan Nagel, Ji Yeol Jimmy Oh (discussant), Markus Pelger, Nikolai Roussanov, Gustavo Schwenkler, Stefan Voigt, Dacheng Xiu, Guofu Zhou, the seminar and conference participants at CityUHK and Central University of Finance and Economics, 2025 Global AI Finance Research Conference, 2025 Melbourne Asset Pricing Meeting, 2025 FinEML Conference, 2025 HKUST IAS-SBM Joint Workshop Financial Econometrics in the Big Data Era, 2025 SoFiE Financial Machine Learning Summer School at Yale, 2025 INFORMS International Meeting, 2025 China International Conference in Finance, 2025 Ninth PKU-NUS Annual International Conference on Quantitative Finance and Economics, and 2024 First Macau International Conference on Business Intelligence and Analytics for their valuable comments. Avramov (E-mail: doron.avramov@runi.ac.il) is at Reichman University; Feng (E-mail: gavin.feng@cityu.edu.hk), He (E-mail: jingyuhe@cityu.edu.hk) and Xiao (E-mail: shxiao3-c@my.cityu.edu.hk) are at the City University of Hong Kong.

# 1 Introduction

Modern cross-sectional asset pricing confronts a high-dimensional aggregation problem: expected returns and factor exposures are forecast from a large number of firm characteristics, even though the underlying economic structure may be low-dimensional. The resulting "factor zoo" (e.g., Harvey et al., 2016; Green et al., 2017) has motivated the use of sparse methods to distill a small set of meaningful drivers from a large pool of candidates. Empirically, sparse modeling has proven effective in explaining average returns and forecasting risk premia (e.g., Gu et al., 2020; Feng et al., 2020; Freyberger et al., 2020).

Nevertheless, the universality of the sparsity assumption has recently been called into question. Giannone et al. (2021) document that Bayesian posteriors rarely concentrate on sparse models in various economic datasets, referring to this phenomenon as the *Illusion of Sparsity*. Kozak et al. (2020) argue that the stochastic discount factor (SDF) may load on a broader set of firm characteristics, and Martin and Nagel (2022) provide a theoretical foundation for this view. He et al. (2024) propose a formal statistical test of the sparsity of observed factors and find limited support.

At the same time, recent advances in high-dimensional modeling, encompassing nonlinear prediction rules, latent factor representations, and deep learning, show that richer formulations can deliver strong empirical performance (e.g., Kelly et al., 2024; Didisheim et al., 2024).[1] Consistent with this view, Shen and Xiu (2024) show that ridge regression outperforms the Lasso in weak-signal environments, which are common for return prediction, highlighting the advantages of denser methods.

This distinction reflects modeling differences beyond the binary choice of sparse versus dense representations. Traditional approaches typically impose *ex ante* commitments to sparsity (via $L_1$ penalties) or density (via $L_2$ shrinkage). For example, implementations of Instrumented Principal Component Analysis (IPCA) often adopt either a dense characteristics structure (Kelly et al., 2019) or a sparse selection specifi-

[1]Other PCA latent factor models include Lettau and Pelger (2020) and Huang et al. (2022), while deep learning approaches include Chen et al. (2024), Feng et al. (2024), and Fan et al. (2025).

cation (Bybee et al., 2023). Each approach has clear strengths, but each also encodes an *ex ante* structural commitment about the characteristic-to-loading map.

In characteristic-based asset pricing, sparsity is a statement about coefficients in an *observed* characteristics basis. When observed characteristics are noisy proxies for latent forces, the effective sparsity suggested by the data can vary across assets and economic conditions, and it can be difficult to justify a single fixed sparsity level *ex ante*. This perspective motivates treating sparsity as an inferential object – allowing the evidence to determine when representations are effectively sparse versus diffuse – rather than adopting sparsity or density as a maintained restriction.[2]

We formalize this idea through *Schrödinger's Sparsity*, a probabilistic framework that treats sparsity as a latent feature of the data rather than a fixed modeling choice. The framework assigns each characteristic a posterior inclusion probability measuring the evidence that it contributes to pricing through the characteristic-to-loading map (and, in our full specification, separately through mispricing). Aggregating these inclusion probabilities yields an interpretable, state-dependent summary of effective sparsity, helping navigate the trade-off between parsimony and explanatory power without committing to fixed thresholds.

We operationalize Schrödinger's Sparsity in a Bayesian implementation of IPCA that combines a latent factor representation with a hierarchical spike-and-slab prior on the characteristic-to-loading map (and, in our full specification, on the characteristic-to-mispricing map). The approach maintains posterior uncertainty about characteristic contributions and updates beliefs as returns are observed, so sparsity is learned from the data via posterior inclusion and exclusion probabilities rather than imposed *ex ante*.

Three key features distinguish our approach. First, the model includes an explicit sparsity probability parameter: the prior probability that a characteristic is excluded, which can be learned from the data or specified exogenously. Second, sparsity is mod-

[2]Section 2.1 and Appendix III provide a structural rationale for this agnostic stance: the latent dimensionality of the economy does not, in general, discipline sparsity in the observed characteristics basis.

eled separately in the alpha and beta channels to reflect their distinct economic roles. Third, the framework accommodates both latent and observable factor structures, including the CAPM and Fama-French benchmarks, as well as exogenous latent factor specifications, within a unified conditional estimation framework.

Our empirical analysis delivers four central findings that establish the practical relevance of this framework.

First, we find empirical support for an interior sparsity level in factor loadings: a broad set of characteristics contributes meaningfully to factor exposures, rejecting both extremely sparse and fully dense specifications. At the same time, the data do not fully support dense specifications. Endogenous sparsity models outperform both dense and fixed-sparsity benchmarks. Fixed sparsity performs well only when the imposed level coincides with the one learned from the data.

These sparsity patterns are not an artefact of latent factor construction. Replacing latent factors with pre-specified observable factors, such as CAPM or the Fama-French five factors, preserves the tendency toward intermediate sparsity, although the exact levels depend on the factor specification. This robustness confirms that Schrödinger's sparsity reflects a general feature of the pricing problem rather than a mechanical byproduct of how factors are extracted.

Second, sparsity varies systematically across test assets. Portfolios with small CAPM or FF5 alphas are priced with relatively few characteristics and therefore exhibit high sparsity. Portfolios with larger alphas or higher Sharpe ratios, by contrast, tend to rely on a broader set of predictors and display lower sparsity.

Third, sparsity is time-varying and state-dependent. During recessions, posterior inclusion probabilities concentrate on a narrow set of characteristics, indicating that macroeconomic stress compresses the effective dimensionality of relevant predictors. In normal periods, the model selects more characteristics, demonstrating that model dimensionality responds endogenously to changing economic conditions.

Finally, allowing for characteristic-driven mispricing clarifies the division of labor in the return decomposition: factor loadings capture common variations, while a

smaller subset of characteristics loads into residual mispricing. Empirically, the mispricing coefficients are markedly sparser than the factor-loading coefficients, indicating that the mispricing channel operates through a more selective set of characteristics than the systematic-risk channel. Crucially, this framework allows the data to determine how concentrated each channel is, and therefore whether the cross section is best viewed as primarily exposure-driven, primarily alpha-driven, or a combination of both. This pattern directly addresses the challenge of distinguishing alpha from beta under model uncertainty.

When mispricing is ruled out, the time-varying loadings are forced to absorb residual return variation that the factor structure cannot explain, producing unstable and sometimes distorted beta estimates. This separation sharpens the alpha signal, improves out-of-sample performance, and yields a return decomposition that aligns more closely with the patterns in the data.

This paper contributes to the empirical asset pricing literature by offering a principled probabilistic resolution to the "Illusion of Sparsity" debate through a structural decomposition of cross-sectional returns. By demonstrating that the optimal model dimensionality occupies a state-dependent "interior region" that shifts with economic regimes, we reconcile the observed density of the SDF with the practical need for predictive parsimony.

Furthermore, our framework refines the estimation of latent risk factors and their time-varying exposures, highlighting the critical role of a selective alpha channel in preserving a stable and interpretable systematic risk structure. These results speak directly to recent discussions regarding the interpretability and limits of flexible specifications (e.g., Kelly et al., 2024) while contributing to the growing literature on sparse modeling (e.g., Chinco et al., 2019; Gu et al., 2020; Feng et al., 2020; Freyberger et al., 2020).

This unified framework – which nests characteristic-driven factor loadings and characteristic-driven mispricing within a single hierarchical prior – also generalizes conditional factor models (e.g., Lettau and Ludvigson, 2001; Gagliardini et al., 2016) by

embedding them in a Bayesian setting with time-varying coefficients and component-specific sparsity, yielding a richer description of how exposures evolve with economic conditions. It also complements and extends the Bayesian model selection and shrinkage literature (e.g., Avramov, 2002; Barillas and Shanken, 2018; Chib et al., 2020, 2024; Bryzgalova et al., 2023; Cong et al., 2023) by shifting attention from averaging over a pre-specified set of models to letting the data discipline the degree of sparsity directly. In particular, whereas model uncertainty is typically handled by averaging across a large space of competing factor-model specifications, we take a complementary approach by treating sparsity in the characteristic space as a latent object to be learned.

Finally, our framework provides a new perspective on regime-dependent pricing: during periods of macroeconomic stress, the characteristic space that drives factor loadings becomes effectively lower-dimensional, with the posterior concentrating on a smaller set of characteristics (e.g., Ang and Kristensen, 2012).

The paper proceeds as follows. Section 2 introduces the model and its theoretical framework. Section 3 describes the data sources and key characteristics. Section 4 presents our main empirical results under the specification without mispricing, examining how sparsity varies across assets and over time, and then extends the framework to observable factor models. Section 5 evaluates the framework once mispricing is allowed, documenting the magnitude of the estimated mispricing component and its out-of-sample predictive performance. Section 6 concludes with a summary of findings and implications.

## 2 Methodology

We investigate the structural origins of cross-sectional returns by treating sparsity as an inferential object. One view is that dispersion in expected returns is organized around a small number of characteristic channels that capture a common component and a mispricing component. Another view is that the cross section is genuinely high-dimensional, with many small effects that remain economically meaningful. Rather than committing to either view *a priori*, we build a framework that learns the degree

and uncertainty of sparsity directly from the data.

Sparsity can be studied within the two main empirical frameworks for asset pricing: beta pricing models and stochastic discount factor (pricing kernel) models. For clarity and comparability, we build on the conditional latent factor model of Kelly et al. (2019), which links firm characteristics to time-varying covariances and provides a flexible structure for modeling asset returns. We further integrate it with the Bayesian latent factor model of Geweke and Zhou (1996), and introduce a hierarchical prior that induces a posterior distribution over sparsity in both alphas and betas. This formulation yields interpretable posterior summaries of (i) the relevance of individual characteristics and (ii) the overall degree of sparsity in the cross-sectional returns.

To make these objects operational, we specify (i) a conditional return-generating structure in which characteristics govern both abnormal returns and factor exposures, and (ii) a hierarchical prior that treats the extent of sparsity as an unknown parameter to be learned from the data. The key outputs of the framework are characteristic-level posterior relevance measures and posterior distributions over the *degree* of sparsity in the alpha and beta channels, which later sections translate into economically meaningful summaries for cross-sectional returns.

A central advantage of the conditional factor representation is that it separates two distinct sources of cross-sectional expected-return dispersion. In the *beta channel*, characteristics matter because they shape exposures to the latent factors that organize common return variation and its time variation. In the *alpha channel*, characteristics matter because they generate systematic components of expected returns that are not captured by those factor exposures.

### 2.1 Structural Motivation: Latent Economic Dimension and Sparsity

This section complements the empirical motivation in the Introduction with a structural rationale for Schrödinger's Sparsity. Empirically, the central challenge is that the evidence for exclusion versus inclusion is not uniform: across assets and economic conditions, some characteristics can appear weakly related to expected returns

in one setting yet matter in another, and modest changes in specification or thresholding can alter which characteristics are classified as "active". As a result, approaches that hard-wire a fixed sparsity level – whether very sparse or fully dense – can materially change the implied decomposition of expected returns into systematic and residual components. This measurement-driven ambiguity motivates treating sparsity as an inferential object rather than a maintained restriction.

The key structural observation is that sparsity is a statement about coefficients in the *observed* characteristics basis, not about the underlying economic primitives. When characteristics are noisy measurements of latent forces, the (unobserved) measurement system can map the same underlying structure into a representation that is either sparse or diffuse in the observed basis.

As shown in Appendix III, the latent economic dimension does not identify whether the population-optimal representation in the observed basis is sparse or dense. Even when the economy is driven by a single latent primitive, the optimal linear mapping from observed characteristics to exposures can be exactly sparse or fully dense, depending on the measurement map and the noise structure. The same logic applies when the latent economy is higher-dimensional: a richer set of primitives does not, by itself, justify a fully dense representation, nor does a smaller set of primitives justify committing ex ante to sparsity. Recognizing this nonidentification reframes the econometric problem: rather than choosing between a "sparse" and a "dense" model a priori, the researcher faces model uncertainty over admissible sparsity patterns.

Appendix III further shows that maintaining a superposition of such patterns – that is, a Bayesian mixture over sparse and diffuse representations – enjoys a standard decision-theoretic guarantee under log-score loss. Operationally, a spike-and-slab prior provides a natural implementation: it induces a posterior distribution over sparsity patterns and weights them by their empirical support. The resulting posterior inclusion (or exclusion) probabilities deliver an interpretable, state-dependent measure of effective sparsity and its uncertainty.

These considerations above motivate treating sparsity as a latent inferential object

rather than a fixed structural restriction, and learning it through a Bayesian framework with a spike-and-slab prior.

### 2.2 Model Setting

We first establish notation. Let $r_{i,t}$ denote the excess return of asset $i$ at time $t$, and define the cross section of excess returns at time $t$ as $\mathbf{r}_t = (r_{1,t}, \ldots, r_{N_t,t})^\top$, where $N_t$ represents the number of assets available at time $t$. The dataset spans $T$ time periods and forms an unbalanced panel, with $N_t$ varying over time. For each asset $i$, let $\mathbf{Z}_{i,t-1} = (z_{i,1,t-1}, \ldots, z_{i,L,t-1})^\top$ denote the $L$-dimensional vector of lagged characteristics observed at time $t-1$.

The common factors $\mathbf{f}_t \in \mathbb{R}^K$ evolve over time and capture systematic sources of return variation across assets. Our baseline specification treats the factors as latent; however, the framework can readily incorporate observable return-spread factors (e.g., the market or Fama-French factors). In the empirical analysis, we primarily focus on latent factors, while also examining a variant specification with observed factors.

The conditional factor model specifies time-varying alphas and betas as functions of the firm characteristics:

$$\begin{aligned} r_{i,t} &= \boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) + \boldsymbol{\beta}(\mathbf{Z}_{i,t-1})^\top \mathbf{f}_t + \epsilon_{i,t}, \\ \boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) &= \alpha_0 + \boldsymbol{\alpha}_1^\top \mathbf{Z}_{i,t-1}, \\ \boldsymbol{\beta}(\mathbf{Z}_{i,t-1}) &= \boldsymbol{\beta}_0 + \left(\mathbb{I}_K \otimes \mathbf{Z}_{i,t-1}^\top\right) \boldsymbol{\beta}_1. \end{aligned} \tag{1}$$

Here, $\alpha_0$ is the potential time-invariant mispricing, $\boldsymbol{\alpha}_1$ is an $L \times 1$ vector of characteristic loadings for the intercept (mispricing) component, $\boldsymbol{\beta}_0$ is a $K \times 1$ vector of baseline factor loadings, and $\boldsymbol{\beta}_1$ is a $KL \times 1$ vector capturing how loadings on each of the $K$ factors vary with the $L$ lagged characteristics. The Kronecker product $\mathbb{I}_K \otimes \mathbf{Z}_{i,t-1}^\top$ stacks the $L$ characteristics separately for each factor, so $\boldsymbol{\beta}_1$ spans all $K \times L$ factor-characteristic interactions.

Our main focus is on sparsity in $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$, which identify the relevant characteristics for explaining alphas and betas. We develop two priors. The primary one,

used to infer the probability of sparsity, is detailed in Section 2.3. As a benchmark for comparison, the second assumes fixed sparsity, as discussed in Section 2.4.

We formulate the model within a Bayesian framework that generalizes the unconditional latent factor structure of Geweke and Zhou (1996). For each period $t$, the latent factors are assigned a normal prior with $\mathbb{E}[\mathbf{f}_t] = \mathbf{0}$, $\mathbb{E}[\mathbf{f}_t\mathbf{f}_t^\top] = \mathbb{I}_K$, and $\mathbb{E}[\boldsymbol{\epsilon}_t \mid \mathbf{f}_t] = \mathbf{0}$. The residuals are assumed to follow a multivariate normal distribution, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_{N_t}^2)$ is a diagonal covariance matrix. We place conjugate inverse-gamma priors on the residual variances of each asset: $\sigma_i^2 \sim \mathcal{IG}\left(v_{0,i}/2, S_{0,i}/2\right)$.

Substituting the characteristic-based specifications of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ into the return equation yields a pooled regression representation of the conditional factor model:

$$r_{i,t} = \alpha_0 + \boldsymbol{\alpha}_1^\top \mathbf{Z}_{i,t-1} + \boldsymbol{\beta}_0^\top \mathbf{f}_t + \boldsymbol{\beta}_1^\top \left[\mathbf{f}_t \otimes \mathbf{Z}_{i,t-1}\right] + \epsilon_{i,t}, \tag{2}$$

Inference is then carried out in a Bayesian framework via Markov chain Monte Carlo (MCMC).

Equation (1) integrates two distinct conditional factor model formulations, enabling a formal test of the factor spanning restriction. Setting the mispricing component to zero,

$$\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) = \mathbf{0}, \tag{3}$$

i.e., $\alpha_0 = 0$ and $\boldsymbol{\alpha}_1 = \mathbf{0}$, enforces the strong constraint that characteristics affect expected returns only through factor exposures, a key assumption in structural factor models like IPCA. All predictable variation in $\mathbb{E}[r_{i,t} \mid \mathbf{Z}_{i,t-1}]$ must then be absorbed by the characteristic-driven exposures $\boldsymbol{\beta}(\mathbf{Z}_{i,t-1})$ and the associated factor premia. Therefore, the latent factors are required to span the cross-sectional returns, and the model operates as a structural factor specification in the spirit of Kelly et al. (2019): the mapping $\mathbf{Z}_{i,t-1} \mapsto \boldsymbol{\beta}(\mathbf{Z}_{i,t-1})$ encodes systematic exposure to common return variation, and the latent factors admit a direct interpretation as priced sources of systematic variation.

In contrast, allowing for a nonzero mispricing function,

$$\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) \neq \mathbf{0}, \tag{4}$$

relaxes the spanning restriction. Characteristics are allowed to enter expected returns directly, in addition to their role in shaping factor exposures. In this more flexible formulation, $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1})$ captures cross-sectional expected-return components not accounted for by the factor exposure mapping, while $\boldsymbol{\beta}(\mathbf{Z}_{i,t-1})$ and factors absorb systematic common variation in returns. Consequently, the model provides a flexible data-generating representation for expected returns, rather than a pure factor model.

These two cases imply different locations for sparsity: the restriction $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) = \mathbf{0}$ forces all characteristic content into the exposure mapping $\boldsymbol{\beta}(\mathbf{Z}_{i,t-1})$, whereas $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) \neq \mathbf{0}$ allows characteristic content to operate through both channels. Our inference framework is designed to quantify the degree of concentration in each channel via the posterior distribution. Accordingly, Sections 4 and 5 analyze the two specifications in parallel: Section 4 focuses on the factor-only model with $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) = \mathbf{0}$, while Section 5 introduces a nonzero $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1})$ and evaluates how allowing $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) \neq \mathbf{0}$ alters sparsity, interpretation, and performance.

**Model Identification.** We follow the identification strategy of Kelly et al. (2019). Let $\Gamma_\alpha = [\alpha_0, \boldsymbol{\alpha}_1]$ and define

$$\Gamma_\beta = \begin{bmatrix} \boldsymbol{\beta}_0^\top \\ \text{vec}_{L,K}^{-1}(\boldsymbol{\beta}_1) \end{bmatrix} \in \mathbb{R}^{(1+L)\times K}, \tag{5}$$

where $\text{vec}_{L,K}^{-1}(\cdot)$ reshapes a $KL \times 1$ vector into an $L \times K$ matrix using column-stacking (so that $\text{vec}(\text{vec}_{L,K}^{-1}(x)) = x$).

We impose the constraints

$$\Gamma_\beta^\top \Gamma_\beta = \mathbb{I}_K, \qquad \Gamma_\alpha^\top \Gamma_\beta = \mathbf{0}_{1\times K}, \tag{6}$$

to ensure that the unconditional second-moment matrix of $\mathbf{f}_t$ is diagonal with descending diagonal entries and to eliminate the usual scale and rotation indeterminacies in latent factor models.

To resolve sign indeterminacy, we implement a deterministic identification by multiplying both the factor and corresponding elements in the coefficients $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ by $-1$ whenever the sample factor mean is negative. This transformation is a one-to-one reparameterization that leaves the product $\Gamma_\beta \mathbf{f}_t$ (and hence the likelihood and posterior distribution) unchanged. It is imposed purely for identification, ensuring that posterior summaries of $\mathbf{f}_t$ and $\Gamma_\beta$ are stable and interpretable. It is not equivalent to imposing a prior that factors must have positive means, as this would alter the probability space. To preserve the structure of $\Gamma_\beta$, the identifications are applied separately to each factor.

We enforce the orthogonality restriction in (6) by projecting $\Gamma_\alpha$ onto the column space of $\Gamma_\beta$ and replacing $\Gamma_\alpha$ with the projection residual (equivalently, regressing $\Gamma_\alpha$ on $\Gamma_\beta$ and retaining the residual). Intuitively, these restrictions eliminate the usual rotation and scale indeterminacies in factor models: $\Gamma_\beta^\top \Gamma_\beta = \mathbb{I}_K$ normalizes the factors, while $\Gamma_\alpha^\top \Gamma_\beta = \mathbf{0}$ ensures that mispricing is orthogonal to factor exposures. We impose these constraints at each iteration of the sampler.

Next, we introduce a hierarchical spike-and-slab prior to identify the sparsity structure of the characteristics influencing both alphas and betas, which are central to the model's framework.

### 2.3 Priors Over Sparsity Structures

We employ hierarchical spike-and-slab priors for $\boldsymbol{\alpha}(\mathbf{Z})$ and $\boldsymbol{\beta}(\mathbf{Z})$ effects rather than pre-specifying shrinkage parameters or penalty estimators. Under this kind of prior, the set of relevant characteristics, the magnitude of their effects, and the overall degree of sparsity are all objects of posterior inference.

The prior combines exact exclusion (a "spike" at zero) with continuous shrinkage for included coefficients (a Gaussian "slab"). Let $d_l^\alpha \in \{0, 1\}$ indicate whether

characteristic $l$ enters the mispricing function $\boldsymbol{\alpha}(\mathbf{Z})$, i.e., whether $\alpha_{1,l}$ is included. Let $d_l^\beta \in \{0,1\}$ indicate whether characteristic $l$ enters the factor-loading function $\boldsymbol{\beta}(\mathbf{Z})$. For $\boldsymbol{\beta}(\mathbf{Z})$ we impose grouped selection: a characteristic is either excluded from *all* factor loadings or included in *all* of them. [3]

Formally, for each $l = 1, \ldots, L$,

$$\alpha_{1,l} \mid d_l^\alpha, \gamma_\alpha^2 \sim (1 - d_l^\alpha)\, \delta_0 + d_l^\alpha\, \mathcal{N}(0, \gamma_\alpha^2), \qquad \mathbf{b}_l \mid d_l^\beta, \gamma_\beta^2 \sim (1 - d_l^\beta)\, \delta_{\mathbf{0}} + d_l^\beta\, \mathcal{N}(\mathbf{0}, \gamma_\beta^2 \mathbb{I}_K), \quad (7)$$

where $\delta_0$ denotes a point mass at zero, $\mathbf{b}_l \in \mathbb{R}^K$ stacks the $K$ coefficients in $\boldsymbol{\beta}_1$ associated with characteristic $l$, and $\delta_{\mathbf{0}}$ denotes a point mass at $\mathbf{0} \in \mathbb{R}^K$. Equivalently, $d_l^\alpha$ controls whether $\alpha_{1,l}$ is exactly zero, while $d_l^\beta$ controls whether all $K$ coefficients linked to characteristic $l$ in $\boldsymbol{\beta}_1$ are exactly zero.

Conditional on inclusion, coefficients follow Gaussian slab priors that shrink estimates toward zero without forcing exact sparsity. The slab variances $\gamma_\alpha^2$ and $\gamma_\beta^2$ control the degree of shrinkage and are assigned inverse-gamma hyperpriors:

$$\gamma_\alpha^2 \sim \mathcal{IG}\left(A_{\gamma_\alpha}/2, B_{\gamma_\alpha}/2\right), \qquad \gamma_\beta^2 \sim \mathcal{IG}\left(A_{\gamma_\beta}/2, B_{\gamma_\beta}/2\right). \quad (8)$$

Each inclusion indicator follows a Bernoulli prior,

$$d_l^\alpha \sim \text{Bernoulli}(1 - q_\alpha), \qquad d_l^\beta \sim \text{Bernoulli}(1 - q_\beta), \quad (9)$$

where $q_\alpha$ and $q_\beta$ govern the *exclusion* probabilities (hence the overall sparsity rates) for $\boldsymbol{\alpha}(\mathbf{Z})$ and $\boldsymbol{\beta}(\mathbf{Z})$, respectively; equivalently, $1 - q_\alpha$ and $1 - q_\beta$ are inclusion probabilities. While standard spike-and-slab specifications (Mitchell and Beauchamp, 1988) treat these probabilities as fixed hyperparameters, we follow Giannone et al. (2021)

[3] The grouped rule imposes a common set of selected characteristics across the $K$ factor loadings. This aligns with the interpretation of sparsity at the level of characteristics, and it avoids factor-by-factor selection, which is harder to interpret economically.

and assign independent Beta hyperpriors:

$$q_\alpha \sim \text{Beta}(a_{q_\alpha}, b_{q_\alpha}), \qquad q_\beta \sim \text{Beta}(a_{q_\beta}, b_{q_\beta}). \tag{10}$$

This hierarchy lets the data inform the degree of sparsity, while still allowing the researcher to encode weak prior beliefs through $(a_q, b_q)$.

Different hyperparameter choices correspond to different prior beliefs about sparsity. These beliefs are not binding: the posterior distributions of $q_\alpha$ and $q_\beta$ reflect both the priors and the observed data. Notably, the resulting posteriors are supported on $(0, 1)$ and therefore deliver a *probabilistic* description of sparsity rather than a single binary claim about whether the model is "sparse" or "dense." This formulation avoids sharp thresholding and instead quantifies uncertainty about overall sparsity. While one can compute sparsity *ex post* by counting selected coefficients (He et al., 2024), our approach treats sparsity probabilities as primitive inferential objects and reports their posterior uncertainty directly.

One might argue that traditional variable-selection methods, such as the Lasso, also "learn" sparsity by selecting a penalty parameter via cross-validation. Cross-validation, however, is a model-selection tool based on data splitting rather than an inferential procedure for the full-sample likelihood. Once a penalty is chosen, it is treated as fixed, producing a single sparsity pattern with no direct quantification of uncertainty about the degree of sparsity. Moreover, cross-validation delivers one penalty value optimized for predictive performance under a chosen loss; it neither yields a distribution over sparsity levels nor averages across alternative specifications. In contrast, our Bayesian learning-sparsity framework produces posterior sparsity probabilities that are directly interpretable and integrates over the space of sparsity structures. a unified probabilistic formulation.

### 2.4 Prior for Exogenous Fixed Sparsity Level

To highlight the value of learning sparsity probabilistically, we also consider a benchmark prior in which the sparsity level is fixed *ex ante* rather than inferred from the data. This specification reflects the conventional approach that presumes a predetermined degree of sparsity or density. Conceptually, it mirrors best subset selection with a fixed cardinality and shares a limitation with penalized regression methods such as the Lasso: once a tuning parameter (e.g., the penalty $\lambda$) is fixed, the model is confined to a predetermined region of the parameter space, preventing sparsity from adapting to the data as in our hierarchical framework.

In our fixed-sparsity benchmark, this restriction is implemented by placing a hard constraint on the number of included predictors. Specifically, the inclusion indicators must satisfy

$$\sum_{l=1}^{L} d_l^{\alpha} = M_{\alpha}, \qquad \sum_{l=1}^{L} d_l^{\beta} = M_{\beta}, \tag{11}$$

where $M_{\alpha}$ and $M_{\beta}$ are user-specified integers. Formally, the joint priors are given by

$$\begin{aligned}
(d_1^{\alpha}, d_2^{\alpha}, \cdots, d_L^{\alpha}) &\propto \left[\prod_{l=1}^{L} \text{Bernoulli}(1-q_{\alpha})\right] \times \mathbb{I}\left(\sum_{l=1}^{L} d_l^{\alpha} = M_{\alpha}\right), \\
(d_1^{\beta}, d_2^{\beta}, \cdots, d_L^{\beta}) &\propto \left[\prod_{l=1}^{L} \text{Bernoulli}(1-q_{\beta})\right] \times \mathbb{I}\left(\sum_{l=1}^{L} d_l^{\beta} = M_{\beta}\right),
\end{aligned} \tag{12}$$

so that only $M_{\alpha}$ characteristics can enter the mispricing function and only $M_{\beta}$ can enter the factor-loading function. By making model size deterministic, the indicator constraint prevents the sparsity level from adapting to the data. By comparing this fixed-sparsity benchmark with our learning-sparsity prior, we can assess the empirical and economic implications of treating sparsity as an unknown quantity rather than a predetermined one.

To perform inference on all model parameters and latent factors, we draw samples from the joint posterior distribution using a Gibbs sampling algorithm. Detailed information on the sampling procedure is provided in Appendix I.

## 3 Data

**Test Assets.** We study the sparsity of asset pricing models using a broad range of U.S. test assets, including individual stocks and constructed portfolios. Our database covers monthly returns for 21,165 individual stocks from January 1980 through December 2024.[4] Portfolios are constructed from these stocks, and we also consider a representative subset of individual stocks as test assets.

We also utilize test assets constructed using the Panel-Tree (P-Tree) method of Cong et al. (2025). The procedure grows 40 trees with 10 leaves each, producing 400 portfolios that generalize multi-way characteristic sorts through a transparent, data-driven partitioning of the firm-characteristics space. P-Tree portfolios are particularly suitable for studying sparsity because they are generated sequentially to expand the span of the test-asset space in an economically disciplined way. Early trees exploit the most informative interactions among characteristics, while later trees extract increasingly subtle return patterns. As a result, enlarging the P-Tree set provides a systematic way to increase the empirical "difficulty" of the pricing problem, making it a natural environment to examine how sparsity in factor exposures adapts as the cross section becomes progressively harder to explain.

To mitigate look-ahead bias, we follow Cong et al. (2025) in constructing the P-Tree portfolios using data from 1980 to 1989. We then fix the tree structure and apply it to generate test assets from January 1990 to December 2024, which forms the main sample for our empirical analysis. For consistency, the sample period for all other test assets is also limited to January 1990 through December 2024, yielding 420 months.

Our primary test assets consist of the first 100 P-Tree portfolios, which exhibit higher Sharpe ratios compared to the subsequent 300 portfolios. Figure A.1 compares the cross-sectional dispersion of 100 P-Tree portfolios with the 25 ME/BM portfolios, highlighting the greater dispersion of the former.

[4]We apply standard filters following Fama and French (1992), including: (1) restricting the sample to stocks listed on the NYSE, AMEX, or NASDAQ for at least one year; (2) selecting firms with CRSP share codes of 10 or 11; and (3) excluding stocks with negative book equity or lagged market equity. The sample begins in 1980 to ensure sufficient I/B/E/S coverage.

In addition to P-Tree portfolios, we examine three commonly used sets of test assets: (i) the 25 ME/BM portfolios, (ii) 360 bivariate-sorted portfolios (Bi360),[5] (iii) 610 univariate-sorted portfolios (Uni610),[6] based on individual characteristics.

**Characteristics.** To construct the instrument set, we compile 20 representative firm-level characteristics, grouped into eight themes: size, value, investment, momentum, profitability, liquidity, volatility, and intangibles. Table A.1 provides detailed descriptions. At the individual-stock level, each characteristic is standardized cross-sectionally to lie within the range $[-1, 1]$.

All portfolio returns and characteristics are value-weighted averages of their constituent stocks, where the portfolio characteristics are formed from the raw firm-level characteristic values. We then re-standardize the portfolio-level characteristics to lie within the range $[-1, 1]$, cross-sectionally within each test-asset class and month.

**Regime.** To examine how model sparsity varies across macroeconomic environments, we incorporate regime-specific analyses. For structural regimes, we follow Smith and Timmermann (2021), who identify major breakpoints in the U.S. stock market and provide corresponding calendar dates. We treat these dates as exogenous partitions and define three regimes accordingly: Regime 1 spans from January 1990 to July 1998 (103 months), Regime 2 extends from August 1998 to June 2010 (143 months), and Regime 3 covers from July 2010 to December 2024 (174 months).

Additionally, we employ the Sahm Rule Recession Indicator from FRED to capture business cycle dynamics. This metric identifies recessions when the three-month average unemployment rate surpasses its 12-month low by at least 0.5 percentage points. Recessionary months identified by the indicator are aggregated into a single recession regime, contrasted with normal periods. The recession regime spans 88 months, encompassing October 1990-November 1992, May 2001-October 2002, March 2008-May 2010, March 2020-March 2021, and June 2024-September 2024.

[5] Constructed under a $2 \times 3 \times 60$ scheme. Two groups remain empty due to the absence of stock assignments; dependent sorting is applied in these cases.

[6] Constructed by independently sorting stocks into 10 deciles for each of a broad set of 61 characteristics, yielding 610 portfolios.

# 4 Schrödinger's Sparsity

A central goal of this paper is to treat sparsity as an inferential object rather than an imposed modeling choice. By adopting a fully probabilistic framework, we bridge the dichotomy between sparse and dense representations, let the data determine how rich or parsimonious the characteristic structure should be, and we quantify uncertainty about that structure.

We begin by studying the specification in Equation (2) under the restriction $\boldsymbol{\alpha}(\mathbf{Z}) \equiv \mathbf{0}$. Throughout this section, we impose this *no-mispricing* constraint: all predictable return variation must be absorbed by characteristic-driven factor loadings $\boldsymbol{\beta}(\mathbf{Z})$ and the associated factor premia. This restriction provides a clean laboratory to examine (i) how the data allocate explanatory power across characteristics, (ii) how sparsity emerges in the construction of latent factors, and (iii) how posterior uncertainty about sparsity varies across test assets and economic conditions. We then evaluate fit and out-of-sample performance under this restricted specification.

In Section 5, we relax $\boldsymbol{\alpha}(\mathbf{Z}) \equiv \mathbf{0}$ and allow characteristic-driven mispricing. This extension changes the "division of labor" in the model – some predictable variation is captured directly by $\boldsymbol{\alpha}(\mathbf{Z})$ rather than being forced into $\boldsymbol{\beta}(\mathbf{Z})$ – and therefore provides a complementary perspective on how sparsity is learned and interpreted.

## 4.1 Learning Sparsity

We first fit the model with a representative set of 100 P-Tree test assets. This choice balances economic richness and transparency: P-Tree portfolios span a wide cross section with high dispersion and interpretable construction characteristics. Starting from this tractable yet informative set allows us to (i) analyze the posterior sparsity of factor loadings and (ii) assess the gains from learning $q_\beta$ relative to fixed-sparsity and dense benchmarks. Broader asset classes and regime analyses follow in Section 4.2.

**Models Are Not Very Sparse.** Panel A of Table 1 highlights a central finding of our sparsity analysis: the factor-loading function is not highly sparse, reinforcing the *Illu-*

*sion of Sparsity* perspective (Giannone et al., 2021) in empirical asset pricing. Specifically, the posterior mean number of selected characteristics consistently falls between 11 and 13 out of 20, demonstrating that neither extreme sparsity nor full density is empirically supported. This tight band indicates that the structure implied by the data requires a meaningful, but not excessive, set of characteristics to describe factor exposures. Put differently, the model learns an interior level of concentration in the mapping from characteristics to factor loadings: neither a handful of dominant predictors nor a fully dense set of small effects.

Table 1: Performance for Various Models

This table reports results out-of-sample (OOS) cross-sectional $R^2$ ($\text{CSR}^2$), OOS Sharpe ratio (SR) of the latent factor tangency portfolio, and the posterior mean of the in-sample sparsity probabilities ($q_\beta$). Values in parentheses below $q_\beta$ denote the posterior number of selected characteristics $\widehat{M}_\beta$, defined as the number of characteristics with posterior inclusion probabilities $> 0.5$. $K$ indicates the number of latent factors. Panel A estimates sparsity endogenously using three Beta priors for $\text{Beta}(9,1)$, $\text{Beta}(5,5)$, and $\text{Beta}(1,9)$ (means 0.9, 0.5, and 0.1). Panel B restricts the number of characteristics $M_\beta$ for each factor loading $\boldsymbol{\beta}_{1,k}$. Panels C and D present results from the Bayesian framework without sparsity and the standard dense IPCA, respectively. The sample consists of a 15-year in-sample estimation period and a 20-year out-of-sample evaluation period.

| | | Out–of-sample | | | | | | In-sample | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\text{CSR}^2$ | | | SR | | | $q_\beta$ and $\widehat{M}_\beta$ | | |
| | | $K=1$ | $K=3$ | $K=5$ | $K=1$ | $K=3$ | $K=5$ | $K=1$ | $K=3$ | $K=5$ |
| *Panel A: Learned Sparsity* | | | | | | | | | | |
| | 0.9 | 14.7 | 63.1 | 68.3 | 0.49 | 0.38 | 0.92 | 0.59 (11) | 0.57 (12) | 0.60 (11) |
| $q_\beta$ prior mean | 0.5 | 14.6 | 61.9 | 68.0 | 0.49 | 0.51 | 0.90 | 0.43 (11) | 0.44 (12) | 0.47 (11) |
| | 0.1 | 14.5 | 62.8 | 68.9 | 0.49 | 0.54 | 0.98 | 0.24 (13) | 0.30 (12) | 0.32 (11) |
| *Panel B: Fixed Sparsity* | | | | | | | | | | |
| | 2 | 13.6 | 63.1 | 62.7 | 0.50 | 0.23 | 0.63 | / | / | / |
| $M_\beta$ | 10 | 13.8 | 62.6 | 64.9 | 0.49 | 0.60 | 0.66 | / | / | / |
| | 18 | 14.9 | 64.0 | 66.2 | 0.49 | 0.54 | 0.55 | / | / | / |
| *Panel C: No Sparsity* | | | | | | | | | | |
| $M_\beta$ | 20 | 14.4 | 62.8 | 65.4 | 0.49 | 0.45 | 0.45 | / | / | / |
| *Panel D: IPCA* | | | | | | | | | | |
| $M_\beta$ | 20 | 17.8 | 61.7 | 70.8 | 0.33 | 0.50 | 0.74 | / | / | / |

Although the posterior estimate of sparsity probability $q_\beta$ changes with different prior means (e.g., at $K = 1$, the posterior values are 0.59, 0.43, and 0.24 when

the prior means are 0.9, 0.5, and 0.1, corresponding to strongly sparse, agnostic, and strongly dense prior views, respectively), these differences leave the substantive conclusion unchanged: the model persistently selects roughly half of the characteristics. Furthermore, the cross-sectional $R^2$s and Sharpe ratios remain largely stable.[7] This reflects standard Bayesian updating: researchers may start from different views about sparsity, but once confronted with evidence that around 11 characteristics matter on average, their posterior beliefs move toward this intermediate level. More dogmatic priors simply adjust more gradually.

For interpretability and neutrality, all subsequent discussions of sparsity levels use the $\text{Beta}(5, 5)$ prior with a mean of 0.5, which reflects an *agnostic stance* that each characteristic is equally likely to be included or excluded ex ante. Thus, the posterior sparsity level is driven primarily by the data rather than biased by a dogmatic prior.

Furthermore, the sparsity pattern varies systematically with the number of latent factors $K$. Under the agnostic prior, the posterior sparsity probability rises from 0.43 to 0.44 and 0.47 as $K$ increases from 1 to 3 to 5. This modest but monotonic pattern is robust across priors.

Economically, this reflects how the model reallocates characteristic information as the number of factors increases. With fewer factors, each loading must capture broad cross-sectional heterogeneity, requiring a richer set of characteristics. As $K$ grows, that heterogeneity can be distributed across factors, so each loading relies on a narrower subset of characteristics. The resulting structure is multi-dimensional but not excessively so: additional factors help separate distinct components of common variation, while learning sparsity prevents overfitting by pruning weak characteristic-loading links as factor dimensionality increases.

**Learning Sparsity Matters.** Having shown that the model learns a stable, mid-range sparsity level (roughly 11 selected characteristics), a natural question arises: what hap-

[7]The cross-sectional $R^2$ is computed as: $\text{CSR}^2 = 1 - \frac{\sum_{i=1}^{N}\left(\frac{1}{T_i}\sum_{t=1}^{T_i}(r_{i,t}-\widehat{r}_{i,t})\right)^2}{\sum_{i=1}^{N}\left(\frac{1}{T_i}\sum_{t=1}^{T_i}(r_{i,t}-\widehat{\beta}_i\text{MktRF}_t)\right)^2}$, where $\widehat{r}_{i,t}$ denotes the predicted mean return, $\widehat{\beta}_i$ is the estimated market beta, and $\text{MktRF}$ is the market factor.

pens if sparsity is not learned but fixed ex ante? To answer this question, we further illustrate a different prior where the number of characteristics is fixed at given levels, $M_\beta \in \{2, 10, 18\}$, and Panel B of Table 1 summarizes the out-of-sample results. All other elements of the Bayesian structure are held the same.

The results are revealing. When sparsity is fixed at levels that are too sparse ($M_\beta = 2$) or too dense ($M_\beta = 18$), model performance deteriorates notably. For example, at $K = 5$, the out-of-sample Sharpe ratio is 0.63 when only two characteristics are allowed, improves to 0.66 when $M_\beta = 10$, and then falls to 0.55 when the model becomes overly dense at $M_\beta = 18$. These choices misalign the imposed structure with the dimensionality required by the data, leading to lower performance. The fixed $M_\beta = 10$ specification performs best as it approximates the sparsity level learned endogenously. Thus, exogenously fixing sparsity is effectively equivalent to choosing model dimensionality before seeing the data: it works well only when the chosen level happens to coincide with the structure implied by the data. The learning sparsity approach mitigates this risk by adapting to the data rather than imposing a fixed structure.

Beyond fixed sparsity benchmarks, it is also informative to compare our framework with models that impose no sparsity at all, which is effectively setting $M_\beta = 20$ and treating all characteristics as relevant. Panel C of Table 1 shows that the fully dense Bayesian specification is uniformly dominated by the learning-sparsity model. Without the ability to prune weak signals, the dense model delivers lower out-of-sample Sharpe ratios and cross-sectional $R^2$, indicating that some level of sparsity is still essential for capturing the structure of the cross section.

Panel D provides a benchmark based on the dense IPCA specification. At $K = 5$, IPCA achieves a $CSR^2$ of 70.8%, and the learning-sparsity model remains close at 68.9%, indicating comparable cross-sectional explanatory power. The two approaches diverge more in out-of-sample tangency performance: the best Sharpe ratio in Panel A at $K = 5$ is 0.98, versus 0.74 for IPCA. This gap highlights that differences in economic performance can be larger than what $CSR^2$ alone would imply, and it motivates evaluating fit and investability jointly.

Treating sparsity as a latent quantity – rather than imposing full density or an arbitrary fixed level – allows the model to adapt to the underlying structure of the data, which in turn improves pricing fit and the out-of-sample performance of the implied factor portfolios. Across priors and factor dimensions, the posterior favors an interior level of sparsity: roughly 11-13 of 20 characteristics are effectively active, consistent with an "illusion of sparsity." Since these conclusions remain stable across priors, we utilize the neutral Beta$(5, 5)$ prior as our baseline going forward.

### 4.2 Schrödinger's Sparsity Everywhere

We now move beyond the baseline test-asset design and ask how the inferred loading structure responds to changes in the information set. Specifically, we examine whether model-implied sparsity is stable across asset universes and whether it varies systematically over time. Section 4.2.1 analyzes alternative test assets and relates their heterogeneity to posterior sparsity. Section 4.2.2 studies macroeconomic regime dependence and evaluates how the estimated factor loadings evolve over time and across asset classes.

#### 4.2.1 Test Assets and Sparsity

It is well recognized that the choice of test assets plays a pivotal role in asset pricing analysis, determines the spanning and identifiability of the return space, and, in turn, influences how model explanatory power is allocated in factor loadings. We hypothesize that model sparsity systematically varies with the heterogeneity and construction of test assets: simpler portfolio sorts, such as the ME/BM 25 portfolios, rely on fewer characteristics to explain returns. In contrast, more informationally rich or heterogeneous test assets, such as those constructed via the P-Tree framework, may require a broader set of predictors due to their increased dimensionality.

To examine this hypothesis, we estimate the model with five latent factors and a prior mean of $q_\beta = 0.5$ on three distinct asset groups: (i) P-Tree portfolios constructed with varying numbers of leaves; (ii) individual stocks split by average market eq-

uity (Big 500 vs. Small 500); [8] and (iii) benchmark portfolios including the ME/BM25, Bi360, and Uni610 sets.

We begin by examining how posterior sparsity probabilities change in response to variations in the cross-sectional size of test assets within the same asset class. In Panel A of Table 2, as the number of P-Tree portfolios increases from 100 to 400, the posterior mean of the sparsity probability in factor loadings declines from 0.50 to 0.37 and then to 0.26. This pattern suggests that as more test assets are added to span the efficient frontier, the model requires a richer set of characteristics to explain expected returns, resulting in lower overall sparsity. Besides, the decline in sparsity exhibits diminishing marginal effects: as the test-asset universe becomes sufficiently granular, the marginal independent information contributed by additional assets diminishes, and the required intensity of desparsification can be correspondingly relaxed.

Table 2: Sparsity across Test Assets

This table reports performance metrics for five-factor models (excluding mispricing) across various test assets: individual stocks, P-Tree assets, and three portfolio sets (ME/BM25, Bi360, and Uni610). Reported values include cross-sectional $R^2$ ($CSR^2$), the latent factor tangency portfolio Sharpe ratio (SR), posterior means of $q_\beta$, and the number of selected characteristics $\widehat{M}_\beta$.

| | Full-sample | | | |
|---|---|---|---|---|
| | $CSR^2$ | SR | $q_\beta$ | $\widehat{M}_\beta$ |
| *Panel A: P-Tree Portfolios* | | | | |
| 100 | 59.6 | 1.12 | 0.50 | 10 |
| 200 | 69.4 | 0.68 | 0.37 | 14 |
| 400 | 63.3 | 1.01 | 0.26 | 17 |
| *Panel B: Individual Stocks* | | | | |
| Big500 | 46.9 | 1.54 | 0.30 | 16 |
| Small500 | 30.1 | 4.16 | 0.42 | 12 |
| *Panel C: Other Portfolios* | | | | |
| ME/BM25 | 53.6 | 0.82 | 0.50 | 10 |
| Bi360 | 71.6 | 1.15 | 0.21 | 19 |
| Uni610 | 66.1 | 0.87 | 0.23 | 18 |

Next, we examine whether pricing difficulty influences sparsity while holding the

[8] We begin by selecting stocks with at least ten years of return data. From this group, we rank firms by average market equity and form two equally sized groups: the "Big 500" (ranks 1-500) and the "Small 500" (ranks 501-1000). Small-cap stocks are generally more difficult to price due to lower liquidity and higher idiosyncratic risk.

asset class and the number of test assets fixed. Panel B of Table 2 compares two groups of individual stocks, each comprising 500 securities but with different levels of average market equity. For individual stocks, the underlying factor structure is less transparent than for P-Tree portfolios, where the construction characteristics are known. Despite this difference, the sparsity estimates vary sharply across the two groups.

For factor loadings, the posterior mean of $q_\beta$ differs across size groups (0.30 for large-cap stocks versus 0.42 for small-cap stocks), indicating a meaningful shift in how concentrated the characteristic-to-loading mapping is across firm size. Consistent with this pattern, the model selects fewer characteristics to explain $\beta_1$ among small-cap stocks, suggesting a more concentrated set of predictive channels for their factor exposures (e.g., Daniel and Titman, 1997). While small caps are often viewed as harder to price, they also exhibit substantially higher idiosyncratic volatility, which lowers the signal-to-noise ratio of firm-level predictors. As a result, weaker characteristic signals are effectively drowned out, and only the most dominant characteristics are retained to describe systematic factor exposures. Later in the text, we revisit the relationship between pricing difficulty and sparsity.

Finally, this pattern extends to other commonly used portfolio-based test assets. Panel C of Table 2 shows that the posterior sparsity probability for factor loadings is 0.50 for the ME/BM25 portfolios, compared to about 0.20 for the Bi360 and Uni610 portfolios. The ME/BM portfolios are constructed using only two characteristics, whereas the Bi360 and Uni610 portfolios incorporate a substantially richer set of characteristics. The lower sparsity observed for the latter thus reflects the model's need to access a broader array of characteristics to capture return variation. These results reinforce the idea that the sparsity of the estimated factor-loading structure closely tracks the heterogeneity and informational richness of the test assets.

Overall, Table 2 supports our hypothesis, demonstrating that the structure and heterogeneity of the test assets play a pivotal role in shaping the degree of sparsity, interpretability, and explanatory power of asset pricing models. The results underscore that sparsity is not an inherent model attribute but rather one that adjusts to

the informational richness and design of the test assets. Consequently, effective asset pricing requires striking a balance between parsimony and fit to accurately reflect the underlying process that generates returns.

**Pricing Difficulty versus Sparsity.** The preceding analysis shows that posterior sparsity differs across test-asset universes. To relate these differences to a measurable economic dimension, we introduce a simple empirical proxy for "pricing difficulty": the magnitude of pricing errors relative to standard benchmark models. Specifically, we use the absolute alpha from the CAPM and the Fama-French five-factor (FF5) model. Larger $|\alpha|$ indicates that returns are harder to reconcile with these benchmarks, and therefore correspond to a more challenging pricing environment.

Operationally, we sort the Bi360, Uni610, and P-Tree400 test assets into 6, 10, and 8 groups, respectively, based on their CAPM/FF5 $|\alpha|$, yielding group sizes of 60, 61, and 50 portfolios. [9] For each group, we compute the average absolute CAPM/FF5 alpha and estimate our model with five latent factors to obtain posterior estimates of the sparsity probabilities $q_\beta$.

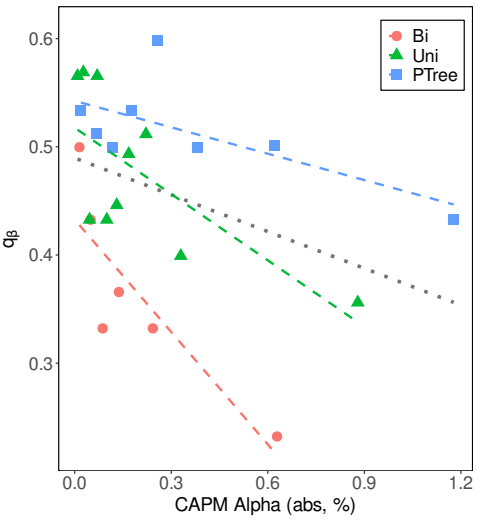
Figure 1 reveals a consistent and economically intuitive pattern: test assets with larger benchmark pricing errors exhibit lower sparsity in factor loadings (lower $q_\beta$). Intuitively, when standard benchmarks fit returns well (smaller absolute CAPM/FF5 alpha), the model can represent factor exposures with a more parsimonious characteristic-to-loading mapping. By contrast, assets that are harder to fit with these benchmarks lead the model to draw on a broader set of characteristics to capture time variation in loadings, reducing sparsity in $\boldsymbol{\beta}_1$. The monotonic nature of this relationship appears across test-asset constructions and across benchmark choices, underscoring that sparsity is not fixed but varies systematically across assets, and pricing difficulty is an important dimension of that variation. These results provide direct evidence for our central thesis: sparsity should be treated as an object of inference, not imposed *ex ante*.

Complementing our analysis of pricing difficulty and sparsity, we also examine

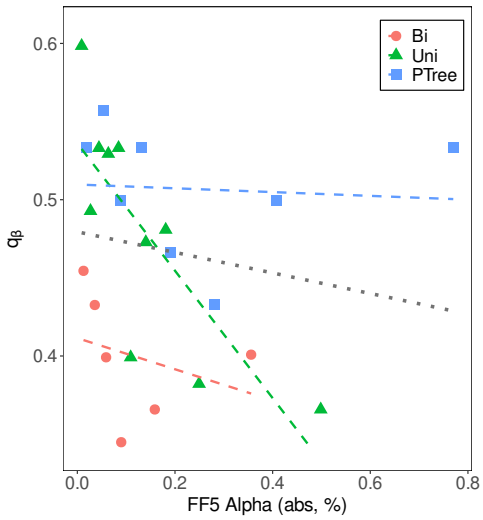
[9] To isolate systematic pricing difficulties from the idiosyncratic noise that dominates individual stocks, we focus this analysis on portfolios.

Figure 1: Sparsity and Pricing Difficulty

This figure illustrates the relationship between sparsity in factor loadings and the difficulty of test asset pricing. Panels (a) and (b) relate $q_\beta$ estimates to average absolute CAPM and FF5 $\alpha$s, respectively. Markers represent bi-sorted (red dots), uni-sorted (green triangles), and P-Tree (blue squares) portfolios. Colored dashed lines denote group-specific regressions; the black dotted line denotes the overall trend.




(a) Group by CAPM $|\alpha|$




(b) Group by FF5 $|\alpha|$

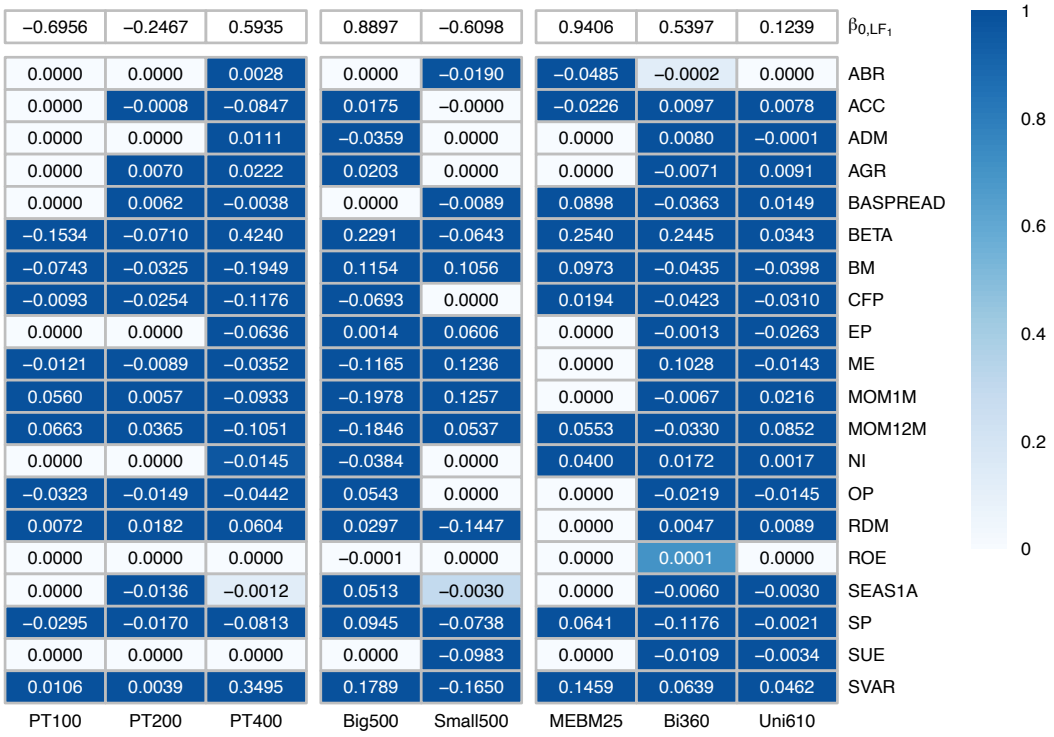
how sparsity relates to the return profile of the test assets. Figure A.2 shows that portfolios with higher Sharpe ratios tend to exhibit denser factor-loading structures. Intuitively, high Sharpe ratio portfolios are often associated with diversified exposure to return premia. Consequently, the model utilizes a broader set of characteristics to span the dynamics of factor loadings, resulting in lower sparsity. This evidence reinforces our central message that sparsity adapts to the asset's economic profile.

**Characteristics Importance.** Having demonstrated that the probability of loading sparsity varies systematically with the economic richness of the test-asset space, we next examine its composition, inquiring which characteristics consistently organize time-varying factor loadings across test-asset sets and which become important only in specific cases. Figure 2 reports the posterior *inclusion* probabilities alongside the corresponding coefficient estimates ($\boldsymbol{\beta}_{1,l}$ for factor loadings) for each characteristic across the various test asset sets.

Several characteristics are consistently selected for factor loadings across all test-

Figure 2: Characteristics Importance across Different Test Assets

This figure illustrates the inclusion probabilities and coefficients of characteristics across test assets. For brevity, we show results for the first latent factor. Each cell reports the coefficient estimate, with color intensity representing its posterior inclusion probability. The corresponding intercept coefficients ($\beta_{0,\mathrm{LF}_1}$) are also provided.




asset sets, with posterior inclusion probabilities close to one. These characteristics include market beta (BETA), book-to-market ratio (BM), twelve-month momentum (MOM12M), sales-to-price ratio (SP), and return variance (SVAR). These variables are strongly associated with the common component of returns, reflecting how portfolios comove with market-wide and macroeconomic fluctuations. Their consistently high inclusion probabilities across test assets highlight their central role in explaining time variation in factor exposures.

Meanwhile, several characteristics exhibit context-dependent importance. For instance, one-month momentum (MOM1M), and the R&D-to-market ratio (RDM) are not selected for the ME/BM 25 portfolios, but contribute meaningfully to return variation in the Bi360, Uni610, and P-Tree portfolios. This pattern reinforces the idea that characteristic relevance depends on the test-asset universe. As the universe becomes more

heterogeneous and the number of distinct return components expands, the model must mobilize a broader set of characteristics in the loading function, which lowers the posterior sparsity probability for factor loadings.

Averaging across all specifications, approximately 70% of the available characteristics are selected for factor loadings, indicating an interior degree of sparsity rather than an extremely sparse or fully dense structure. The probabilistic nature of the Bayesian selection mechanism allows the model to express uncertainty about inclusion while remaining economically interpretable. Furthermore, the estimated coefficients on the first latent factor reveal economically intuitive loadings, particularly for characteristics such as momentum, value, and volatility, demonstrating that the latent factors capture meaningful structure in the cross section of returns.

To summarize, our evidence across multiple sets of test assets reveals that sparsity is not a fixed property, but rather varies systematically with the difficulty of the pricing task. Heterogeneous assets tend to exhibit lower sparsity probabilities. In contrast, simpler and more homogeneous asset sets, such as the ME/BM 25 portfolios, exhibit higher sparsity probabilities, reflecting their limited need for elaborate explanatory structures.

These findings highlight that treating asset pricing models as either intrinsically sparse or intrinsically dense creates an overly simplistic dichotomy. Sparsity is an inferential property determined by the data rather than an intrinsic model attribute. Our probabilistic framework endogenizes the degree of sparsity, thereby providing a flexible and empirically grounded way to characterize expected returns across diverse test-asset environments.

#### 4.2.2 Macro Regimes and Sparsity

Beyond cross-sectional variation, we hypothesize that sparsity also evolves over time, particularly in response to changing macroeconomic conditions. A substantial body of literature documents the time-varying nature of asset pricing relationships. For instance, Avramov and Chordia (2006a) show that factor loadings (betas) vary

systematically with business cycle indicators and offer a business-cycle-based explanation for the role of momentum in the cross section of returns. Likewise, Avramov and Chordia (2006b) demonstrate that the alpha and beta of mean-variance portfolios respond differently to characteristics across economic regimes, with variables related to distress, profitability, and momentum exhibiting heterogeneous effects during expansions and recessions. Furthermore, structural breaks identified by Smith and Timmermann (2021) suggest the presence of distinct economic regimes that necessitate different model specifications, while Li et al. (2023) underscore the temporal variation in return predictability.

These findings suggest that the probability of sparsity, which essentially reflects the relevance of characteristics for factor loadings, may itself be time-varying. Therefore, to investigate this hypothesis, we empirically examine whether sparsity changes across structural breaks and macroeconomic regimes, and assess whether different sets of characteristics contribute to return dynamics at different points in time.

Table 3: Sparsity in Structural Breaks and Business Cycles

This table reports $CSR^2$, Sharpe ratio (SR), and posterior sparsity $q_\beta$ and the posterior number of selected characteristics $\widehat{M}_\beta$ for the P-Tree 100 test assets. To accommodate the limited sample, a three-factor specification is used throughout. Panel A uses sequential segmentation with breakpoints following Smith and Timmermann (2021). Panel B focuses on macro-driven recession periods (88 months) identified by the Sahm Rule. Panel C covers the full sample from January 1990 to December 2024.

| | $CSR^2$ | SR | $q_\beta$ | $\widehat{M}_\beta$ |
|---|---|---|---|---|
| *Panel A: Sequential Segmentation* | | | | |
| Regime1 | 52.9 | 1.40 | 0.53 | 9 |
| Regime2 | 36.6 | 0.74 | 0.53 | 9 |
| Regime3 | 68.9 | 0.53 | 0.50 | 10 |
| *Panel B: Macro-driven Segmentation* | | | | |
| Normal | 61.7 | 0.83 | 0.49 | 10 |
| Recession | 21.9 | 1.01 | 0.56 | 8 |
| *Panel C: Full Period* | | | | |
| Whole | 52.1 | 0.73 | 0.46 | 11 |

Focusing on the three structural regimes identified by Smith and Timmermann (2021), Panel A of Table 3 shows that the explanatory power of characteristics for betas

varies across regimes: $CSR^2$ declines from 52.9% to 36.6%, before rising to 68.9% in Regime 3. [10] $\widehat{M}_\beta$ is roughly stable at 9 in the first two regimes and modestly increases to 10 in the recent regime, indicating that factor loadings now depend on a richer set of characteristics. These patterns indicate that the magnitude and dimensionality of priced risk are time-varying rather than static.

Second, we further examine how sparsity evolves across business cycles using the real-time Sahm Rule Recession Indicator. Panel B of Table 3 reveals a pronounced deterioration in model performance during recessions. $CSR^2$ drops sharply from 61.7% in normal periods to just 21.9%, reflecting a contraction in return predictability under macroeconomic stress. Meanwhile, the posterior sparsity probability significantly rises from 0.49 in normal periods to 0.56 during recessions. [11]

During recessions, the model relies on fewer characteristics while exhibiting weaker cross-sectional explanatory power, potentially because heightened volatility reduces the signal-to-noise ratio in returns, making it more difficult to extract stable pricing relations. Such dimensionality compression during recessions suggests that, under heightened stress, investors shift their focus away from complex, idiosyncratic signals to concentrate on a narrow set of systemic risk factors, such as solvency and liquidity. In other words, recessions diminish the explanatory power of fundamentals for factor loadings.

Several mechanisms may explain this pattern. (i) Heightened uncertainty can shift investor focus from firm-specific variables to broader macro-level risks, reducing the relevance of cross-sectional predictors (Avramov et al., 2013). (ii) During downturns, increased risk aversion may lead to a reallocation of capital toward safer assets (Baker and Wurgler, 2007) or amplify demand for downside protection (Avramov et al., 2013), thereby weakening firm-level predictors. (iii) Large macroeconomic shocks can dis-

[10] The decline in Regime 2 may be attributable to its coverage of the aftermath of the 1997 Asian Financial Crisis and the entirety of the 2008 Global Financial Crisis, two episodes of severe macroeconomic stress. We later examine more systematically how stress episodes affect the cross-sectional explanatory power of the model.

[11] We provide formal distributional comparisons in Table A.2, indicating that the recession and non-recession posteriors for $q_\beta$ are unlikely to arise from the same underlying distribution.

rupt the firm-level return-generation process, undermining predictability during crises (Chordia and Shivakumar, 2002). [12]

### 4.3 Conditional Model for Observable Factors and Sparsity

Thus far, we have estimated separate models for each test asset set and obtained the corresponding posterior sparsity probabilities for factor loadings. In the baseline specifications, latent factors are estimated separately for each asset universe and are therefore tailored to the structure of the corresponding test assets.

In this section, we ask whether the Schrödinger's sparsity patterns documented earlier are specific to the dataset-dependent latent factors, or instead reflect more general features of the pricing problem. To address this question, we replace asset-specific latent factors with common, pre-specified factors that are shared across test-asset sets, while continuing to learn sparsity endogenously through the Bayesian prior. This design makes sparsity measures directly comparable across asset universes and yields a more transparent interpretation of how sparsity varies with pricing difficulty rather than with factor construction.

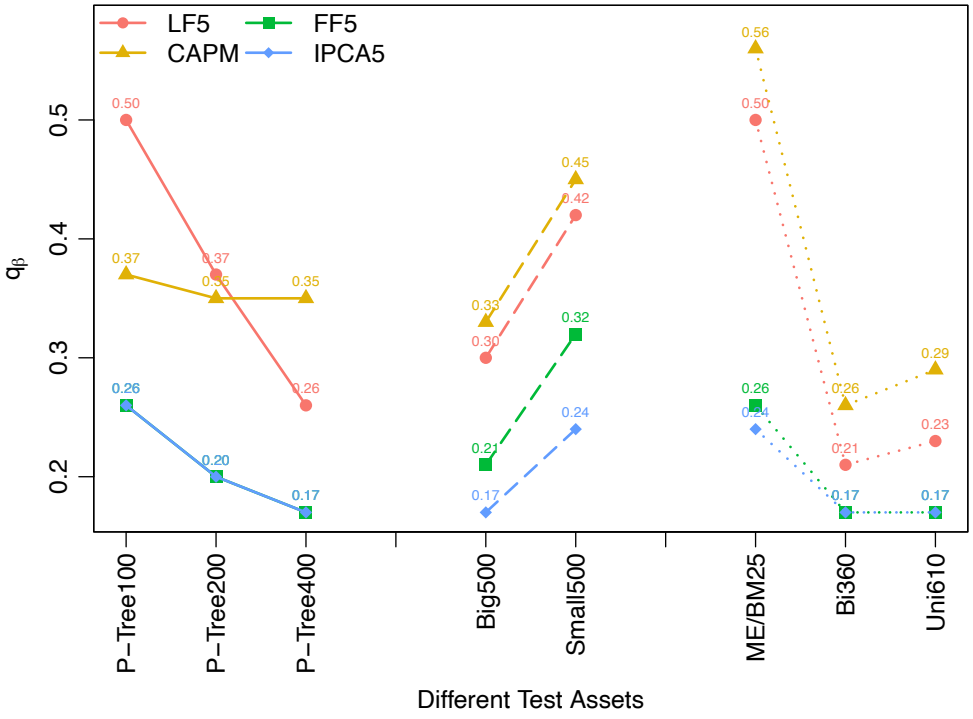
We consider two classes of pre-specified factors and treat them as observed: (i) benchmark models, including the CAPM and the Fama-French five factors (FF5); and (ii) a five-factor specification extracted via IPCA from individual stock returns (IPCA5). Figure 3 reports posterior means of $q_\beta$ across test-asset sets for the baseline five-factor latent specification, as well as for the CAPM, FF5, and IPCA5 variants. The corresponding $CSR^2$ and related performance metrics are reported in Table A.3.

We first ask whether the Schrödinger's sparsity patterns survive when the factor space is held constant across datasets. Figure 3 indicate that they largely do. Within each family of test assets, the posterior mean of $q_\beta$ declines as the test-asset universe becomes richer. For P-Tree portfolios, $q_\beta$ falls from 0.26 to 0.17 when the number of test assets increases from 100 to 400 under IPCA5, and a similar monotonic decline appears

[12] To unpack the economic drivers of time-varying betas, our Internet Appendix examines the characteristic set conditioning on structural regimes and business cycles. The analysis reveals that while factor loadings are driven by a stable core of characteristics, others exhibit significant state dependence: some fundamental signals lose explanatory power during recessionary episodes.

Figure 3: Sparsity across Factor Specifications

This figure plots $q_\beta$ estimates for eight test asset sets under four factor specifications: LF5 (red dots), CAPM (yellow triangles), FF5 (green squares), and IPCA5 (blue diamonds). The $x$-axis displays P-Tree portfolios, individual-stock universes (Big500 and Small500), and characteristic-sorted portfolios (ME/BM25, Bi360, and Uni610), connected by solid, dashed, and dotted lines, respectively. For P-Tree assets, the FF5 and IPCA5 estimates nearly coincide, rendering only the blue markers and lines visible.




under other factor specifications. The same pattern holds for individual stocks and for the three commonly used portfolio sets, where ME/BM25 exhibits much higher sparsity than Bi360 or Uni610. These results confirm that our core patterns are not artefacts of dataset-specific latent factors, but are robust to the imposition of a common factor structure exogenously.

Across factor specifications, two additional regularities emerge. First, conditional on a given set of test assets, the latent factor model (LF5) typically resides on the "efficiency frontier" of parsimony and fit: it delivers the strongest or near-strongest $CSR^2$ while preserving relatively high posterior sparsity in factor loadings. For example, for the P-Tree200 portfolios, LF5 attains a $CSR^2$ of 69.4% with $q_\beta = 0.37$, whereas FF5 and IPCA5 yield lower $CSR^2$ (61.2% and 46.2%) and much denser loadings ($q_\beta = 0.20$).

A similar pattern holds for the Big500 and Small500 stocks, where FF5 and IPCA5 both underperform LF5 in $CSR^2$ even though they rely on lower $q_\beta$. This suggests "inefficient density", where pre-specified factors are misaligned with the process that generates returns. The model compensates by activating more characteristic-driven loadings, but the resulting added characteristic does not improve pricing performance.

Second, pre-specified factor models differ substantially in how they strike a balance between fit and sparsity. As shown in Table A.3, the CAPM often produces the high posterior $q_\beta$ but low $CSR^2$ for granular test assets such as Big500, reflecting an extremely restrictive one-factor space in which the posterior finds little reliable signal and shrinks most characteristic-based betas toward zero. Moving to FF5 or IPCA5 introduces additional systematic risk dimensions and lowers $q_\beta$, yet for several datasets (notably Bi360 and Uni610) the resulting denser loading structures still deliver weak $CSR^2$. Within the family of pre-specified factor models, this pattern of very sparse but misspecified CAPM versus inefficiently dense FF5/IPCA5 reinforces the conclusion that factor design, not just the level of sparsity, is critical for pricing performance.

These evidences point to a disciplined interaction between factor design and learned sparsity. When the factor space is flexible and well-aligned with the data, as in the latent factor specification, our Bayesian procedure learns an interior sparsity level that jointly delivers a strong cross-sectional fit and parsimonious loading structures. When the factor space is too coarse or conditionally misaligned with certain test-asset environments, the posterior either contracts sharply toward sparse loadings or exhibits inefficient density. In all cases, sparsity remains an object of inference rather than a maintained assumption, and the qualitative Schrödinger's sparsity patterns documented for latent factors re-emerge under common, pre-specified factor structures.

## 5 Learning Sparsity with Mispricing

The previous section imposed the strict restriction that conditional expected returns are fully spanned by characteristic-driven factor exposures. Consequently, all predictable variation in expected returns must operate through time-varying factor

loadings and the associated factor premia.

Allowing for mispricing shifts the framework from a pure factor-pricing restriction toward a more general data-generating representation. Equity characteristics may enter expected returns both indirectly through factor loadings and directly through a mispricing component orthogonal to factor exposures. Expected returns are thus driven jointly by common variation and characteristic-based terms, so the model delivers a predictive decomposition of returns rather than a purely factor-based explanation.

In this section, we study sparsity in the presence of a mispricing term and evaluate the model through the lens of prediction and data generation. The focus shifts from assessing the latent factors in isolation to asking how well the model forecasts expected returns when both $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ are governed by a learnable sparsity structure.

Specifically, Section 5.1 evaluates the magnitude of the estimated mispricing component. Section 5.2 then assesses out-of-sample performance of portfolios constructed from the model's conditional mean forecasts. Finally, Section 5.3 provides complementary diagnostics that separate the factor-loading channel from the mispricing channel using (i) an adjusted cross-sectional $R^2$ based on traded factor proxies and (ii) two implementable alpha portfolios built directly from the estimated mispricing signal.

### 5.1 Magnitude of Mispricing

Since we now allow for a characteristic-driven mispricing component, we begin by quantifying its magnitude. In our framework, mispricing is conditional and varies across assets and time through the function $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1})$, so its economic importance cannot be assessed from individual coefficients alone. We therefore construct aggregated measures of the mispricing component and evaluate them using the full posterior distribution. Taking advantage of the Bayesian framework, we approximate the posterior of each summary statistic by mapping the sequence of parameter draws from the Gibbs sampler into the corresponding metric. This design allows us to quantify not only the level of mispricing but also the uncertainty surrounding its magnitude in a

fully probabilistic manner.

Let $g = 1, \ldots, G$ index the posterior draws from the MCMC sampler. For each posterior draw $g$, the conditional mispricing term is given by

$$\alpha_{i,t}^{(g)} = \alpha_0^{(g)} + \boldsymbol{\alpha}_1^{(g)\top} \mathbf{Z}_{i,t-1}.$$

Accordingly, we consider two objects of interest: (i) the scale of the mispricing-coefficient vector $\Gamma_\alpha = [\alpha_0,\ \boldsymbol{\alpha}_1]$ and (ii) the scale of the implied mispricing $\alpha_{i,t}$. For each draw $g$, we compute the corresponding value of these summary statistics, and the collection of $\{\cdot^{(g)}\}_{g=1}^{G}$ naturally yields their posterior distributions.

The first metric is based on Kelly et al. (2019). We measure the size of the mispricing coefficients $\Gamma_\alpha = [\alpha_0,\ \boldsymbol{\alpha}_1]$ by computing, for each posterior draw $g$,

$$\widehat{W}_\alpha^{(g)} = \widehat{\Gamma}_\alpha^{(g)\top} \widehat{\Gamma}_\alpha^{(g)}.$$

This scalar represents the squared norm of the coefficient vector, summarizing the extent to which characteristics contribute to mispricing. Since all characteristics are standardized, $W_\alpha$ is proportional to the variance of the characteristic-driven part of mispricing across assets. A larger value, therefore, indicates a larger contribution of characteristics to mispricing.

The second metric directly examines the scale of conditional mispricing. For each MCMC draw $g$, we compute

$$\widehat{\alpha}_{it}^{(g)} = \widehat{\alpha}_0^{(g)} + \widehat{\boldsymbol{\alpha}}_1^{(g)\top} \mathbf{Z}_{i,t-1},$$

and summarize aggregate mispricing as

$$\text{RMSE}(\widehat{\boldsymbol{\alpha}}^{(g)}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \widehat{\alpha}_{it}^{(g)} \right)^2}.$$

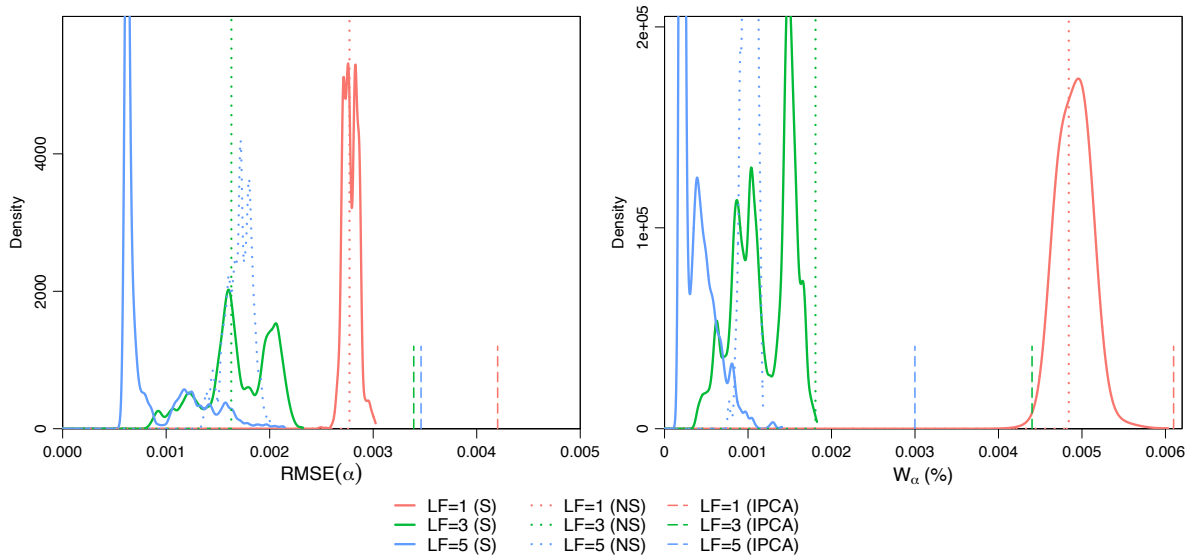
This measure is the cross-sectional root mean square of each asset's average mispric-

ing, capturing the overall magnitude of pricing errors in economic terms.

Figure 4 reports the posterior distributions of the two metrics for models with different numbers of latent factors. As a frequentist method, IPCA yields only point estimates, which appear as short vertical lines in the figure without associated uncertainty. Furthermore, for comparison, we include a Bayesian factor model without sparsity priors but just normal priors, which is dense.

Figure 4: Posterior Mispricing for Various Models

This figure displays posterior distributions (or point estimates) for the mispricing RMSE($\boldsymbol{\alpha}$) (Panel a) and the coefficient norm $W_\alpha$ (Panel b). Colors denote models with 1 (red), 3 (green), and 5 (blue) latent factors. Line styles indicate the model specification: sparsity-aware (S, solid), non-sparse (NS, dotted), and IPCA (dashed). Short vertical lines mark IPCA point estimates. The $y$-axis is truncated to better compare the distribution shapes across models.




(a) Density / value of RMSE($\boldsymbol{\alpha}$)

(b) Density / value of $W_\alpha(\%)$

Three key patterns emerge. First, across specifications and factor counts, the posterior distributions of the two mispricing metrics are distinct from zero; the high-density posterior intervals for both indicators exclude zero.

Second, holding the number of factors fixed, the sparsity-learning Bayesian specification (S) dominates the dense Bayesian benchmark (NS). Its posterior distributions for both metrics shift markedly to the left, with thinner tails and little overlap, indicating that selective shrinkage reduces the strength of mispricing rather than just tightening confidence bands. The IPCA point estimates lie even further to the right,

often outside the central mass of NS, which is consistent with larger residual mispricing when model uncertainty is ignored, and no targeted shrinkage is applied.

Third, increasing the number of latent factors pushes the posterior distributions of both metrics closer to zero, making them more concentrated. Additional latent factors absorb more systematic variation, and the spike and slab prior filters out weak signal characteristics. Mispricing declines through the combined effect of richer factor structure and adaptive sparsity.

The evidence supports the concept of Schrödinger's sparsity. The mispricing channel is neither shut down nor left unconstrained; instead, it is disciplined by posterior learning toward an interior region. As model capacity increases, adaptive shrinkage and factor absorption work in tandem to compress mispricing to economically small magnitudes. By contrast, ignoring sparsity or ignoring model uncertainty weakens this discipline and yields systematically larger coefficient norms and mispricing.

### 5.2 Investment Performance

Interpreting the model with mispricing as a data-generating process, we evaluate out-of-sample performance of the full model – rather than the Sharpe ratios of the latent factors – because the factors are no longer required to span the test assets. We therefore construct portfolios based on the model's return forecasts and assess their performance under three weighting schemes. For the equal-weighted and value-weighted strategies, we form long and short positions according to the sign of the model-implied expected returns and allocate using absolute weights. We also form forecast-weighted portfolios that scale positions directly with the model's predicted returns, subject to a leverage constraint.

Specifically, let the portfolio weights $w_{i,t}$ be equal-weighted or value-weighted, where $i$ denotes the stock dimension and $t$ represents the time period. These three

portfolio weighting schemes are as follows:

$$\text{Sign-adjusted Equal/Value-weighted: } \widehat{w}_{i,t} = \begin{cases} \frac{w_{i,t}}{\sum_j w_{j,t}\mathbb{I}_{(\mathbb{E}[r_{j,t}|\mathbf{Z}_{j,t-1}]\geq 0)}}, & \text{if } \mathbb{E}[r_{i,t} \mid \mathbf{Z}_{i,t-1}] \geq 0, \\ \frac{-w_{i,t}}{\sum_j w_{j,t}\mathbb{I}_{(\mathbb{E}[r_{j,t}|\mathbf{Z}_{j,t-1}]<0)}}, & \text{if } \mathbb{E}[r_{i,t} \mid \mathbf{Z}_{i,t-1}] < 0. \end{cases}$$

$$\text{Forecast-weighted: } \widehat{w}_{i,t} = \frac{\mathbb{E}\left[r_{i,t} \mid \mathbf{Z}_{i,t-1}\right]}{\sum_i \left|\mathbb{E}\left[r_{i,t} \mid \mathbf{Z}_{i,t-1}\right]\right|}.$$

Thus, we define the forecast-implied portfolio as $R_{i,t} = \sum_i \widehat{w}_{i,t} r_{i,t}$. As for $\mathbb{E}\left[r_{i,t} \mid \mathbf{Z}_{i,t-1}\right]$, we compute it using the following procedure:

$$\mathbb{E}\left[r_{i,t} \mid \mathbf{Z}_{i,t-1}\right] = \widehat{\alpha}_0 + \widehat{\boldsymbol{\alpha}}_1^\top \mathbf{Z}_{i,t-1} + \widehat{\boldsymbol{\beta}}_0^\top \mathbb{E}_{t-1}\left[\mathbf{f}_t\right] + \widehat{\boldsymbol{\beta}}_1^\top \left[\mathbb{E}_{t-1}\left[\mathbf{f}_t\right] \otimes \mathbf{Z}_{i,t-1}\right], \tag{13}$$

where $\widehat{(\cdot)}$ denotes the in-sample parameter estimates, and $\mathbb{E}_{t-1}\left[\mathbf{f}_t\right]$ is estimated using the sample mean of the factors from periods 1 through $t-1$.

Before turning to the out-of-sample forecasting and portfolio performance results, we first examine how the introduction of characteristic-driven mispricing reshapes the posterior sparsity structure. Panel A of Table 4 shows that the mispricing and factor loading channels exhibit clearly differentiated structural roles. The posterior means of $q_\alpha$ are systematically above 0.5 and in several specifications approach 0.8, whereas $q_\beta$ remains consistently below 0.5. This pattern reflects the interaction between the two channels: once a characteristic plays a meaningful role in shaping the factor structure through time-varying betas, its marginal contribution to the remaining pricing error becomes much weaker. The finding that alpha is significantly sparser than beta suggests that the factor structure efficiently absorbs most characteristic-driven variation, leaving only a highly selective, small set of characteristics to drive the orthogonal mispricing component.

We next examine when both characteristic-driven alphas and time-varying betas are used to form conditional mean forecasts, which modeling choice delivers the strongest out-of-sample performance? Across all portfolio weighting schemes in Ta-

Table 4: Out-of-sample Forecast-Implied Investment Performance for Various Models

This table reports the out-of-sample annualized Sharpe ratios for three forecast-implied strategies based on Equation (13): sign-adjusted value-weighted (Sign-adj. VW), sign-adjusted equal-weighted (Sign-adj. EW), and forecast-weighted (Forecast-W). $K$ denotes the number of latent factors. Panel A features the "learning-sparsity" setting with Beta$(5,5)$ priors for both $q_\alpha$ and $q_\beta$, including their posterior estimates and selected characteristic counts ($\widehat{M}_\alpha$, $\widehat{M}_\beta$). Note that $\widehat{M}_\alpha = 0$ indicates that no characteristic's posterior inclusion probability exceeds 0.5. Panel B imposes fixed restrictions on the number of characteristics affecting $\boldsymbol{\alpha}_1$ ($M_\alpha$) and each factor loading $\boldsymbol{\beta}_{1,k}$ ($M_\beta$). Panels C and D present results for the non-sparse Bayesian and standard dense IPCA benchmarks, respectively. The sample comprises a 15-year in-sample estimation period, followed by a 20-year out-of-sample evaluation period.

| | | Out-of-sample Sharpe Ratio | | | | | |
|---|---|---|---|---|---|---|---|
| | | Sign-adj. VW | | Sign-adj. EW | | Forecast-W | |
| | | $K=1$ | $K=5$ | $K=1$ | $K=5$ | $K=1$ | $K=5$ |
| *Panel A: Learned Sparsity* | | | | | | | |
| $(q_\alpha, q_\beta)$ prior mean | 0.5,0.5 | 0.59 | 0.83 | 0.42 | 0.76 | 0.46 | 0.78 |
| | | In-sample estimates of $q_\alpha$, $q_\beta$ and $\widehat{M}_\alpha, \widehat{M}_\beta$ | | | | | |
| | | $K=1$: (0.57,0.44) and (7,11); | | | | | |
| | | $K=5$: (0.80,0.44) and (0,12) | | | | | |
| *Panel B: Fixed Sparsity* | | | | | | | |
| | 2,2 | 0.59 | 0.63 | 0.43 | 0.43 | 0.42 | 0.48 |
| | 10,2 | 0.68 | 0.73 | 0.48 | 0.55 | 0.50 | 0.56 |
| | 18,2 | 0.70 | 0.71 | 0.52 | 0.57 | 0.53 | 0.59 |
| | 2,10 | 0.58 | -0.61 | 0.42 | -0.42 | 0.40 | -0.39 |
| $(M_\alpha, M_\beta)$ | 10,10 | 0.63 | 0.61 | 0.43 | 0.42 | 0.47 | 0.39 |
| | 18,10 | 0.68 | 0.61 | 0.49 | 0.42 | 0.51 | 0.38 |
| | 2,18 | 0.57 | -0.50 | 0.43 | -0.37 | 0.40 | -0.31 |
| | 10,18 | 0.65 | 0.61 | 0.44 | 0.42 | 0.47 | 0.47 |
| | 18,18 | 0.68 | 0.16 | 0.49 | -0.12 | 0.51 | -0.14 |
| *Panel C: No Sparsity* | | | | | | | |
| $(M_\alpha, M_\beta)$ | 20 | 0.67 | 0.74 | 0.46 | 0.51 | 0.51 | 0.54 |
| *Panel D: IPCA* | | | | | | | |
| $(M_\alpha, M_\beta)$ | 20 | 0.66 | 0.74 | 0.52 | 0.56 | 0.55 | 0.57 |

ble 4, models that learn sparsity in both alpha and beta consistently outperform dense and fixed sparsity specifications out-of-sample. With Beta$(5,5)$ priors on $(q_\alpha, q_\beta)$ and five latent factors, the learning sparsity model delivers annualized Sharpe ratios of 0.83, 0.76, and 0.78, showing that it extracts a forecasting signal that remains effective out-of-sample.

Panel B of Table 4 clarifies that these gains do not arise from any particular fixed sparsity choice. Fixed sparsity specifications display substantial performance disper-

sion across different $(M_\alpha, M_\beta)$ configurations, with Sharpe ratios varying sharply even for modest changes in thresholds. While a small subset of configurations yields moderately positive returns, their performance remains well below that of the learning sparsity benchmark and is highly sensitive to the imposed structure.

More importantly, when $K = 5$, no fixed sparsity level consistently approximates the performance of the learning specification across portfolio constructions or factor dimensions. Several fixed configurations generate sharply deteriorated or even negative Sharpe ratios, highlighting the fragility of ex ante sparsity restrictions. These results suggest that sparsity in conditional asset pricing is not adequately characterized by a single fixed level, but rather must be inferred from the data to remain both economically meaningful and empirically stable.

The dense Bayesian and IPCA specifications with $K = 5$ in Panels C and D of Table 4 achieve Sharpe ratios in the 0.50 to 0.75 range, but they fall short of the probabilistic sparsity model. The ranking is robust across sign-adjusted and forecast-weighted portfolios. Sign-based strategies are somewhat more stable when sparsity is misspecified, suggesting that most economic value comes from correctly identifying winners and losers rather than from precisely estimating return magnitudes.

Learning sparsity in $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ has clear and sizable effects on investment performance. Their forecast-based portfolios deliver much higher Sharpe ratios than fixed sparsity specifications, dense Bayesian models, or IPCA. This result reinforces the importance of estimating sparsity from the data rather than imposing it directly. [13]

### 5.3 Decomposing Risk and Mispricing

While the Sharpe ratios of the forecast-implied portfolios provide a direct measure of the economic value of our return predictions, they combine the systematic component and mispricing jointly. To separate these two, we supplement the portfolio results with model-based diagnostics that identify the contributions of the factor-loading

[13] The drawdown profiles in our Internet Appendix follow the same pattern. Learning sparsity models achieve the highest Sharpe ratios with moderate drawdowns, while rigid or misspecified sparsity levels lead to weaker performance and larger tail losses.

channel and the mispricing channel. We construct (i) an adjusted cross-sectional $R^2$ that attributes variation in average returns solely to the factor-loading channel, and (ii) two alpha portfolios that isolate the traded component of characteristic-driven mispricing. All results are evaluated out-of-sample.

As discussed above, once mispricing is allowed, factor premia no longer span expected returns. The conditional expectation of the latent factor is

$$\mathbb{E}\left[\mathbf{f}_t \mid \boldsymbol{\alpha}(\mathbf{Z}_{t-1}), \boldsymbol{\beta}(\mathbf{Z}_{t-1}), \boldsymbol{\Sigma}, \mathbf{r}_t\right] = \boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top \left[\boldsymbol{\beta}(\mathbf{Z}_{t-1})\boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top + \boldsymbol{\Sigma}\right]^{-1} \left[\mathbf{r}_t - \boldsymbol{\alpha}(\mathbf{Z}_{t-1})\right],$$

which shows that the latent factors are not directly traded, since they are constructed from returns net of the mispricing component.

Following Kelly et al. (2019) and Chen et al. (2025), we therefore construct a traded (realized) factor proxy by projecting raw excess returns onto the factor space,

$$\mathbb{E}\left[\mathbf{f}_t^{\text{traded}} \mid \boldsymbol{\alpha}(\mathbf{Z}_{t-1}), \boldsymbol{\beta}(\mathbf{Z}_{t-1}), \boldsymbol{\Sigma}, \mathbf{r}_t\right] = \boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top \left[\boldsymbol{\beta}(\mathbf{Z}_{t-1})\boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top + \boldsymbol{\Sigma}\right]^{-1} \mathbf{r}_t.$$

Therefore, the fitted return obtained from the factor-loading channel ($\tilde{r}_{i,t}$) is

$$\widetilde{r}_{i,t} = \widehat{\boldsymbol{\beta}}_0^\top \mathbf{f}_t^{\text{traded}} + \widehat{\boldsymbol{\beta}}_1^\top \left(\mathbf{f}_t^{\text{traded}} \otimes \mathbf{Z}_{i,t-1}\right),$$

and we compute the adjusted cross-sectional $R^2$,

$$\text{CSR}^2_{\text{adj}} = 1 - \frac{\sum_{i=1}^{N} \left(\frac{1}{T_i}\sum_{t=1}^{T_i}(r_{i,t} - \widetilde{r}_{i,t})\right)^2}{\sum_{i=1}^{N} \left(\frac{1}{T_i}\sum_{t=1}^{T_i}(r_{i,t} - \widehat{\beta}_i \,\text{MktRF}_t)\right)^2}, \tag{14}$$

which attributes cross-sectional variation in average returns to the fitted factor-loading component.

We also construct two alpha strategies that load only on the estimated mispricing component. Let $\widehat{\Gamma}_\alpha = [\widehat{\alpha}_0, \widehat{\boldsymbol{\alpha}}_1]$ collect the mispricing coefficients, and define $\widetilde{\mathbf{Z}}_{t-1} = [\mathbf{1}, \mathbf{Z}_{t-1}]$ to incorporate the common mispricing term $\widehat{\alpha}_0$. First, the pure-alpha strategy

projects $\widehat{\Gamma}_\alpha$ onto the characteristic space to obtain signal weights,

$$\boldsymbol{w}_{t-1}^{\text{PA}} = \widetilde{\mathbf{Z}}_{t-1}\big(\widetilde{\mathbf{Z}}_{t-1}^\top \widetilde{\mathbf{Z}}_{t-1}\big)^{-1}\widehat{\Gamma}_\alpha,$$

and applies these weights to the return residuals that remain after removing traded factor exposures,

$$R_t^{\text{PA}} = (\boldsymbol{w}_{t-1}^{\text{PA}})^\top(\mathbf{r}_t - \tilde{\mathbf{r}}_t), \tag{15}$$

where $\tilde{\mathbf{r}}_t \equiv (\tilde{r}_{1,t}, \ldots, \tilde{r}_{N_t,t})^\top$. Unlike Kelly et al. (2019), this strategy does not invest in returns; instead, it isolates the mispricing by hedging out the factor-driven component.

Second, we construct an alpha long-short portfolio by replacing the portfolio weights in Equation (15) with

$$\boldsymbol{w}_{t-1}^{\text{LS}} = c_{t-1}\left(\widetilde{\mathbf{Z}}_{t-1}\widehat{\Gamma}_\alpha - \overline{\widetilde{\mathbf{Z}}_{t-1}\widehat{\Gamma}_\alpha}\mathbf{1}\right),$$

where $\overline{\widetilde{\mathbf{Z}}_{t-1}\widehat{\Gamma}_\alpha}$ is the cross-sectional mean of the signal, and $c_{t-1}$ is used to ensure the gross leverage equals one.

Across specifications, with $K = 5$, the learning-sparsity model attains out-of-sample $\text{CSR}^2_{\text{adj}} = 69.3\%$, which exceeds all fixed sparsity configurations (typically in the 63-67% range for $K = 5$). Hence, letting the model determine its effective sparsity does not sacrifice explanatory power for parsimony.

The alpha portfolio results demonstrate the extent to which additional traded structure is recovered when sparsity is learned rather than imposed. Under learning sparsity with $K = 5$, the pure-alpha strategy delivers a Sharpe ratio of 0.86, and the alpha long-short strategy reaches 1.04. Fixed-sparsity models can occasionally match one of these values. For example, $(M_\alpha, M_\beta) = (10, 2)$ or $(10, 10)$ yield long-short Sharpe ratios around 0.88-1.00, but only at the cost of substantially lower $\text{CSR}^2_{\text{adj}}$. Dense alternatives perform even worse on the alpha dimension: the no-sparsity Bayesian model delivers a pure-alpha Sharpe ratio of 0.30 and an alpha long-short Sharpe ratio of 0.67, while IPCA reaches 0.43 and 0.77, respectively, despite achieving similar $\text{CSR}^2_{\text{adj}}$.

Table 5: Out-of-sample Performance for Various Models with Mispricing

This table reports out-of-sample results for P-Tree 100 test assets, including the adjusted cross-sectional $R^2$ ($\text{CSR}^2_{\text{adj}}$), the pure-alpha strategy Sharpe ratio (Pure-alpha SR), and the long-short alpha strategy Sharpe ratio (Alpha LS SR). $K$ denotes the number of latent factors. Panel A features "learning-sparsity" settings where $q_\alpha$ and $q_\beta$ both follow $\text{Beta}(5,5)$ priors. Panel B imposes fixed sparsity by limiting the number of characteristics influencing $\boldsymbol{\alpha}_1$ ($M_\alpha$) and each factor loading $\boldsymbol{\beta}_{1,k}$ ($M_\beta$). Panels C and D provide benchmarks from the non-sparse Bayesian framework and standard dense IPCA, respectively. The sample is divided into a 15-year in-sample estimation period and a 20-year out-of-sample evaluation period.

| | | $\text{CSR}^2_{\text{adj}}$ | | Pure-alpha SR | | Alpha LS SR | |
|---|---|---|---|---|---|---|---|
| | | $K=1$ | $K=5$ | $K=1$ | $K=5$ | $K=1$ | $K=5$ |
| *Panel A: Learned Sparsity* | | | | | | | |
| $(q_\alpha, q_\beta)$ prior mean | 0.5,0.5 | 13.3 | 69.3 | 0.50 | 0.86 | 0.81 | 1.04 |
| *Panel B: Fixed Sparsity* | | | | | | | |
| | 2,2 | 13.3 | 63.7 | 0.04 | 0.46 | 0.55 | 0.48 |
| | 10,2 | 13.6 | 63.7 | 0.55 | 0.86 | 0.92 | 0.88 |
| | 18,2 | 13.9 | 63.7 | 0.53 | 0.64 | 1.00 | 0.75 |
| | 2,10 | 13.2 | 66.6 | 0.04 | 0.54 | 0.54 | 0.93 |
| $(M_\alpha, M_\beta)$ | 10,10 | 12.8 | 63.9 | 0.54 | 0.64 | 0.85 | 1.00 |
| | 18,10 | 13.4 | 65.9 | 0.54 | 0.40 | 0.97 | 0.79 |
| | 2,18 | 13.4 | 65.8 | 0.01 | 0.10 | 0.53 | 0.44 |
| | 10,18 | 13.9 | 67.3 | 0.54 | 0.56 | 0.85 | 0.92 |
| | 18,18 | 12.9 | 65.0 | 0.54 | 0.56 | 0.97 | 0.90 |
| *Panel C: No Sparsity* | | | | | | | |
| $(M_\alpha, M_\beta)$ | 20 | 12.2 | 68.4 | 0.49 | 0.30 | 0.87 | 0.67 |
| *Panel D: IPCA* | | | | | | | |
| $(M_\alpha, M_\beta)$ | 20 | 16.4 | 69.3 | 0.67 | 0.43 | 0.81 | 0.77 |

Learning sparsity, therefore, produces a sharper and more exploitable mispricing signal than either fixed-sparsity or dense benchmarks.

These diagnostics show that learning sparsity jointly in alpha and beta preserves the factor-based fit of expected returns while sharply strengthening the mispricing component. The posterior settles on an interior decomposition of the model-implied mean return into a systematic component and a mispricing component, creating a selective alpha channel and an informative beta structure that together deliver stronger cross-sectional explanatory power and more robust alpha strategies.

Our results reveal an economically meaningful "division of labor" between the two primary channels of return predictability: while factor loadings exhibit moderate

density to capture systematic risk, mispricing coefficients are systematically sparser. Critically, this dual-channel framework ensures that the dense systematic risk component is not forced into an overly parsimonious structure, which would otherwise result in significant pricing errors and distorted factor exposures. This selective alpha channel suggests that once a characteristic contributes to the common factor structure, its role in explaining residual mispricing diminishes. This refinement enables the model to extract a more precise and actionable mispricing signal compared to traditional dense or fixed-sparsity benchmarks.

## 6 Conclusion

This paper offers a probabilistic perspective on sparsity in asset pricing. We develop a Bayesian framework that treats sparsity as a latent feature learned from the data, bypassing the traditional binary of strictly sparse or fully dense specifications. Within a conditional factor structure, hierarchical spike-and-slab priors enable the data to determine which equity characteristics are relevant for factor exposures and which are relevant for mispricing.

Our results support an interior view of sparsity ("Schrödinger's sparsity"). Factor loadings exhibit moderate density, as a broad characteristic set contributes to systematic risk. At the same time, fully dense representations are typically unnecessary. Endogenous sparsity models outperform dense and fixed-sparsity benchmarks, suggesting that optimal asset pricing specifications occupy an interior region between these extremes.

Allowing for characteristic-driven mispricing further sharpens this conclusion. When mispricing is excluded, factor loadings are forced to absorb residual structure that is not captured by the factor component alone. Admitting mispricing facilitates a stable division of labor: factor exposures capture common variation through a rich characteristic structure, while a selective set of characteristics governs mispricing. This reallocation improves return forecasts and translates into stronger out-of-sample investment performance.

These conclusions are robust to the factor specification. Similar sparsity patterns arise when latent factors are replaced by observable benchmarks, and augmenting observable factor models with latent components further improves fit. Our evidence suggests that sparsity is neither illusory nor absolute, but a dynamic empirical object, varying across assets and regimes, that is best learned endogenously rather than imposed ex ante. Ultimately, this framework advances the field by refining the identification of latent risk factors and their time-varying exposures within a high-dimensional conditional modeling framework, providing a robust path forward for navigating model uncertainty in cross-sectional returns.

## References


Ang, A. and D. Kristensen (2012). Testing conditional factor models. *Journal of Financial Economics 106*(1), 132–156.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics 64*(3), 423–458.

Avramov, D. and T. Chordia (2006a). Asset pricing models and financial market anomalies. *Review of Financial Studies 19*(3), 1001–1040.

Avramov, D. and T. Chordia (2006b). Predicting stock returns. *Journal of Financial Economics 82*(2), 387–415.

Avramov, D., T. Chordia, G. Jostova, and A. Philipov (2013). Anomalies and financial distress. *Journal of Financial Economics 108*(1), 139–159.

Baker, M. and J. Wurgler (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives 21*(2), 129–151.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *Journal of Finance 73*(2), 715–754.

Barron, A., J. Rissanen, and B. Yu (1998). The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory 44*(6), 2743–2760.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Bryzgalova, S., J. Huang, and C. Julliard (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Journal of Finance 78*(1), 487–557.

Bybee, L., B. Kelly, and Y. Su (2023). Narrative asset pricing: Interpretable systematic risk factors from news text. *Review of Financial Studies 36*(12), 4759–4787.

Chen, L., M. Pelger, and J. Zhu (2024). Deep learning in asset pricing. *Management Science 70*(2), 714–750.

Chen, Q., N. Roussanov, and X. Wang (2025). Semiparametric conditional factor models in asset pricing. Technical report, University of Pennsylvania.

Chib, S., X. Zeng, and L. Zhao (2020). On comparing asset pricing models. *Journal of Finance 75*(1), 551–577.

Chib, S., L. Zhao, and G. Zhou (2024). Winners from winners: A tale of risk factors. *Management Science 70*(1), 396–414.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). Sparse signals in the cross-section of returns. *Journal of Finance 74*(1), 449–492.

Chordia, T. and L. Shivakumar (2002). Momentum, business cycle, and time-varying expected returns. *Journal of Finance 57*(2), 985–1019.

Cong, L., G. Feng, J. He, and X. He (2025). Growing the efficient frontier on panel trees. *Journal of Financial Economics 167*, 104024.

Cong, L. W., G. Feng, J. He, and J. Li (2023). Sparse modeling under grouped heterogeneity with an application to asset pricing. Technical report, National Bureau of Economic Research.

Daniel, K. and S. Titman (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance 52*(1), 1–33.

Didisheim, A., S. Ke, B. Kelly, and S. Malamud (2024). APT or "AIPT"? The surprising dominance of large factor models. Technical report, National Bureau of Economic Research.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), 427–465.

Fan, J., Z. T. Ke, Y. Liao, and A. Neuhierl (2025). Structural deep learning in conditional asset pricing. Technical report, Princeton University.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *Journal of Finance 75*(3), 1327–1370.

Feng, G., J. He, N. G. Polson, and J. Xu (2024). Deep learning in characteristics-sorted factor models. *Journal of Financial and Quantitative Analysis 59*(7), 3001–3036.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies 33*(5), 2326–2377.

Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica 84*(3), 985–1046.

George, E. I. and R. E. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies 9*(2), 557–587.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Green, J., J. R. Hand, and X. F. Zhang (2017). The characteristics that provide independent information about average US monthly stock returns. *Review of Financial Studies 30*(12), 4389–4436.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies 33*(5), 2223–2273.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

He, J., L. Zhao, and G. Zhou (2024). No sparsity in asset pricing: Evidence from a generic statistical test. Technical report, Washington University in St. Louis.

Huang, D., F. Jiang, K. Li, G. Tong, and G. Zhou (2022). Scaled PCA: A new approach to dimension reduction. *Management Science 68*(3), 1678–1695.

Kelly, B., S. Malamud, and K. Zhou (2024). The virtue of complexity in return prediction. *Journal of Finance 79*(1), 459–503.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Lettau, M. and S. Ludvigson (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy 109*(6), 1238–1287.

Lettau, M. and M. Pelger (2020). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies 33*(5), 2274–2325.

Leung, G. and A. R. Barron (2006). Information theory and mixing least-squares regressions. *IEEE Transactions on Information Theory 52*(8), 3396–3410.

Li, S. Z., P. Yuan, and G. Zhou (2023). Pockets of factor pricing. Technical report, Washington University in St. Louis.

Martin, I. W. and S. Nagel (2022). Market efficiency in the age of big data. *Journal of Financial Economics 145*(1), 154–177.

Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association 83*(404), 1023–1032.

Robert, C. P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Springer.

Shen, Z. and D. Xiu (2024). Can machines learn weak signals? Technical report, University of Chicago.

Smith, S. C. and A. Timmermann (2021). Break risk. *Review of Financial Studies 34*(4), 2045–2100.

# Appendix

## I Prior setting and Gibbs sampler

Posterior inference in our Bayesian framework is conducted using MCMC methods, specifically a Gibbs sampler. To facilitate notation, conditional on the latent factors $\mathbf{f}_t$ and characteristics $\mathbf{Z}_{i,t-1}$, we express the model in Equation (2) as:

$$r_{i,t} = \mathcal{W}_{i,t}\Gamma + \epsilon_{i,t} \tag{A.1}$$

where $\mathcal{W}_{i,t} = \left[1, \mathbf{Z}_{i,t-1}^\top, \mathbf{f}_t^\top, (\mathbf{f}_t \otimes \mathbf{Z}_{i,t-1})^\top\right]$ and $\Gamma = \left[\alpha_0, \boldsymbol{\alpha}_1^\top, \boldsymbol{\beta}_0^\top, \boldsymbol{\beta}_1^\top\right]^\top$.

For simplicity, we consider $i = 1, \ldots, N$ assets, where each asset $i$ has $T_i$ observations. In matrix notation, we write $\mathbf{R}_i = \mathcal{W}_i\Gamma + \boldsymbol{\epsilon}_i$, with $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 \mathbb{I}_{T_i})$. Stacking all assets, we obtain

$$\mathbf{R} = \mathcal{W}\Gamma + \boldsymbol{\epsilon},$$

where $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N)^\top$, $\mathcal{W} = (\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_N)^\top$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_N)^\top$.

The likelihood function is given by:

$$p(\mathbf{R} \mid -) = \prod_{i=1}^{N} (2\pi\sigma_i^2)^{-\frac{T_i}{2}} \exp\left[-\frac{1}{2\sigma_i^2}(\mathbf{R}_i - \mathcal{W}_i\Gamma)^\top(\mathbf{R}_i - \mathcal{W}_i\Gamma)\right]$$

**Prior Specification.** Recall the spike-and-slab priors for mispricing and factor loadings defined in Equations (8)-(10):

$$[\boldsymbol{\alpha}_1]_l \sim \begin{cases} \mathcal{N}(0, \gamma_\alpha^2) & \text{if } d_l^\alpha = 1 \\ 0 & \text{if } d_l^\alpha = 0 \end{cases} \qquad [\boldsymbol{\beta}_1]_l \sim \begin{cases} \mathcal{N}(0, \gamma_\beta^2) & \text{if } d_l^\beta = 1 \\ 0 & \text{if } d_l^\beta = 0 \end{cases}$$

$$d_l^\alpha \sim \text{Bernoulli}(1 - q_\alpha) \qquad d_l^\beta \sim \text{Bernoulli}(1 - q_\beta)$$

$$q_\alpha \sim \text{Beta}(a_{q_\alpha}, b_{q_\alpha}) \qquad q_\beta \sim \text{Beta}(a_{q_\beta}, b_{q_\beta})$$

$$\gamma_\alpha^2 \sim \mathcal{IG}\left(A_{\gamma_\alpha}/2, B_{\gamma_\alpha}/2\right) \qquad \gamma_\beta^2 \sim \mathcal{IG}\left(A_{\gamma_\beta}/2, B_{\gamma_\beta}/2\right).$$

For the remaining parameters, we adopt standard conjugate priors: $\alpha_0, \boldsymbol{\beta}_0 \overset{iid}{\sim} \mathcal{N}\left(0, \xi^2\right)$, $\xi^2 \sim \mathcal{IG}\left(C/2, D/2\right)$, and $\sigma_i^2 \overset{iid}{\sim} \mathcal{IG}\left(v_{0,i}/2, S_{0,i}/2\right)$.

**Learning Sparsity.** As mentioned in Section 2.3, $d_l^\alpha$ and $d_l^\beta$ are binary indicators for the inclusion of the $l$-th coefficients in the mispricing and factor loading functions, respectively. Furthermore, since we impose grouped selection on $\boldsymbol{\beta}(\mathbf{Z})$, we need only select from the $L$ characteristics to determine the loadings. Let $s(d^\alpha) = \sum_l^L d_l^\alpha$ and $s(d^\beta) = \sum_l^L d_l^\beta$. Then the full conditional distributions are derived as follows:

1. Sampling $d_l^\alpha$ and $d_l^\beta$.

   Let $\widetilde{\mathcal{W}}_i$ and $\widetilde{\Gamma}$ denote the subsets of regressors and coefficients corresponding to the selected variables ($d_l = 1$).

   Marginalizing out $\widetilde{\Gamma}$, the posterior of the indicator vector $\boldsymbol{d}$ is proportional to:

$$p(\boldsymbol{d} \mid -) \propto q_\alpha^{L-s(d^\alpha)}(1-q_\alpha)^{s(d^\alpha)} q_\beta^{L-s(d^\beta)}(1-q_\beta)^{s(d^\beta)} \left(\frac{1}{\gamma_\alpha^2}\right)^{\frac{s(d^\alpha)}{2}} \left(\frac{1}{\gamma_\beta^2}\right)^{\frac{Ks(d^\beta)}{2}} \times |V_1|^{\frac{1}{2}} \left[\prod_{i=1}^N \left(\frac{1}{\sigma_i^2}\right)^{\frac{T_i+v_{0,i}}{2}+1}\right] \times \exp\left(-\sum_{i=1}^N \frac{S_i + S_{0,i}}{2\sigma_i^2} - \frac{1}{2}\sum_{i=1}^N \left(\left(\sigma_i^{-2}\widetilde{\mathcal{W}}_i^\top \mathbf{R}_i\right)^\top \widehat{\widetilde{\Gamma}}_i\right) + \frac{1}{2}\widetilde{\Gamma}_1^\top V_1^{-1}\widetilde{\Gamma}_1\right), \tag{A.2}$$

   where $\widetilde{\Gamma}_1 = V_1\left(\sum_{i=1}^N \sigma_i^{-2}\widetilde{\mathcal{W}}_i^\top \mathbf{R}_i\right)$ is the posterior mean for $\widetilde{\Gamma}$, $\widehat{\widetilde{\Gamma}}_i = (\widetilde{\mathcal{W}}_i^\top \widetilde{\mathcal{W}}_i)^{-1}\widetilde{\mathcal{W}}_i^\top \mathbf{R}_i$ is the OLS estimator for the $i$-th asset, $S_i = \left(\mathbf{R}_i - \widetilde{\mathcal{W}}_i\widehat{\widetilde{\Gamma}}_i\right)^\top \left(\mathbf{R}_i - \widetilde{\mathcal{W}}_i\widehat{\widetilde{\Gamma}}_i\right)$, and $V_1 = \left(\sum_{i=1}^N \sigma_i^{-2}\widetilde{\mathcal{W}}_i^T\widetilde{\mathcal{W}}_i + A\right)^{-1}$ and $A = \text{diag}\left(1/\xi^2, \mathbb{I}_{s(d^\alpha)}/\gamma_\alpha^2, \mathbb{I}_K/\xi^2, \mathbb{I}_{Ks(d^\beta)}/\gamma_\beta^2\right)$.

   When we need to decide whether $d_l = 1$ or $d_l = 0$, we first use Equation (A.2) to compute the respective posterior probabilities, $\pi(d_l = 1 \mid -)$ and $\pi(d_l = 0 \mid -)$. Next, we calculate the relevant probability, compare it with a random draw $U$ from a uniform distribution, and follow Equation (A.3) for the comparison logic.

$$d_l = \begin{cases} 1 & \text{if } U \leq P(d_l = 1) = \frac{\pi(d_l=1|-)}{\pi(d_l=0|-)+\pi(d_l=1|-)} \\ 0 & \text{Otherwise.} \end{cases} \tag{A.3}$$

2. Sampling $q_\alpha$ and $q_\beta$

$$q_\alpha \mid d^\alpha \sim \text{Beta}\left(a_{q_\alpha} + L - s(d^\alpha), b_{q_\alpha} + s(d^\alpha)\right), q_\beta \mid d^\beta \sim \text{Beta}\left(a_{q_\beta} + L - s(d^\beta), b_{q_\beta} + s(d^\beta)\right).$$

3. Sampling $\gamma_\alpha^2$ and $\gamma_\beta^2$

$$\gamma_\alpha^2 \mid d^\alpha, \widetilde{\boldsymbol{\alpha}}_1 \sim \mathcal{IG}\left(\frac{A_{\gamma_\alpha} + s(d^\alpha)}{2}, \frac{B_{\gamma_\alpha} + \widetilde{\boldsymbol{\alpha}}_1^\top \widetilde{\boldsymbol{\alpha}}_1}{2}\right), \gamma_\beta^2 \mid d^\beta, \widetilde{\boldsymbol{\beta}}_1 \sim \mathcal{IG}\left(\frac{A_{\gamma_\beta} + Ks(d^\beta)}{2}, \frac{B_{\gamma_\beta} + \widetilde{\boldsymbol{\beta}}_1^\top \widetilde{\boldsymbol{\beta}}_1}{2}\right),$$

where $\widetilde{\boldsymbol{\alpha}}_1$ and $\widetilde{\boldsymbol{\beta}}_1$ indicate subvector of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ for the selected coefficients only.

4. Sampling $\xi^2$

$$\xi^2 \mid \alpha_0, \boldsymbol{\beta}_0 \sim \mathcal{IG}\left(\frac{C + (1+K)}{2}, \frac{D + \alpha_0^2 + \boldsymbol{\beta}_0^\top \boldsymbol{\beta}_0}{2}\right).$$

5. Sampling $\sigma_i^2$

$$\sigma_i^2 \mid d^\alpha, d^\beta, \mathbf{Z}, \widetilde{\Gamma} \sim \mathcal{IG}\left(\frac{v_{0,i} + T_i}{2}, \frac{S_{0,i} + S_i + \left(\widetilde{\Gamma} - \widehat{\widetilde{\Gamma}}_i\right)^\top \widetilde{\mathcal{W}}_i^\top \widetilde{\mathcal{W}}_i \left(\widetilde{\Gamma} - \widehat{\widetilde{\Gamma}}_i\right)}{2}\right).$$

6. Sampling $\widetilde{\Gamma}$

$$\widetilde{\Gamma} \mid d^\alpha, d^\beta, \mathbf{Z}, \sigma_i^2, \gamma^2, \xi^2 \sim \mathcal{MVN}\left(\widetilde{\Gamma}_1, V_1\right).$$

7. Note that sampling the latent factors conditional on coefficients and data is standard in the Bayesian literature (e.g., Geweke and Zhou, 1996). We sample latent factors from the conditional distribution. For each time $t$, conditional on all other parameters, the joint distribution of $\mathbf{f}_t$ and $\mathbf{r}_t$ is Gaussian:

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{r}_t \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{0}_{K\times 1} \\ \boldsymbol{\alpha}(\mathbf{Z}_{t-1}) \end{pmatrix}, \begin{pmatrix} \mathbb{I}_K & \boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top \\ \boldsymbol{\beta}(\mathbf{Z}_{t-1}) & \boldsymbol{\beta}(\mathbf{Z}_{t-1})\boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top + \boldsymbol{\Sigma} \end{pmatrix}\right]$$

The latent factor is drawn from a conditional normal distribution with the following mean and covariance matrix:

$$\begin{aligned} \mathbb{E}\left[\mathbf{f}_t \mid \boldsymbol{\alpha}(\mathbf{Z}_{t-1}), \boldsymbol{\beta}(\mathbf{Z}_{t-1}), \boldsymbol{\Sigma}, \mathbf{r}_t\right] &= \boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top \left[\boldsymbol{\beta}(\mathbf{Z}_{t-1})\boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top + \boldsymbol{\Sigma}\right]^{-1} \left[\mathbf{r}_t - \boldsymbol{\alpha}(\mathbf{Z}_{t-1})\right], \\ \mathrm{Cov}\left[\mathbf{f}_t \mid \boldsymbol{\alpha}(\mathbf{Z}_{t-1}), \boldsymbol{\beta}(\mathbf{Z}_{t-1}), \boldsymbol{\Sigma}, \mathbf{r}_t\right] &= \mathbb{I}_K - \boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top \left[\boldsymbol{\beta}(\mathbf{Z}_{t-1})\boldsymbol{\beta}(\mathbf{Z}_{t-1})^\top + \boldsymbol{\Sigma}\right]^{-1} \boldsymbol{\beta}(\mathbf{Z}_{t-1}). \end{aligned}$$

**Fixed Sparsity.** In the fixed-sparsity specification, the number of selected characteristics in $\boldsymbol{\alpha}(\mathbf{Z})$ and $\boldsymbol{\beta}(\mathbf{Z})$ are constrained ex ante to $s(d^\alpha) = M_\alpha$ and $s(d^\beta) = M_\beta$. Relative to the learning-sparsity sampler in Step 1 above, only the update of $d_l^\alpha$ and $d_l^\beta$ dif-

fers: instead of drawing each indicator from its Bernoulli full conditional, we update $\boldsymbol{d} = (d^{\alpha}, d^{\beta})$ using a Metropolis-Hastings step.

At each iteration, we propose a candidate $\boldsymbol{d}'$ that satisfies the $(M_{\alpha}, M_{\beta})$ cardinality constraints and compute the acceptance probability

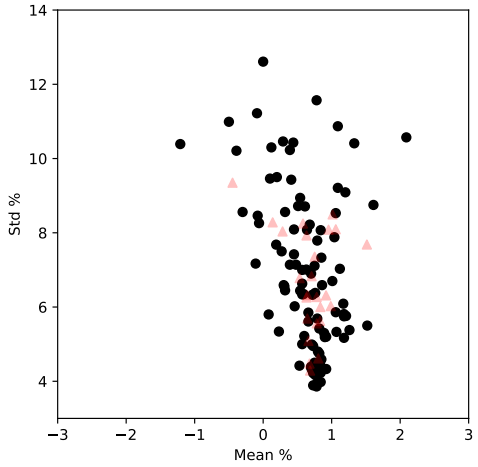
$$\alpha(\boldsymbol{d} \to \boldsymbol{d}') = \min\Big\{1,\, p(\boldsymbol{d}' \mid -)/p(\boldsymbol{d} \mid -)\Big\},$$

where $p(\boldsymbol{d} \mid -)$ is given in Equation (A.2). The proposal is accepted with probability $\alpha(\boldsymbol{d} \to \boldsymbol{d}')$. All remaining parameters are updated using the same Gibbs steps as in the learning-sparsity specification.

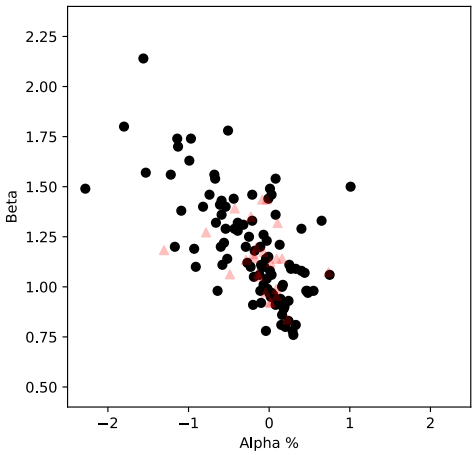
# II Data Overview

Figure A.1: Diversified P-Tree Test Assets

This figure compares 100 P-Tree test assets (black dots) with 25 ME/BM portfolios (light-red triangles). Panel (a) plots mean monthly returns against their standard deviations, and Panel (b) plots CAPM $\alpha$ against $\beta$. All return-based metrics are expressed in percentages.




(a) Mean v.s. Standard Deviation




(b) CAPM Alpha-Beta

Table A.1: Equity Characteristics

This table provides descriptions for the full set of 61 characteristics used for portfolio formation, including characteristic sorts and the construction of P-Tree portfolios. Bold text denotes the representative subset of 20 characteristics used in the primary empirical analysis.

| No. | Characteristics | Description | Category |
|---|---|---|---|
| 1 | **ABR** | Abnormal returns around earnings announcement | Momentum |
| 2 | **ACC** | Operating accruals | Investment |
| 3 | **ADM** | Advertising expense-to-market | Intangibles |
| 4 | **AGR** | Asset growth | Investment |
| 5 | ALM | Quarterly asset liquidity | Intangibles |
| 6 | ATO | Asset turnover | Profitability |
| 7 | **BASPREAD** | Bid-ask spread (3 months) | Frictions |
| 8 | **BETA** | Beta (3 months) | Frictions |
| 9 | **BM** | Book-to-market ratio | Value-versus-growth |
| 10 | BM_IA | Industry-adjusted book to market | Value-versus-growth |
| 11 | CASH | Cash holdings | Value-versus-growth |
| 12 | CASHDEBT | Cash to debt | Value-versus-growth |
| 13 | **CFP** | Cashflow-to-price | Value-versus-growth |
| 14 | CHCSHO | Change in shares outstanding | Investment |
| 15 | CHPM | Change in Profit margin | Profitability |
| 16 | CHTX | Change in tax expense | Momentum |
| 17 | CINVEST | Corporate investment | Investment |
| 18 | DEPR | Depreciation/ PP&E | Momentum |
| 19 | DOLVOL | Dollar trading volume | Frictions |
| 20 | DY | Dividend yield | Value-versus-growth |
| 21 | **EP** | Earnings-to-price | Value-versus-growth |
| 22 | GMA | Gross profitability | Investment |
| 23 | GRLTNOA | Growth in long-term net operating assets | Investment |
| 24 | HERF | Industry sales concentration | Intangibles |
| 25 | HIRE | Employee growth rate | Intangibles |
| 26 | ILL | Illiquidity rolling (3 months) | Frictions |
| 27 | LEV | Leverage | Value-versus-growth |
| 28 | LGR | Growth in long-term debt | Investment |
| 29 | MAXRET | Maximum daily returns (3 months) | Frictions |
| 30 | **ME** | market equity value | Frictions |
| 31 | ME_IA | Industry-adjusted size | Frictions |
| 32 | **MOM1M** | Previous month return | Momentum |
| 33 | **MOM12M** | Cumulative returns in the past (2-12) months | Momentum |
| 34 | MOM36M | Cumulative returns in the past (13-36) months | Momentum |
| 35 | MOM60M | Cumulative returns in the past (13-60) months | Momentum |
| 36 | MOM6M | Cumulative returns in the past (2-6) months | Momentum |
| 37 | **NI** | Net equity issue | Investment |
| 38 | NINCR | Number of earnings increases | Momentum |
| 39 | NOA | Net operating assets | Investment |
| 40 | **OP** | Operating profitability | Profitability |
| 41 | PCTACC | Percent operating accruals | Investment |
| 42 | PM | Profit margin | Profitability |
| 43 | PS | Performance Score | Profitability |
| 44 | RD_SALE | R&D-to-sales | Intangibles |
| 45 | **RDM** | R&D-to-market | Intangibles |
| 46 | RE | Revisions in analysts' earnings forecasts | Intangibles |
| 47 | RNA | Return on net operating assets | Profitability |
| 48 | ROA | Return on assets | Profitability |
| 49 | **ROE** | Return on equity | Profitability |
| 50 | RSUP | Revenue surprise | Momentum |
| 51 | RVAR_CAPM | Idiosyncratic volatility -CAPM (3 months) | Frictions |
| 52 | RVAR_FF3 | Res. var. - Fama-French 3 factors (3 months) | Frictions |
| 53 | **SEAS1A** | 1-Year Seasonality | Intangibles |
| 54 | SGR | Sales growth | Value-versus-growth |
| 55 | **SP** | Sales-to-price | Value-versus-growth |
| 56 | STD_DOLVOL | Std of dollar trading volume (3 months) | Frictions |
| 57 | STD_TURN | Std.of Share turnover (3 months) | Frictions |
| 58 | **SUE** | Standardized unexpected quarterly earnings | Momentum |
| 59 | **SVAR** | Return variance (3 months) | Frictions |
| 60 | TURN | Shares turnover | Frictions |
| 61 | ZEROTRADE | Number of zero-trading days (3 months) | Frictions |

# III Structural Motivation: Latent Economic Dimension and Sparsity

This appendix formalizes the argument discussed in Section 2.1, namely that economic dimension and sparsity in the observed characteristic basis are distinct objects. Sparsity is a property of the representation in the characteristic space, whereas economic dimension refers to the latent fundamentals that generate returns. When characteristics are noisy, and the measurements of these fundamentals are rotated, the measurement system is unobserved, and the latent dimension does not identify whether the population-optimal representation is sparse or dense.

This section formalizes two points. First, latent dimension does *not* identify whether the population-optimal representation in the observed basis is *sparse or dense*: even when the economic primitives are one-dimensional, the optimal linear map from observed characteristics to exposures can be exactly sparse or fully dense, depending on the measurement system; and the same is true when the latent economy is high-dimensional. Second, once this nonidentification is acknowledged, the econometric problem takes a decision-theoretic form (Robert, 2007; Berger, 2013): the researcher faces *model uncertainty* over sparsity patterns.

A Schrödinger formulation with spike-and-slab prior (Mitchell and Beauchamp, 1988; George and McCulloch, 1993; Giannone et al., 2021) and summarized by posterior sparse probabilities corresponds to Bayesian uncertainty over these patterns, with an oracle-style guarantee under log-score loss.

## III.1 Underlying Economic Fundamentals and Sparsity

Consider an economy where firm $i$'s exposures are governed by a vector of latent fundamentals $u_i \in \mathbb{R}^J$, while the true dimension $J$ is unknown. The econometrician observes characteristics $z_i \in \mathbb{R}^L$ that provide noisy measurements of $u_i$:

$$z_i \;=\; Au_i + \eta_i, \qquad \mathbb{E}[\eta_i] = 0, \quad \text{Var}(\eta_i) = \Sigma_\eta, \tag{A.4}$$

where $A \in \mathbb{R}^{L\times J}$ is an unobserved measurement map and $\eta_i$ is measurement noise.

**Remark 1.** The "latent fundamentals" $u_i$ are the unobserved economic primitives that generate observed characteristics through a measurement system. These are conceptually distinct from latent pricing factors in IPCA-style return models: IPCA factors are statistical common components in returns, and need not coincide with the primitives $u_i$. The identification issue here concerns sparsity *in the observed characteristic basis,* not whether the return-factor dimension is small or large.

The true exposures are driven by the latent fundamentals as

$$\beta_i = Hu_i, \tag{A.5}$$

where $H$ is an unknown matrix (dimensions suppressed), applied row by row when $\beta_i$ is vector-valued.

Furthermore, the population-optimal linear characteristic representation, $B^\star$, is defined as the projection of these exposures onto the observed characteristics $z_i$:

$$B^\star \;=\; \mathrm{Cov}(Hu, z)\,\mathrm{Var}(z)^{-1} \;=\; H\Sigma_u A^\top \big(A\Sigma_u A^\top + \Sigma_\eta\big)^{-1}, \tag{A.6}$$

where $\Sigma_u = \mathrm{Var}(u_i)$ and we assume $\mathrm{Cov}(u_i, \eta_i) = 0$.

Equation (A.6) isolates the identification issue: whether $B^\star$ is sparse or dense is not determined by $J$ alone; it depends on the (unobserved) measurement map $A$, the covariance $\Sigma_u$, and measurement noise $\Sigma_\eta$, even if researchers know the exact values of those unobserved parameters.

**Proposition 1.** *Fix any menu size $L \geq 2$.*

*(i)* ***Small latent dimension ($J = 1$).*** *There exist measurement systems $(A, \Sigma_\eta)$ and exposure maps $H$ such that the population-optimal characteristic representation $B^\star$ in Equation (A.6) is* exactly sparse. *There also exist (possibly different) $(\tilde{A}, \tilde{\Sigma}_\eta)$ and $\tilde{H}$ such that $B^\star$ is* fully dense.

*(ii)* ***Large latent dimension (e.g., $J = L$).*** *There exist measurement systems $(A, \Sigma_\eta)$ and exposure maps $H$ such that $B^\star$ is* exactly sparse *even though the latent economy is L-*

*dimensional. There also exist (possibly different)* $(\tilde{A}, \tilde{\Sigma}_\eta)$ *and* $\tilde{H}$ *such that* $B^\star$ *is* fully dense.

*Proof.* Recall that $\text{Var}(z) = A\Sigma_u A^\top + \Sigma_\eta$ and $\text{Cov}(Hu, z) = H\Sigma_u A^\top$.

**(i) Small latent dimension:** $J = 1$**.** Set $\Sigma_u = 1$ and take $H = 1$, so $\beta_i = Hu_i = u_i$.

*(i.a) Exactly sparse.* $B^\star$ *with* $J = 1$. Let $A = e_1 \in \mathbb{R}^{L\times 1}$ and $\Sigma_\eta = \text{diag}(\sigma_1^2, \ldots, \sigma_L^2)$ with $\sigma_\ell^2 > 0$. Then $\text{Var}(z) = AA^\top + \Sigma_\eta = \text{diag}(1 + \sigma_1^2, \sigma_2^2, \ldots, \sigma_L^2)$ is invertible and $\text{Cov}(Hu, z) = H\Sigma_u A^\top = e_1^\top$. Hence $B^\star = e_1^\top \text{Var}(z)^{-1} = \dfrac{1}{1+\sigma_1^2}\, e_1^\top$, which has exactly one nonzero entry.

*(i.b) Fully dense.* $B^\star$ *with* $J = 1$. Let every characteristic measure the same scalar fundamental: $A = \mathbf{1}_L \in \mathbb{R}^{L\times 1}$ and $\Sigma_\eta = \sigma^2 I_L$ with $\sigma^2 > 0$. Then $\text{Var}(z) = \mathbf{1}_L\mathbf{1}_L^\top + \sigma^2 I_L$ is invertible and $\text{Cov}(Hu, z) = \mathbf{1}_L^\top$. Using the Sherman-Morrison formula,

$$(\sigma^2 I_L + \mathbf{1}_L\mathbf{1}_L^\top)^{-1} = \frac{1}{\sigma^2}\left(I_L - \frac{\mathbf{1}_L\mathbf{1}_L^\top}{\sigma^2 + L}\right) \Rightarrow B^\star = \mathbf{1}_L^\top(\sigma^2 I_L + \mathbf{1}_L\mathbf{1}_L^\top)^{-1} = \frac{1}{\sigma^2 + L}\,\mathbf{1}_L^\top,$$

which is fully dense.

**(ii) Large latent dimension:** $J = L$**.** Set $\Sigma_u = I_L$ and take $H = \mathbf{1}_L^\top$, so $\beta_i = \mathbf{1}_L^\top u_i$ loads on all $L$ fundamentals.

*(ii.a) Exactly sparse.* $B^\star$ *with* $J = L$. Let $A = I_L$ and $\Sigma_\eta = \text{diag}(\sigma_1^2, \ldots, \sigma_L^2)$. Then $\text{Var}(z) = I_L + \Sigma_\eta$ is diagonal and invertible, while $\text{Cov}(Hu, z) = H\Sigma_u A^\top = \mathbf{1}_L^\top$. Hence

$$B^\star = \mathbf{1}_L^\top(I_L + \Sigma_\eta)^{-1} = \left(\frac{1}{1+\sigma_1^2}, \ldots, \frac{1}{1+\sigma_L^2}\right),$$

which becomes exactly sparse by taking $\sigma_\ell^2 \to \infty$ for all but a chosen subset of indices.

*(ii.b) Fully dense.* $B^\star$ *with* $J = L$. Let $A = I_L$ and $\Sigma_\eta = \sigma^2 I_L$ with $\sigma^2 > 0$. Then

$$B^\star = \mathbf{1}_L^\top(I_L + \sigma^2 I_L)^{-1} = \frac{1}{1+\sigma^2}\,\mathbf{1}_L^\top,$$

which is fully dense. □

Proposition 1 implies that the latent dimension identifies neither a sparse nor a

dense representation in the observed basis. Thus, "low-$J$" concept does not justify committing ex ante to a sparse map, and "large-$J$" concept does not justify committing ex ante to a fully dense map. This result formally demonstrates that the "Illusion of Sparsity" is not merely an empirical regularity but a structural feature of any system where observed characteristics are noisy, rotated proxies of latent economic primitives.

### III.2 Decision-theoretic Rationale under Log-score Loss

The nonidentification result in Proposition 1 implies that sparsity is not a structural object but a representation choice made by the econometrician. Committing ex-ante to a sparse or a dense specification, therefore, becomes a decision under model uncertainty: a sparse model may be misspecified when the representation is in fact dense, while a dense model may overfit when the signal is concentrated. The difficulty is amplified in practice because both the measurement system and its parameters are unknown, further obscuring which representation is closest to the population objective.

Thus, the question is not which single pattern is "true," but how to construct forecasts that are robust across alternative sparsity structures. We therefore recast the problem in a decision-theoretic framework, in which each sparsity pattern is treated as a candidate model, and performance is evaluated based on the log predictive score (Gneiting and Raftery, 2007).

Let $\mathcal{S}$ be a collection of candidate sparsity patterns, where each $S \in \mathcal{S}$ indexes a restricted model (e.g., the set of characteristics allowed to have nonzero coefficients). Let $\pi(S)$ denote a prior over patterns, with $\sum_{S\in\mathcal{S}} \pi(S) = 1$. Order the observed data as a sequence $\mathcal{D}_T = (y_1, \ldots, y_T)$. For each pattern $S$, let $p_S(y_t \mid y_{1:t-1})$ denote the one-step-ahead Bayesian predictive density implied by model $S$:

$$p_S(y_t \mid y_{1:t-1}) = \int p(y_t \mid y_{1:t-1}, \theta_S, S)\, d\Pi_S(\theta_S \mid y_{1:t-1}), \tag{A.7}$$

where $\theta_S$ denotes model-specific parameters and $\Pi_S$ is the prior on $\theta_S$.

The mixture predictive density, representing the Bayesian model average in predictive form, is defined as:

$$p_{\text{mix}}(y_t \mid y_{1:t-1}) = \sum_{S \in \mathcal{S}} \pi(S \mid y_{1:t-1})\, p_S(y_t \mid y_{1:t-1}), \quad \pi(S \mid y_{1:t-1}) \propto \pi(S)\, p_S(y_{1:t-1}), \tag{A.8}$$

where $p_S(y_{1:t-1}) = \prod_{\tau=1}^{t-1} p_S(y_\tau \mid y_{1:\tau-1})$ is the model-$S$ joint predictive density of the history.

**Lemma 1** (Log-score regret bound for the Bayes mixture)**.** *For any realized sample $\mathcal{D}_T = (y_1, \ldots, y_T)$ and any fixed pattern $S \in \mathcal{S}$,*

$$-\sum_{t=1}^{T} \log p_{\text{mix}}(y_t \mid y_{1:t-1}) \;\le\; -\sum_{t=1}^{T} \log p_S(y_t \mid y_{1:t-1}) \;-\; \log \pi(S). \tag{A.9}$$

*Proof.* By construction, the joint mixture probability of the realized sequence satisfies

$$p_{\text{mix}}(y_{1:T}) = \sum_{S' \in \mathcal{S}} \pi(S')\, p_{S'}(y_{1:T}) \;\ge\; \pi(S)\, p_S(y_{1:T}).$$

Taking $-\log$ and factorizing each joint probability into one-step-ahead predictive densities: $p(y_{1:T}) = \prod_{t=1}^{T} p(y_t \mid y_{1:t-1})$, which yields (A.9). □

**Interpretation of the $\log(1/\pi(S))$ term.** The log score at time $t$ is $-\log p(\, y_t \mid y_{1:t-1}\,)$, the negative log probability assigned to the realized observation. Thus, the left-hand side of (A.9) is the mixture's *cumulative predictive loss* (negative log predictive density). Inequality (A.9) says: relative to any fixed candidate pattern $S$, the mixture suffers at most an *additive* penalty $-\log \pi(S) = \log(1/\pi(S))$ over the entire sample. This penalty is the cost of not committing ex ante to a single pattern: if $\pi(S)$ is not too small for patterns considered plausible, the worst-case extra cumulative loss is modest. For example, if the prior is uniform over $|\mathcal{S}|$ candidate patterns, then $\log(1/\pi(S)) = \log|\mathcal{S}|$.

Lemma 1 is a formal "remaining agnostic" guarantee, closely related to standard oracle inequalities for Bayesian mixtures (see, e.g., Barron et al., 1998; Leung and Barron, 2006): regardless of whether the best-supported pattern is sparse, dense, or inte-

rior, mixture prediction adapts to the pattern supported by the data, and the cost of not knowing that regime ex ante is bounded by $\log(1/\pi(S))$. By utilizing a hierarchical spike-and-slab prior, the model endogenously searches for the optimal $S$ in this combinatorial space, ensuring the predictive loss remains near-optimal regardless of whether the true structure favors sparsity or density.

### III.3 Connection to Spike-and-Slab and Posterior Inclusion/Exclusion Probabilities.

To implement this mixture, we employ a hierarchical spike-and-slab prior as in Section 2.3. The binary indicators $d_\ell \in \{0, 1\}$ encode whether characteristic $\ell$ is active.

$$d_\ell \mid q \sim \text{Bernoulli}(1-q), \qquad q \sim \text{Beta}(a, b), \qquad \ell = 1, \ldots, L, \tag{A.10}$$

together with a conditional prior for coefficients given $d = (d_1, \ldots, d_L)$. This induces a prior over sparsity patterns $S = \{\ell : d_\ell = 1\}$:

$$\pi(S) = \Pr(d_\ell = 1 \; \forall \ell \in S, \; d_\ell = 0 \; \forall \ell \notin S) = \frac{B\left(a + L - |S|, \; b + |S|\right)}{B(a, b)}, \tag{A.11}$$

where $B(\cdot, \cdot)$ is the Beta function. The posterior exclusion probability, $\Pr(d_\ell = 0 \mid \mathcal{D}_T)$, summarizes the posterior mass assigned to patterns that exclude characteristic $\ell$:

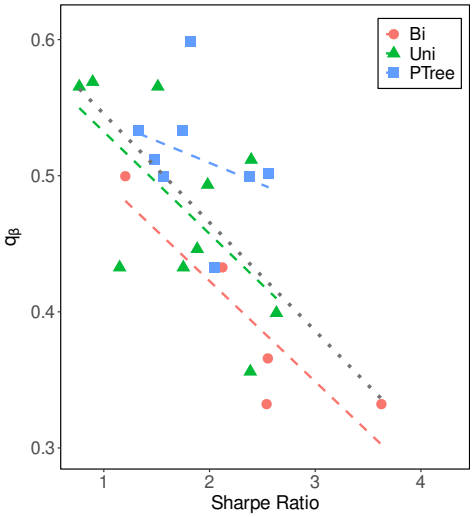
$$\Pr(d_\ell = 0 \mid \mathcal{D}_T) = 1 - \sum_{S \in \mathcal{S}:\, \ell \in S} \Pr(S \mid \mathcal{D}_T).$$

Aggregating these exclusion probabilities yields an interpretable, state-dependent summary of sparsity and provides a principled way to navigate the trade-off between parsimony and explanatory power without hard-wiring a fixed sparsity level ex ante.

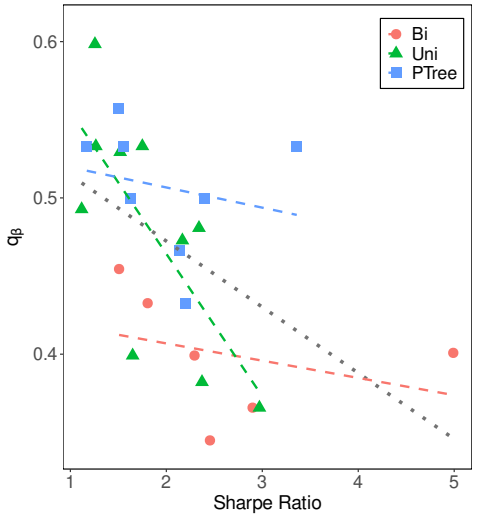
# IV Additional Tables and Figures

Figure A.2: Sparsity and Sharpe Ratio

This figure plots $q_\beta$ estimates against test asset Sharpe ratios, grouped by average absolute CAPM $\alpha$ (Panel a) and FF5 $\alpha$ (Panel b). Markers denote portfolio types: bi-sorted (red dots), uni-sorted (green triangles), and P-Tree (blue squares). Colored dashed lines represent group-specific linear fits, while the black dotted line shows the aggregate trend across all assets.




(a) Group by CAPM $|\alpha|$




(b) Group by FF5 $|\alpha|$

Table A.2: Pairwise Comparison of Distributional Properties across Regimes

This table reports pairwise comparisons of model-implied sparsity distributions across regimes. For each pair, Regimes 1 vs. 2, 1 vs. 3, 2 vs. 3, and Recession vs. Normal, we report the Wasserstein-1 distance, the Kolmogorov-Smirnov (KS) statistic with its associated $p$-value, and the posterior probability that a random draw from the first regime ($a$) exceeds one from the second ($b$), denoted as Prob($a > b$).

| Metrics | Regime1 vs. 2 | Regime1 vs. 3 | Regime2 vs. 3 | Recession vs. Normal |
|---|---|---|---|---|
| Wasserstein-1 distance | 0.006 | 0.027 | 0.033 | 0.071 |
| KS stat. | 0.030 | 0.122 | 0.148 | 0.313 |
| KS $p$-value | 0.349 | 0.000 | 0.000 | 0.000 |
| Prob. $a > b$ | 0.024 | 0.499 | 0.585 | 0.716 |

## Table A.3: Performance for Various Factor Specifications

This table reports performance metrics for four specifications: LF5, CAPM, FF5, and IPCA5 (estimated without mispricing). The analysis spans P-Tree assets of varying sizes (Panel A), individual stocks grouped by market equity (Panel B), and three standard portfolio sets (Panel C). Reported metrics include cross-sectional $R^2$ ($\text{CSR}^2$) and posterior estimates for $q_\beta$ and $\widehat{M}_\beta$.

| | Full-sample | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LF5 | | | CAPM | | | FF5 | | | IPCA5 | | |
| | $\text{CSR}^2$ | $q_\beta$ | $\widehat{M}_\beta$ | $\text{CSR}^2$ | $q_\beta$ | $\widehat{M}_\beta$ | $\text{CSR}^2$ | $q_\beta$ | $\widehat{M}_\beta$ | $\text{CSR}^2$ | $q_\beta$ | $\widehat{M}_\beta$ |
| *Panel A: P-Tree Portfolios* | | | | | | | | | | | | |
| 100 | 59.6 | 0.50 | 10 | 15.6 | 0.37 | 14 | 51.7 | 0.26 | 17 | 43.9 | 0.26 | 17 |
| 200 | 69.4 | 0.37 | 14 | 20.4 | 0.35 | 14 | 61.2 | 0.20 | 19 | 46.2 | 0.20 | 19 |
| 400 | 63.3 | 0.26 | 17 | 20.6 | 0.35 | 14 | 58.9 | 0.17 | 20 | 34.2 | 0.17 | 20 |
| *Panel B: Individual Stocks* | | | | | | | | | | | | |
| Big500 | 46.9 | 0.30 | 16 | -2.3 | 0.33 | 15 | 7.7 | 0.21 | 19 | 39.2 | 0.17 | 20 |
| Small500 | 30.1 | 0.42 | 12 | 1.2 | 0.45 | 10 | 9.5 | 0.32 | 16 | 21.9 | 0.24 | 17 |
| *Panel C: Other Portfolios* | | | | | | | | | | | | |
| ME/BM25 | 53.6 | 0.50 | 10 | 0.5 | 0.56 | 6 | 49.1 | 0.26 | 17 | 53.3 | 0.24 | 18 |
| Bi360 | 71.6 | 0.21 | 19 | 28.6 | 0.26 | 17 | 70.1 | 0.17 | 20 | 15.3 | 0.17 | 20 |
| Uni610 | 66.1 | 0.23 | 18 | 34.8 | 0.29 | 16 | 64.3 | 0.17 | 20 | -8.4 | 0.17 | 20 |

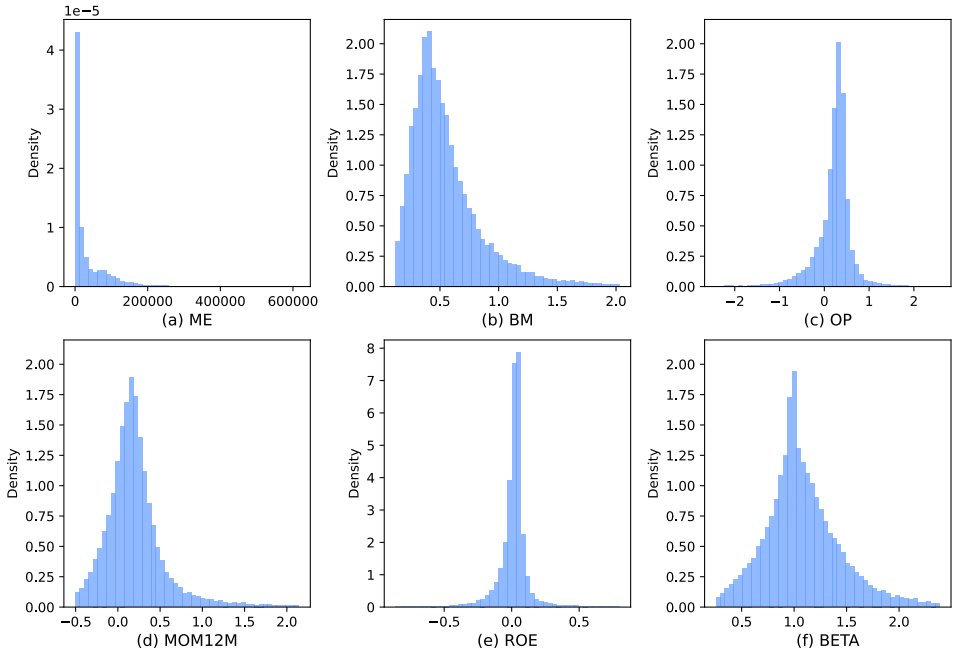
# Internet Appendix

# Schrödinger's Sparsity in the Cross Section of Stock Returns

## I Figures and Tables

Figure IA.1: Histograms of Some Characteristics

This figure illustrates a subset of characteristics from the P-Tree 100 portfolios. Note that these are portfolio-level, weighted raw characteristics. Our method, however, employs cross-sectionally standardized portfolio characteristics.




## II Regime-based Characteristics Importance

In this subsection, we investigate which specific firm characteristics are relevant across different economic environments. Figure IA.2 presents the posterior *inclusion* probabilities and associated coefficient estimates for the first latent factor from different regimes. We report results conditional on both the endogenously estimated structural regimes defined by Smith and Timmermann (2021) and business cycle states (recession vs. non-recession) identified by the Sahm Rule.

Figure IA.2: Regime-based Characteristics Importance

This figure illustrates the inclusion probabilities of characteristics across test assets, along with the associated coefficient estimates. For brevity, results are shown for the first latent factor only. Each cell reports the coefficient estimate, with color intensity indicating the inclusion probability. The $\beta_0$ coefficients corresponding to the first latent factor ($\beta_{0,\mathrm{LF}_1}$) are also presented.


We first focus on the identified structural regimes. Notably, a small set of characteristics – in particular market beta (BETA), book-to-market (BM), cashflow-to-price (CFP), market equity (ME), momentum (MOM1M, MOM12M), and return variance (SVAR) show high inclusion probabilities and economically meaningful loadings across most regimes. These variables capture persistent sources of systematic risk.

Second, the figure reveals strong state dependence in the signs and magnitudes of several coefficients. For example, the loading on BETA is strongly positive in Regime 1, but turns negative in later structural regimes, implying the factor tilts dynamically

between low- and high-beta stocks depending on the risk environment. Even for persistently selected characteristics like ME and SVAR, the coefficients exhibit substantial variation across these structural breaks, indicating that the compensation for these characteristics is strongly regime dependent. Other variables, such as net equity issue (NI) and sales-to-price (SP), become relevant only in specific regimes.

Furthermore, under recession and non-recession regimes as defined by the Sahm Rule Recession Indicator, we observe a significant shift in the drivers of factor loading. During recessionary periods, asset growth (AGR) and R&D-to-market (RDM) exhibit substantially lower posterior inclusion probabilities than in normal times. This decline suggests that in times of heightened macroeconomic stress, the dominant priced risk dimension shifts away from long-horizon growth and investment fundamentals toward downside and liquidity risk. The reduced relevance of AGR likely reflects that in recessions, investment is cut back more uniformly (e.g., Cooper et al., 2008); likewise, the diminished role of RDM is consistent with evidence that market participants shift their focus to systemic risk and solvency, rather than firm characteristics.

In addition, characteristics such as operating profitability (OP) lose much of their ability to explain systematic risk, aligning with the hypothesis that the signal-to-noise ratio of accounting data deteriorates during crises due to write-downs and earnings volatility (e.g., Novy-Marx, 2013), or implies that during recessions, returns become increasingly driven by macroeconomic factors, thereby weakening the role of firm-specific characteristics in capturing variation in dynamic factor loadings.

## III Performance of With-Mispricing Models

We analyze the downside risk of the forecast-implied portfolios using the maximum drawdown (MDD) results reported in Table IA.1, interpreted alongside the Sharpe ratios in Table 4.

Within Panel A, the learning-sparsity configuration with $K = 5$, which achieves Sharpe ratios of 0.83, 0.76, and 0.78 in Table 4, features MDDs of roughly 0.30, 0.22, and 0.25 for the three portfolios, respectively. By contrast, some fixed-sparsity con-

Table IA.1: Forecast-Implied Investment Performance (MDD) for Various Models

This table reports out-of-sample performance for portfolios constructed from return forecasts Equation (13) using the model with mispricing. We consider sign-adjusted value-weighted (Sign-adj. VW), equal-weighted (Sign-adj. EW), and forecast-weighted (Forecast-W) strategies. $K$ denotes the number of latent factors, and we report the maximum drawdown (MDD) for each strategy.

Panel A implements learning-sparsity with $\text{Beta}(5,5)$ priors on $q_\alpha$ and $q_\beta$; we also report posterior estimates for $\{q, \widehat{M}\}$ across $\alpha$ and $\beta$. Note that $\widehat{M} = 0$ indicates no selection probability exceeds 0.5, not an empty set. Panel B imposes fixed sparsity limits $(M_\alpha, M_\beta)$, while Panels C and D provide benchmarks from the non-sparse Bayesian framework and standard IPCA, respectively. The first 15 years are used for estimation, with the subsequent 20 years for out-of-sample evaluation.

| | | Out-of-sample Maximum Drawdown | | | | | |
|---|---|---|---|---|---|---|---|
| | | Sign-adj. VW | | Sign-adj. EW | | Forecast-W | |
| | | $K=1$ | $K=5$ | $K=1$ | $K=5$ | $K=1$ | $K=5$ |
| *Panel A: Learned Sparsity* | | | | | | | |
| $(q_\alpha, q_\beta)$ prior mean | 0.5,0.5 | 0.42 | 0.30 | 0.47 | 0.22 | 0.53 | 0.25 |
| | | | In-sample estimates of $q_\alpha$, $q_\beta$ and $\widehat{M}_\alpha, \widehat{M}_\beta$<br>$K=1$: (0.57,0.44) and (7,11);<br>$K=5$: (0.80,0.44) and (0,12) | | | | |
| *Panel B: Fixed Sparsity* | | | | | | | |
| | 2,2 | 0.52 | 0.53 | 0.51 | 0.60 | 0.59 | 0.57 |
| | 10,2 | 0.37 | 0.39 | 0.39 | 0.40 | 0.47 | 0.49 |
| | 18,2 | 0.36 | 0.34 | 0.37 | 0.21 | 0.44 | 0.31 |
| | 2,10 | 0.51 | 0.94 | 0.51 | 0.94 | 0.59 | 0.94 |
| $(M_\alpha, M_\beta)$ | 10,10 | 0.37 | 0.55 | 0.41 | 0.62 | 0.50 | 0.63 |
| | 18,10 | 0.36 | 0.55 | 0.39 | 0.62 | 0.45 | 0.64 |
| | 2,18 | 0.51 | 0.91 | 0.50 | 0.92 | 0.59 | 0.94 |
| | 10,18 | 0.37 | 0.55 | 0.41 | 0.62 | 0.49 | 0.60 |
| | 18,18 | 0.35 | 0.22 | 0.38 | 0.63 | 0.45 | 0.82 |
| *Panel C: No Sparsity* | | | | | | | |
| $(M_\alpha, M_\beta)$ | 20 | 0.31 | 0.33 | 0.41 | 0.32 | 0.46 | 0.39 |
| *Panel D: IPCA* | | | | | | | |
| $(M_\alpha, M_\beta)$ | 20 | 0.35 | 0.29 | 0.29 | 0.26 | 0.36 | 0.35 |

figurations with $K = 5$ produce near-catastrophic drawdowns of approximately 0.9, indicating that hard-wiring the sparsity level can create fragile portfolios, even when average returns appear acceptable.

Dense benchmarks in Panels C and D tend to have somewhat smaller MDDs, typically in the 0.3-0.4 range, but at the cost of materially lower Sharpe ratios. Relative to these alternatives, the learning-sparsity specification strikes a more attractive balance

between reward and risk: it transforms characteristic-driven forecasts into portfolios that simultaneously achieve high out-of-sample Sharpe ratios and avoid the extreme tail events observed under poorly tuned sparsity. In this sense, treating sparsity as a latent quantity to be learned, rather than as a fixed tuning parameter, improves not only average performance but also the resilience of forecast-based strategies to large drawdowns.

## References


Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *Journal of Finance 63*(4), 1609–1651.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

Smith, S. C. and A. Timmermann (2021). Break risk. *Review of Financial Studies 34*(4), 2045–2100.