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

Doron Avramov Guanhao Feng Jingyu He Shuhua Xiao

November 19, 2025

Abstract

This paper introduces *Schrödinger's Sparsity*, a Bayesian asset pricing framework that models sparsity as a latent feature to be inferred rather than imposed. By combining spike-and-slab priors with Bayesian updating, the model learns posterior probabilities of sparsity in both mispricing (alphas) and factor loadings (betas). Empirically, we find that alphas exhibit consistently greater sparsity than betas; however, neither collapses to full sparsity nor full density. Instead, an intermediate level of sparsity emerges endogenously, varying systematically with asset complexity, pricing difficulty, and macroeconomic regimes, and tightening during economic recessions. By allowing complexity to arise endogenously, the framework delivers interpretable, adaptive, and high-performing predictions, thereby bridging the gap between sparse selection and dense shrinkage approaches.

Key Words: Bayes; conditional model; Slack and Slab prior; latent factors; mispricing; sparsity.

JEL Classification: C12, 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, 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 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.

1 Introduction

Sparsity has become a foundational concept in high-dimensional modeling, enhancing interpretability, parsimony, and predictive performance. In empirical asset pricing, the proliferation of risk factors and anomalies (e.g., Harvey et al., 2016; Green et al., 2017) has motivated the adoption of sparse models to select a small number of factors from a large pool of potential candidates. Empirical studies demonstrate that sparse modeling effectively explains cross-sectional returns and improves the forecasting of risk premia (e.g., Gu et al., 2020; Feng et al., 2020; Freyberger et al., 2020).

Recent research questions the universality of the sparsity assumption in characteristic-based models. Giannone et al. (2021) show that Bayesian posteriors rarely concentrate on sparse models, describing this as the "illusion of sparsity." Kozak et al. (2020) propose that the stochastic discount factor (SDF) may rely on a broader set of firm characteristics beyond those typically captured by low-dimensional, ad hoc factor models. Similarly, He et al. (2024) empirically test the hypothesis of sparse factor models and find limited supporting evidence. Advances in high-dimensional modeling, including nonlinear, latent, and deep learning models, show strong empirical performance in empirical asset pricing.[1] For example, Shen and Xiu (2024) find that Ridge regression outperforms Lasso in weak-signal environments typical of return prediction, underscoring the benefits of denser regularization in such contexts.

Together, these findings highlight fundamental modeling philosophies in financial econometrics. Traditional approaches typically impose an *ex ante* commitment to either sparsity (via L_1 penalties) or density (via L_2 shrinkage), shaping both model structure and interpretability. For instance, implementations of Instrumented Principal Component Analysis (IPCA) typically adopt either a dense factor structure (Kelly et al., 2019) or a sparse penalized specification (Bybee et al., 2023), each reflecting distinct structural assumptions and advantages. This dichotomy raises an important methodological question: Can the degree of sparsity be learned adaptively from the

[1] Recent studies include Lettau and Pelger (2020), Gu et al. (2021), Huang et al. (2022), Chen et al. (2024), Avramov et al. (2023), Didisheim et al. (2023), Feng et al. (2024), and Kelly et al. (2024).

data through Bayesian updating rather than being predetermined?

We introduce the concept of Schrödinger's Sparsity, a framework for asset pricing models that balances sparse and dense structures without committing to either a priori. Statistical inference resolves this ambiguity by estimating the posterior distribution over model complexity, enabling a principled trade-off between parsimony and explanatory power. In this framework, the prior acts as a wavefunction, encoding beliefs about variable inclusion, which are updated using data to yield interpretable posterior inclusion probabilities. This approach systematically navigates the tension between model simplicity and explanatory richness.

To implement this idea, we extend the latent factor model of Geweke and Zhou (1996) and Kelly et al. (2019) into a Bayesian framework with a hierarchical spike-and-slab prior. The model captures the joint evolution of mispricings (alphas) and factor loadings (betas) based on lagged firm characteristics. Observed returns update the posterior distribution dynamically, analogous to wavefunction collapse in quantum mechanics. This probabilistic mechanism ensures flexibility, allowing sparsity to emerge naturally through learning rather than relying on fixed thresholds. The framework accommodates diverse asset classes and investment horizons, enabling analysis of sparsity dynamics, mispricing, and portfolio optimization. Its Bayesian structure enhances interpretability by quantifying uncertainty in variable inclusion, balancing explanatory power with parsimony.

Three key innovations distinguish our approach: (i) The prior inclusion probability, analogous to an initial amplitude for variable presence, can be learned from data or set exogenously. (ii) Sparsity patterns for alphas and betas are modeled separately to reflect their distinct economic roles. (iii) The method accommodates both latent and observable factor structures, including CAPM, Fama–French, and exogenous latent factor specifications, within a unified conditional estimation framework.

Regularization techniques, such as Lasso and Elastic Net, play a pivotal role in asset pricing, as they extract robust signals from high-dimensional data. By tuning penalization strength via cross-validation or information criteria, these methods adapt

to diverse assets and market conditions. Their core advantage lies in reducing estimation noise and isolating predictive characteristics, enhancing empirical modeling of expected returns. However, once the penalty is set, the model structure becomes fixed: variables are either included or excluded, with sparsity predetermined for the sample. While effective, this approach does not address uncertainty in model complexity.

Our Bayesian framework addresses the limitation of fixed sparsity assumptions by treating sparsity as an inferable parameter. We employ probabilistic priors on predictor relevance, updating beliefs with data to derive economically interpretable posterior inclusion probabilities. These probabilities reflect both the importance of the predictor and the overall sparsity. By averaging across the model space, the approach integrates sparse selection and dense shrinkage while capturing structural uncertainty. This probabilistic treatment enhances risk–return profiles, especially under uncertain sparsity levels.

Our empirical analysis highlights four central findings that demonstrate the practical relevance of the proposed approach. First, sparsity is neither extreme nor uniform. Across test assets and model specifications, posterior probabilities indicate that alphas and betas exhibit intermediate sparsity levels rather than being fully sparse or fully dense. Alphas are consistently sparser than betas, reflecting their distinct economic roles. Moreover, alphas and betas display a complementary pattern: when factor loadings are dense, mispricing is sparse, and vice versa.

Second, sparsity adapts to the complexity of assets and the difficulty of pricing them. Simple test assets, such as ME/BM portfolios, require relatively few characteristics and exhibit high sparsity. In contrast, more complex portfolios (e.g., Panel-Tree portfolios or individual stocks) demand a broader set of predictors, resulting in lower sparsity. Consistent with this, assets with higher Jensen's alphas tend to exhibit denser mispricing terms, while those with higher Sharpe ratios rely on denser factor loadings.

Third, sparsity is time-varying. During recessions, posterior inclusion probabilities concentrate on a narrower set of characteristics, indicating that macroeconomic

stress compresses the dimensionality of relevant predictors. During normal periods, the model selects a more diverse set of characteristics, underscoring that model complexity expands and contracts endogenously in response to the state of the economy.

Fourth, the patterns are robust across latent and observable factor models. Replacing latent factors with observable ones, such as CAPM or Fama–French factors, preserves the broad tendency toward intermediate sparsity, though the exact values depend on the chosen specification. Incorporating latent factors alongside observable ones further improves performance by capturing signals that standard models miss.

Taken together, these findings demonstrate that intermediate sparsity systematically emerges from the data, varies with asset and macroeconomic conditions, and improves both interpretability and predictive performance.

We place the sparsity-versus-density debate at the center of asset pricing by recasting it as a problem of probabilistic inference. Unlike prior work that imposes either sparse or dense structures ex ante, our framework learns the posterior distribution of sparsity itself. In doing so, we establish a first-order contribution that reshapes the understanding of model complexity in finance.

Thus, our formulation moves beyond the existing machine learning literature on sparse modeling (e.g., Chinco et al., 2019; Gu et al., 2020; Feng et al., 2020; Freyberger et al., 2020; Cui et al., 2025) and recent debates about the virtue of complexity (e.g., Kelly et al., 2024; He et al., 2024). Whereas these important studies assess the relative performance of sparse versus dense specifications, we reframe the problem by treating sparsity itself as a latent quantity. In doing so, we allow the data to update flexible priors on the degree of sparsity, thereby unifying sparse selection and dense shrinkage within a single probabilistic framework.

This unification perspective also generalizes conditional factor models (e.g., Jagannathan and Wang, 1996; Lettau and Ludvigson, 2001) by embedding them in a Bayesian framework with time-varying, component-specific sparsity, offering a richer account of dynamic economic conditions. It subsumes the Bayesian model selection and shrinkage literature (e.g., Avramov, 2002; Barillas and Shanken, 2018; Chib et al.,

2020) by shifting the focus from averaging across predefined models to endogenously learning the degree of sparsity itself. Finally, it reconceptualizes regime-dependent pricing, showing that model complexity compresses systematically under macroeconomic stress (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 reports the main empirical results. Section 5 analyzes sparsity variations across cross-sectional and time-series dimensions. Section 6 extends the framework to observable factor models. Section 8 concludes with a summary of findings and implications.

2 Methodology

2.1 Expected Return Dynamics

Sparsity in the cross section of expected asset returns can be studied within the two main empirical asset pricing frameworks: 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 develop this framework by formulating a Bayesian version with a specially designed hierarchical prior that characterizes the probability distribution over the sparsity of characteristics in both alphas and betas. This formulation allows for a probabilistic assessment of both the relevance of individual characteristics and the overall degree of sparsity in the cross section of expected returns.

We first establish the notation used throughout the analysis. 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 firm characteristics observed at time $t-1$. The common asset pricing 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 is flexible and 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 hybrid specification that combines latent and observed factors.

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

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

Here, α_0 is the potential common 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, $\boldsymbol{\beta}_1$ is a $K \times KL$ matrix capturing how the loadings on each of the K factors vary with the L lagged firm characteristics, and $\mathbf{f}_t$ are latent factors. The Kronecker product $\mathbb{I}_K \otimes \mathbf{Z}_{i,t-1}$ ensures that each factor loading responds flexibly to the associated firm-specific information.

Our primary interest lies in the sparsity of the model, specifically in $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$, which determine which characteristics are relevant for explaining alphas and betas, respectively. The prior specification used to infer the probability distribution over sparsity is described in Section 2.2, while the case of fixed sparsity levels is discussed in Section 2.3.

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 $\mathbb{E}[\mathbf{f}_t] = \mathbf{0}$, $\mathbb{E}[\mathbf{f}_t \mathbf{f}_t^\top] = \mathbb{I}$, 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

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

which shows that the conditional factor model admits a pooled regression representation. Inference is performed within a Bayesian framework using Markov chain Monte Carlo (MCMC), which generates draws from the conditional posterior distributions of the model parameters and latent variables.

The model described above explicitly allows for mispricing through the intercept terms. When only $\alpha_0 \neq 0$, the model captures common mispricing. When $\boldsymbol{\alpha}_1 \neq \mathbf{0}$, mispricing varies systematically with firm characteristics. A specification without mispricing is obtained by omitting the alpha-related components, in which case Equation (2) simplifies to:

$$r_{i,t} = \boldsymbol{\beta}_0^\top \mathbf{f}_t + \boldsymbol{\beta}_1^\top [\mathbf{f}_t \otimes \mathbf{Z}_{i,t-1}] + \epsilon_{i,t}. \tag{3}$$

As noted earlier, our framework flexibly accommodates both observable and latent factors. Let $\mathbf{f}_t = [\mathbf{f}_t^O, \mathbf{f}_t^L]$ denote the vector of observable and latent factors, respectively. The associated coefficient functions, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$, can be decomposed into components corresponding to each factor type. When latent factors are included, the model jointly learns the loading coefficients while treating $\mathbf{f}_t^L$ as unobserved variables inferred from the data.

Model identification. We follow the identification strategy of Kelly et al. (2019). Let $\Gamma_\alpha = [\alpha_0, \boldsymbol{\alpha}_1]$ and $\Gamma_\beta = [\boldsymbol{\beta}_0, \boldsymbol{\beta}_1]$. We impose the constraint $\Gamma_\beta^\top \Gamma_\beta = \mathbb{I}_K$, ensure that the unconditional second-moment matrix of $\mathbf{f}_t$ is diagonal with descending diagonal entries, and restrict the mean of $\mathbf{f}_t$ to be non-negative. To preserve the structure of Γ_β, these constraints are applied separately to each factor. We additionally impose $\Gamma_\alpha^\top \Gamma_\beta = \mathbf{0}_{1\times K}$, which is enforced by regressing Γ_α on Γ_β and replacing Γ_α with the residual from this regression.

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.

2.2 Priors Over Sparsity Structures

The prior consists of two components: a normal distribution and a point mass at zero. Let d_i^α and d_i^β be binary indicators for the inclusion of the i-th coefficients in the mispricing and factor loading functions, respectively. If $d_i^\alpha = 1$, the corresponding mispricing coefficient is included in the model and assigned a normal prior $\mathcal{N}(0, \gamma_\alpha^2)$, which induces shrinkage without enforcing exact sparsity. If $d_i^\alpha = 0$, it is excluded from the alpha function via a spike prior that places an infinite mass at zero and fixes the coefficient to zero. The variance hyperparameter γ_α^2 controls the degree of shrinkage and follows an inverse-gamma prior:

$$\gamma_\alpha^2 \sim \mathcal{IG}\left(A_{\gamma_\alpha}/2, B_{\gamma_\alpha}/2\right). \tag{4}$$

This normal prior acts as the Bayesian analogue of an L_2 penalty (ridge regression) due to its quadratic form.

Similarly, if $d_i^\beta = 1$, the corresponding factor loading coefficient is assigned a normal prior $\mathcal{N}(0, \gamma_\beta^2)$; if $d_i^\beta = 0$, the coefficient is excluded from the beta function via a point mass at zero. The shrinkage variance γ_β^2 is also given an inverse-gamma prior:

$$\gamma_\beta^2 \sim \mathcal{IG}\left(A_{\gamma_\beta}/2, B_{\gamma_\beta}/2\right). \tag{5}$$

Each inclusion indicator follows a Bernoulli prior:

$$\begin{aligned} d_i^\alpha &\sim \text{Bernoulli}(1 - q_\alpha), \\ d_i^\beta &\sim \text{Bernoulli}(1 - q_\beta), \end{aligned} \tag{6}$$

where q_α and q_β are sparsity probabilities. While standard spike-and-slab priors (Mitchell and Beauchamp, 1988) treat these probabilities as fixed hyperparameters, we follow

Giannone et al. (2021) in assigning independent Beta hyperpriors:

$$\begin{aligned} q_\alpha &\sim \text{Beta}(a_{q_\alpha}, b_{q_\alpha}), \\ q_\beta &\sim \text{Beta}(a_{q_\beta}, b_{q_\beta}), \end{aligned} \tag{7}$$

allowing the overall level of sparsity to be inferred from the data.

Different choices of prior parameters reflect varying beliefs about model sparsity; however, these beliefs are not binding, as the posterior distributions of q_α and q_β are shaped jointly by the priors and the observed data. Importantly, the posteriors of the sparsity probabilities q_α and q_β lie in the unit interval $(0, 1)$, offering a probabilistic rather than binary view of model complexity. This formulation avoids sharp thresholding and instead quantifies uncertainty about sparsity. While one could compute sparsity *ex post* by counting the number of nonzero coefficients (He et al., 2024), we adopt a fully Bayesian approach that explicitly captures the posterior uncertainty surrounding the sparsity probabilities q_α and q_β.

2.3 Prior for Exogenous Fixed Sparsity Level

We also consider an alternative prior that fixes the sparsity level rather than learning it probabilistically from the data. This serves as a benchmark, reflecting the traditional approach that assumes either a sparse or dense structure *ex ante*. A common example is Lasso regression, where the regularization parameter determines the degree of penalization and, consequently, the implied sparsity level. Once this parameter is set, typically via cross-validation or Bayesian information criterion, the resulting sparsity is fixed: stronger penalization yields a sparser model, while weaker penalization results in a denser specification. However, such methods do not allow the sparsity level to adapt flexibly to the data in a fully probabilistic manner.

For this fixed-sparsity benchmark, we impose a binding constraint on the total number of included predictors by restricting the sum of the inclusion indicators d_i^α

and d_i^β. Specifically, we adopt the following joint priors:

$$
\begin{aligned}
(d_1^\alpha, d_2^\alpha, \cdots, d_L^\alpha) &\sim \left[\prod_{i=1}^{L} \text{Bernoulli}(1-q_\alpha)\right] \times \mathbb{I}\left(\sum_{i=1}^{L} d_i^\alpha = M_\alpha\right), \\
(d_1^\beta, d_2^\beta, \cdots, d_L^\beta) &\sim \left[\prod_{i=1}^{L} \text{Bernoulli}(1-q_\beta)\right] \times \mathbb{I}\left(\sum_{i=1}^{L} d_i^\beta = M_\beta\right),
\end{aligned}
\tag{8}
$$

where $M_\alpha \leq L$ and $M_\beta \leq L$ denote the fixed number of selected characteristics for alphas and betas, respectively. These values directly control the sparsity level in the model and are specified by the researchers.

To perform inference on all model parameters and latent factors, we draw samples from the joint posterior distribution using a Gibbs sampling algorithm. Full details of the sampling procedure are provided in the Appendix.

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.[2] Portfolios are constructed from these stocks, and we also consider a representative subset of individual stocks as test assets.

We also utilize test assets based on the Panel-Tree (P-Tree) method of Cong et al. (2025), which constructs 400 portfolios using a machine learning algorithm that grows 40 trees with 10 "leaves" each. P-Tree portfolios are particularly well-suited to our sparsity analysis for two main reasons. First, they generalize multi-way characteristic sorting while maintaining full transparency on the characteristics used in each portfolio. Second, the portfolios are generated sequentially to expand the existing asset space and enhance the Sharpe ratio. This sequential construction prioritizes informative characteristics early on, resulting in diminishing marginal contributions from later

[2]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 IBES coverage.

portfolios. As a result, it naturally reveals how sparsity evolves with increasing model complexity.

To mitigate look-ahead bias, we follow Cong et al. (2025) in constructing the P-Tree portfolios using data from 1980–1989. 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.2 compares the cross-sectional dispersion of 100 P-Tree portfolios with the standard 25 ME/BM portfolios, highlighting the greater dispersion of the former.

In addition to P-Tree portfolios, we examine three commonly used sets of test assets: (i) the standard 25 ME/BM portfolios, (ii) 360 bivariate-sorted portfolios (Bi360),[3] (iii) 610 univariate-sorted portfolios (Uni610),[4] based on individual characteristics.

Characteristics. We compile 61 firm-level characteristics, grouped into eight themes: size, value, investment, momentum, profitability, liquidity, volatility, and intangibles. Table A.1 provides detailed descriptions. 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. The characteristics are then standardized within each portfolio class. Figure A.3 illustrates the raw (pre-standardization) values of selected characteristics for the P-Tree portfolios.

Regime. To explore how model sparsity evolves across macroeconomic conditions, we incorporate regime variation. For structural regimes, we follow Smith and Timmermann (2021), who identify significant breakpoints in the U.S. stock market and provide corresponding calendar dates. These serve as exogenous partitions for our regime-specific analyses.

[3] 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.

[4] Constructed with 61 characteristics and 10 breakpoints per characteristic, yielding 610 portfolios.

Additionally, we use the real-time Sahm Rule Recession Indicator from FRED. We group all months identified as recessionary into a single regime and contrast model behavior during these periods with non-recessionary (normal) times.

4 Schrödinger's Sparsity

A central objective of this paper is to adopt a probabilistic perspective on model sparsity. Rather than assuming the model to be either sparse or dense, we adopt a fully probabilistic view of sparsity and allow the degree of sparsity to be learned from the data. In short, "sparsity" is an object of inference.

4.1 Learning Sparsity Matters: Mispricing and Sparsity

We begin the analysis 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) compare the posterior sparsity of mispricing and factor loadings and (ii) assess the gains from learning (q_α, q_β) relative to fixed-sparsity and dense benchmarks. Broader asset classes and regime analyses follow in Section 5.

Asset Pricing Evaluation Before examining the nature of sparsity across model specifications and test assets, we begin by evaluating whether enabling the model to learn sparsity through a hierarchical spike-and-slab prior delivers an economically meaningful reduction in mispricing. In asset pricing, mispricing is the principal manifestation of model misspecification; a procedure that merely reduces estimation variance without materially lowering the magnitude of mispricing does not deliver an economic improvement. Guided by this objective, we assess performance along two complementary dimensions: (i) an outcome dimension that measures residual pricing errors in the return space, and (ii) a structural dimension that gauges the strength of the parameter channel that generates mispricing within the model. This distinction separates observed pricing errors from the mechanism that generates them, prevent-

ing variance shrinkage from being mistaken for an economic improvement. We adopt a Bayesian posterior-comparison framework, which enables transparent probability statements about model dominance and effect sizes, rather than relying on frequentist significance tests.

The first metric evaluates the overall magnitude of mispricing across all test assets. Within our model, mispricing is decomposed into two components: a constant intercept term, α_0, which captures persistent pricing errors common across assets, and a vector of characteristic-based intercepts, $\boldsymbol{\alpha}_1$, which reflects systematic mispricing related to firm characteristics. For each posterior draw g from the MCMC algorithm, we compute asset-level mispricing as $\widehat{\boldsymbol{\alpha}}_{it}^{(g)} = \alpha_0^{(g)} + \boldsymbol{\alpha}_1^{(g)\top}\mathbf{Z}_{i,t-1}$, and summarize the aggregate mispricing using

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

This computation yields a distribution over posterior draws, providing not only a point estimate but also a full characterization of the uncertainty surrounding the magnitude of mispricing. By examining the shape and location of the posterior distribution under different model specifications, we assess the effect of sparsity priors on pricing performance without relying on frequentist hypothesis tests.

The second metric quantifies the overall magnitude of the mispricing coefficients, following the approach of (Kelly et al., 2019). Specifically, we examine the scale of coefficients for mispricing $\Gamma_\alpha = [\alpha_0, \boldsymbol{\alpha}_1]$ by computing

$$W_\alpha^{(g)} = \widehat{\Gamma}_\alpha^{(g)\prime}\widehat{\Gamma}_\alpha^{(g)},$$

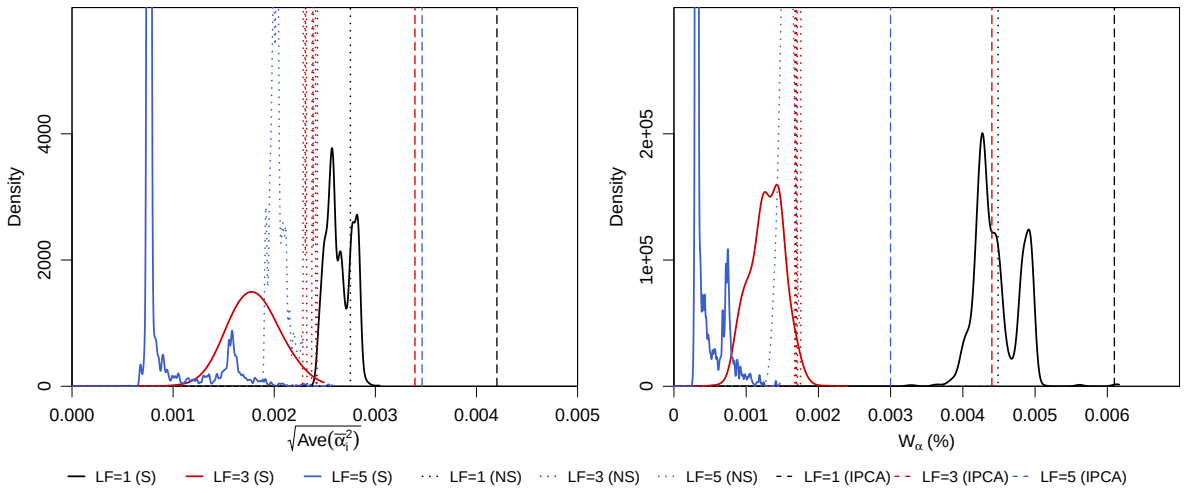
for each posterior draw g from the MCMC algorithm. This scalar captures the squared norm of the coefficient vector, providing a summary measure of how large the mispricings are, conditional of the included characteristics. Evaluating $W_\alpha^{(g)}$ across posterior draws yields its full posterior distribution, which allows us to compare the magnitude

and uncertainty of mispricing across models.

We evaluate the impact of imposing a hierarchical spike-and-slab prior on mispricing estimates by comparing three model specifications: (i) a sparse Bayesian specification (S) that employs a hierarchical spike-and-slab prior; (ii) a non-sparse Bayesian specification (NS) with a conjugate Normal prior and no variable selection; and (iii) the standard IPCA model of Kelly et al. (2019). Figure 1 presents the posterior distributions of the aggregate mispricing measure $\boldsymbol{\alpha}$ and the norm of the mispricing-related coefficients, W_α, under each specification. Since IPCA is a frequentist method, it yields point estimates only; these appear as vertical lines in the figure, without associated posterior uncertainty. We further test whether a given characteristic systematically accounts for the mispricing component. Details appear in Appendix A.I.

Figure 1: Posterior Mispricing under Model Specifications

This figure illustrates the posterior density (or value) of the mispricing component $\boldsymbol{\alpha}$ and norm of coefficients W_α under different model specifications. Panel (a) reports the posterior distributions or values of mispricing across different numbers of latent factors and model specifications. Panel (b) presents the posterior distributions or values of W_α. Black, red, and blue curves correspond to models with 1, 3, and 5 latent factors (LF), respectively. Solid lines depict our sparsity-aware specification (S), dotted lines depict our specification without sparsity (NS), and dashed lines refer to the IPCA method. For parsimony, in the sparsity-enabled specifications, we place Beta$(5, 5)$ priors on both q_α and q_β. Since IPCA delivers a point estimate, we indicate its value with a vertical line. Considering that several posterior densities attain relatively high values, we truncate the y-axis to facilitate a clearer comparison of the shapes of the remaining distributions.



(a) Value / density of $\boldsymbol{\alpha}$ (b) Value / density of $W_\alpha(\%)$

Figure 1 presents three consistent strands of evidence from the posterior comparison: empirical existence, posterior dominance, and mechanism.

First, across specifications and factor counts, the two mispricing measures are posteriorly distinct from zero; the high-density posterior intervals for both indicators exclude zero. Joint evidence on the outcome and structural dimensions rules out an estimation-noise explanation. If variance were merely reduced without addressing mispricing, the distribution of estimated alpha would primarily exhibit scale shrinkage rather than a shift in location.

Second, holding the number of factors fixed, the sparsity-learning Bayesian specification (S) improves upon the dense Bayesian (NS) in both location and shape: the posteriors for metrics shift materially to the left with thinner tails and minimal overlap, indicating that sparsity learning delivers not only narrower uncertainty but also a substantive reduction in the mispricing strength. IPCA provides point estimates only and typically lies to the right of S, often beyond the central mass of NS, consistent with larger residual mispricing and coefficient norms when model uncertainty is not characterized probabilistically and parameters do not receive selective shrinkage.

Moreover, as the number of factors increases, the posteriors for the two metrics move toward zero and become more concentrated, reflecting the complementarity between factor absorption and sparsity-induced shrinkage. Additional latent factors absorb a greater share of systematic variation, limiting the scope for misallocating risk exposure into alpha. The spike-and-slab prior then selects weak-signal characteristics and imposes disciplined shrinkage on marginal terms, while avoiding excessive penalization of salient components. The decline in mispricing thus arises from the coordinated action of both channels under posterior learning.

Taken together, the evidence supports the central claim of Schrödinger's sparsity: the mispricing channel is neither empty nor full, but adapts within an interior region through posterior learning. As model capacity increases, adaptive shrinkage works jointly with factor absorption to compress mispricing to economically acceptable levels. By contrast, ignoring sparsity or ignoring model uncertainty materially weakens this compression, yielding systematically higher mispricing and larger coefficient norms at the point-estimate level.

Probability of Sparsity Having verified the asset-pricing performance of our approach with respect to mispricing, we next study posterior learning of sparsity itself by treating the sparsity probabilities of the mispricing component and the factor loadings (denoted q_α and q_β) as objects of inference. This analysis serves three purposes: (i) to characterize the endogenous allocation of model complexity across the mispricing and risk-exposure channels; (ii) to align statistical improvements with economic content by asking whether more sparse is equivalent to better; and (iii) to identify the empirical footprint of "Schrödinger's sparsity," whereby sparsity neither collapses to an all-sparse corner nor expands to an all-dense corner but instead adapts to an interior region dictated by the data.

Panel A of Table 1 reports three sets of results across varying prior means for the sparsity probabilities (q_α, q_β) and different numbers of latent factors (K): (i) the cross-sectional R^2, (ii) the Sharpe ratio of the tangency portfolio formed from the estimated latent factors, and (iii) the posterior means of the sparsity probabilities for mispricing (q_α) and factor loadings (q_β). The cross-sectional R^2 is computed as:

$$\text{CSR}^2 = 1 - \frac{\sum_{i=1}^{N} \left(\frac{1}{T_i} \sum_{t=1}^{T_i} (r_{i,t} - \widehat{r}_{i,t}) \right)^2}{\sum_{i=1}^{N} \left(\frac{1}{T_i} \sum_{t=1}^{T_i} (r_{i,t} - \hat{\beta}_i \text{MktRF}_t) \right)^2}, \tag{9}$$

where $\widehat{r}_{i,t}$ denotes the predicted mean return, $\widehat{\beta}_i$ is the estimated market beta, and MktRF stands for the market factor.

We summarize the main findings from Panel A below. First, across prior-mean combinations and factor counts K, the posterior sparsity probabilities consistently fall between fully sparse (near 0) and fully dense (near 1). Both the mispricing coefficients ($\boldsymbol{\alpha}_1$) and factor loadings ($\boldsymbol{\beta}_1$) exhibit intermediate levels of sparsity, with mispricing generally being more sparse. For example, under a symmetric Beta prior with mean 0.5, the model with $K = 5$ yields posterior means of 0.79 for q_α and 0.50 for q_β. These posterior probabilities and associated model performance metrics remain stable across diverse prior specifications, demonstrating that the results are primarily data-driven and robust.

Table 1: Model Performance Under Different Priors

This table reports model results under various prior assumptions for sparsity on P-Tree 100 test assets, including cross-sectional R^2, Sharpe ratio from the latent factor tangency portfolio, and posterior means of sparsity levels. K indicates the number of latent factors. Panel A's row labeled "(q_α prior mean, q_β prior mean)" corresponds to unconstrained sparsity settings, where priors on q_α and q_β indicate three Beta distributions: $\text{Beta}(9,1)$, $\text{Beta}(5,5)$, and $\text{Beta}(1,9)$, with means of 0.9, 0.5, and 0.1, respectively. Panel B restricts the number of characteristics driving $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$. Specifically, M_α limits the number of characteristics influencing $\boldsymbol{\alpha}_1$, while M_β restricts the number of characteristics affecting each factor loading $\boldsymbol{\beta}_{1,k}$. Panel C reports results from our Bayesian framework without sparsity constraints, whereas Panel D presents estimates obtained via the standard (dense) IPCA method.

		CSR^2			TP. SR			(q_α, q_β)		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Unrestricted # selected chars.										
(q_α prior mean, q_β prior mean)	0.9,0.9	29.2	43.7	58.9	0.34	0.76	0.65	0.66,0.62	0.83,0.57	0.93,0.64
	0.5,0.9	29.4	43.4	57.0	0.35	0.78	0.88	0.51,0.62	0.68,0.60	0.77,0.64
	0.1,0.9	29.4	43.1	56.6	0.35	1.05	0.86	0.37,0.62	0.53,0.60	0.63,0.66
	0.9,0.5	29.3	44.3	**59.9**	0.34	0.81	0.52	0.66,0.47	0.82,0.47	0.93,0.50
	0.5,0.5	29.5	42.4	58.8	0.35	1.00	0.68	0.52,0.47	0.69,0.43	0.79,0.50
	0.1,0.5	29.5	43.6	58.1	0.35	1.19	0.84	0.37,0.47	0.53,0.46	0.64,0.49
	0.9,0.1	29.5	46.9	58.3	0.34	1.15	0.97	0.66,0.31	0.83,0.31	0.92,0.33
	0.5,0.1	29.6	42.5	57.9	0.35	1.00	0.87	0.52,0.31	0.69,0.30	0.79,0.34
	0.1,0.1	29.7	45.6	53.7	0.35	**1.38**	0.86	0.37,0.31	0.54,0.31	0.62,0.35
Panel B: Fixed # selected chars.										
(M_α, M_β)	2,2	25.4	49.3	48.4	0.44	1.10	0.95	/	/	/
	10,2	28.0	51.1	50.0	0.37	0.56	0.87	/	/	/
	18,2	25.2	46.9	37.8	0.32	0.77	0.67	/	/	/
	2,10	28.8	50.9	59.6	0.42	0.59	0.90	/	/	/
	10,10	29.6	38.3	41.1	0.35	0.87	1.07	/	/	/
	18,10	27.2	40.9	39.5	0.32	0.59	0.87	/	/	/
	2,18	29.8	**54.9**	56.1	**0.43**	0.64	0.89	/	/	/
	10,18	29.9	34.5	51.0	0.36	1.02	**1.10**	/	/	/
	18,18	27.5	39.3	42.1	0.33	0.54	0.92	/	/	/
Panel C: No sparsity										
(M_α, M_β)	(20,20)	**29.9**	36.9	45.2	0.35	0.57	0.95	/	/	/
Panel D: IPCA										
(M_α, M_β)	(20,20)	21.1	30.3	37.8	0.32	0.40	0.57	/	/	/

Second, mispricing and factor loading components exhibit systematically different sparsity patterns, with mispricing showing greater sparsity. For example, across all $K = 5$ specifications in Panel A, the posterior mean of q_α consistently exceeds 0.6 and reaches as high as 0.9, while the corresponding q_β values remain uniformly lower. This difference reflects the theoretical complementarity between mispricing and factor loadings. Once a characteristic meaningfully contributes to the factor structure, its incremental explanatory role in mispricing diminishes. Consequently, lower sparsity in factor loadings implies that a broader set of characteristics is used to explain expected

returns. This leaves less scope for persistent pricing errors, resulting in a sparser $\boldsymbol{\alpha}_1$.

Third, sparsity patterns vary systematically with the number of latent factors K. For instance, under a symmetric prior with (q_α, q_β) means of $(0.5, 0.5)$, the posterior estimates evolve from $(0.52, 0.47)$ at $K = 1$, to $(0.69, 0.43)$ at $K = 3$, and to $(0.79, 0.50)$ at $K = 5$. As K increases, the posterior probability of sparsity for factor loadings (q_β) remains low or declines, while the sparsity of mispricing (q_α) rises substantially. This pattern suggests that with more latent factors absorbing information from characteristics, the role of the mispricing component becomes increasingly limited. Moreover, the spike-and-slab prior filters weak signals via spike selection, applies disciplined group shrinkage to marginal terms, and avoids over-penalizing salient components. These forces operate jointly through two complementary channels, namely factor absorption and sparsity-induced shrinkage, and produce a structural compression of mispricing. Importantly, this trend holds across a wide range of prior configurations, reinforcing the notion that the results are data-driven.

Fourth, we find that model performance, as measured by both cross-sectional R^2 and the Sharpe ratio, is robust across various specifications. For example, at $K = 3$, the cross-sectional R^2 ranges from 42.4 to 46.9, while the Sharpe ratio tends to improve under denser priors on $\boldsymbol{\beta}_1$. More broadly, model fit improves substantially with larger K: R^2 increases from roughly 29% at $K = 1$ to over 55% at $K = 5$. When the factor structure is limited (i.e., small K), performance gains primarily come from the mispricing channel. In contrast, with richer factor structures (larger K), explanatory power shifts toward the factor loadings.

As demonstrated, the sparsity probabilities for mispricing and factor loadings significantly depend on the number of latent factors. This prompts a natural inquiry: what are the consequences of ignoring this "Schrödinger's sparsity" and instead imposing sparsity exogenously? Panel B of Table 1 presents results from fixed-sparsity models, where the number of characteristics driving $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ is constrained by (M_α, M_β). We examine nine combinations of inclusion sizes. For a given number of latent factors, model performance varies substantially with the imposed sparsity level.

In particular, performance is *non-monotonic* in both M_α and M_β: models that are either *too sparse* (e.g., $M = 2$) or *too dense* (e.g., $M = 18$) often underperform. For instance, at $K = 3$, the Sharpe ratio declines from 1.10 when $(M_\alpha, M_\beta) = (2, 2)$ to just 0.54 when both are set to 18.

Compared to Panel A, which endogenously learns sparsity from the data, Panel B shows that fixed-sparsity models perform best when their imposed (M_α, M_β) values align closely with the posterior sparsity levels learned under the probabilistic framework. For example, at $K = 5$, the best-performing probabilistic model implies sparsity near $(1, 10)$, while the best fixed model occurs at $(2, 10)$; for $K = 3$, the optimal posterior sparsity is around $(4, 14)$, while the top-performing fixed case is $(2, 18)$. This inverted-U pattern, along with the consistency with Panel A, highlights the importance of adaptively learning sparsity from the data, rather than imposing it ex ante.

Finally, models with probabilistic sparsity consistently outperform fully dense Bayesian models and the frequentist IPCA benchmark. Panels C and D of Table 1 highlight these comparisons. At $K = 5$, the cross-sectional R^2 is 45.2% for the dense Bayesian model and 37.8% for IPCA, while the corresponding Sharpe ratios are 0.95 and 0.57. Both metrics are substantially lower than those achieved by the probabilistic sparsity model. By treating sparsity as a latent parameter rather than imposing full density or hard-thresholded sparsity, the model can dynamically adjust to the underlying structure of the data. This approach delivers superior statistical and economic performance by enabling more robust learning about sparsity.

Taken together, the evidence supports the central proposition of "Schrödinger's sparsity": sparsity is a learnable latent quantity that adapts with both the number of factors and the information in the sample. Moreover, alpha and beta exhibit stable asymmetry and complementarity in their sparsity structure.

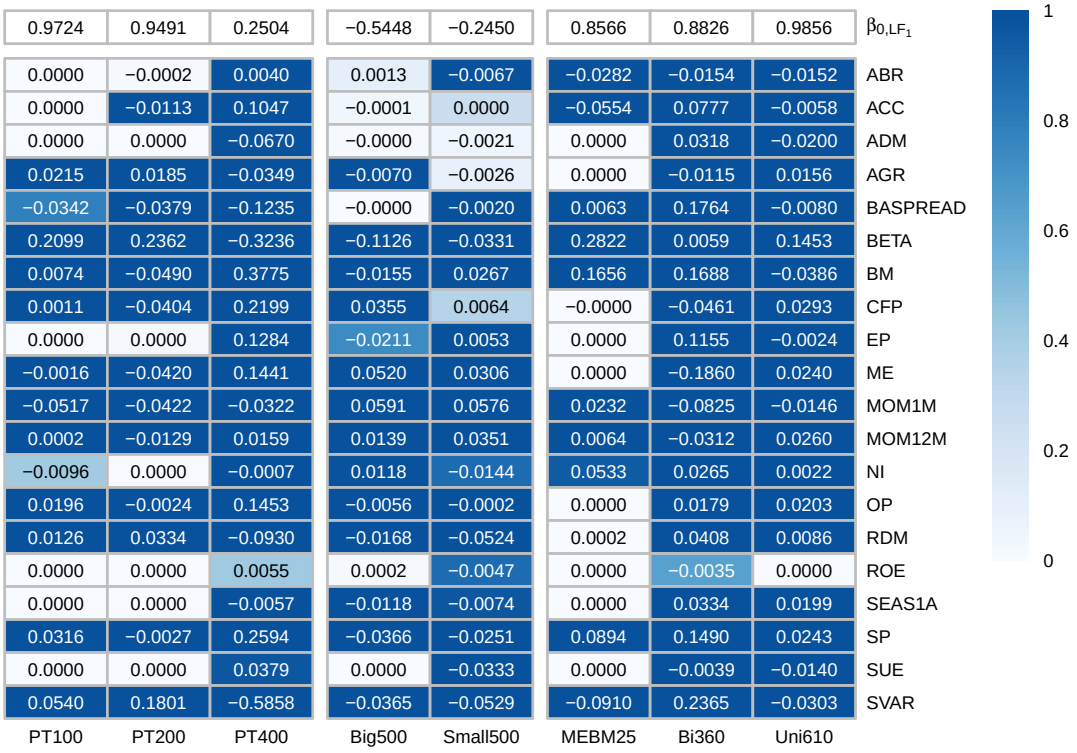
4.2 Models without Mispricing

Prior results demonstrate that incorporating the mispricing channel enables posterior learning of sparsity, thereby enhancing pricing performance. A natural counterfactual follows: if mispricing were excluded and the entire explainable component

of expected returns were assigned to the factor structure, how would information be reallocated? We further assess model performance in the absence of mispricing by estimating conditional latent factor models that attribute all variation in expected returns to the factor structure.

Figure 2: Characteristics Importance in Different Test Assets (without Mispricing)

This figure depicts selection probabilities of characteristics across test assets. Panel (a) shows the probability of selection for explaining factor loadings and the associated $\beta_{1,i}$ estimates. For brevity, results are shown for the first latent factor only. Each cell displays the selection probability, with color intensity reflecting its magnitude.



(a) $\boldsymbol{\beta}_{1,\mathrm{LF}_1}$

Table 2 reports results across a range of prior specifications for $\boldsymbol{\beta}_1$, enabling us to examine the robustness of sparsity patterns in factor loadings. We switch off the α channel and estimate a latent-factor model in which expected returns are fully explained by loadings and latent factor risk premia. Specifically, we consider factor dimensionalities $K \in \{1, 3, 5\}$. In Panel A, we impose Beta priors on q_β with means 0.1/0.5/0.9. In Panel B, we exogenously constrain the number of admissible characteristics per loading, $M_\beta \in \{2, 10, 18\}$. This analysis clarifies the extent to which firm characteristics inform the factor structure when mispricing is omitted.

Table 2: Model Performance under Different Priors (without Mispricing)

This table reports model results under various prior assumptions for sparsity, including cross-sectional R^2, Sharpe ratio from the latent factor tangency portfolio, and the posterior mean of the sparsity level. K indicates the number of latent factors. The priors on q_β indicate three Beta distributions: Beta$(9,1)$, Beta$(5,5)$, and Beta$(1,9)$, with means of 0.9, 0.5, and 0.1, respectively. Panel B restricts the number of characteristics driving beta: M_β limits the characteristics affecting each factor loading $\boldsymbol{\beta}_{1,k}$.

		CSR^2			TP. SR			q_β		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Unrestricted # selected chars.										
	0.9	20.4	52.8	58.1	0.51	0.73	0.69	0.62	0.59	0.60
q_β prior mean	0.5	20.5	53.2	58.3	0.51	0.73	0.99	0.48	0.43	0.50
	0.1	20.6	53.6	60.8	0.51	0.73	0.98	0.32	0.27	0.29
Panel B: Fixed # selected chars.										
	2	15.6	49.5	50.4	0.52	0.83	0.74	/	/	/
M_β	10	20.1	52.1	58.5	0.51	0.75	0.84	/	/	/
	18	19.9	53.8	59.6	0.51	0.44	0.59	/	/	/

Panels A and B of Table 2 show that when the number of latent factors is limited (e.g., $K=1$), models that learn characteristic relevance from the data outperform those with fixed inclusion sizes. However, this advantage diminishes as K increases, with richer models eventually outperforming sparser ones regardless of whether sparsity is learned or imposed.

Comparing Tables 1 and 2 reveals that excluding mispricing leads to lower posterior sparsity in factor loadings. For example, with a prior mean of 0.9 and $K=5$, q_β falls from approximately 0.66 to 0.60, reflecting the shift in explanatory burden from $\boldsymbol{\alpha}_1$ to $\boldsymbol{\beta}_1$. As a result, factor loadings become denser to accommodate the omitted pricing component. While cross-sectional R^2 remains high, reaching 60.8% without mispricing, the Sharpe ratios decline notably, with a maximum of 0.99 compared to over 1.3 when $\boldsymbol{\alpha}_1$ is included. This suggests that although models without mispricing can explain average returns, they deliver weaker economic performance, revealing a trade-off between density and fit, as well as between the Sharpe ratio and interpretability. Moreover, posterior q_β tends to stabilize at relatively low levels (0.27–0.62) across priors, indicating the limited viability of sparse representations in this setting.

Panel B also reveals a familiar inverted-U pattern: when $K=1$, fixed-sparsity models with very low or very high M_β tend to underperform those with intermediate inclusion sizes. This mirrors the pattern observed in Table 1 and further supports the

need to avoid rigid sparsity assumptions. [5] As K increases, the capacity of the loading channel to absorb information rises, the constraint imposed by fixed sparsity weakens, and relatively denser specifications become increasingly dominant. This dominance does not stem from superior uncertainty management; it reflects a greater ability to accommodate higher dimensionality.

Overall, the probabilistic framework performs best when both $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are active and learned jointly. A probabilistic framework that incorporates mispricing can achieve higher overall performance without sacrificing the Sharpe ratio by exploiting a channel-level division of labor, specifically by utilizing sparser alpha and richer beta. Conversely, forcing alpha equal to zero pushes structural differences into beta, which leads to overparameterization and a decline in trading value.

5 Schrödinger's Sparsity Everywhere

We have documented the presence of Schrödinger's Sparsity in the analysis of the 100 P-Tree test assets: the posterior probability of sparsity generally lies between the extremes of a highly sparse and a fully dense model. Furthermore, mispricing components consistently exhibit higher sparsity probabilities than factor loadings. In this section, we examine how these patterns vary across different settings. Section 5.1 analyzes alternative test assets and explores how their "complexity" relates to model-implied sparsity. Section 5.2 investigates the time variation in sparsity across macroeconomic regimes and evaluates how the estimated mispricing and factor loadings evolve over time and across asset classes.

5.1 Test Assets and Sparsity

Thus far, we have examined the probability of sparsity using the 100 P-Tree test assets. However, it is well recognized that the choice of test assets plays a pivotal role in asset pricing analysis. It determines the spanning and identifiability of the return space, and in turn, influences how model complexity is allocated across mispricing

[5] Table A.3 reports OOS performance for specifications that exclude mispricing. Models that estimate the sparsity probability endogenously consistently outperform counterparts that fix the sparsity level exogenously.

and factor loadings. We hypothesize that model sparsity systematically varies with the complexity and construction of test assets. Simpler portfolio sorts, such as the 25 ME/BM portfolios, rely on fewer characteristics to explain returns. In contrast, more complex or heterogeneous test assets, such as those constructed via the P-Tree framework (Cong, Feng, He, and He, 2025), may require a broader set of predictors due to their increased dimensionality.

To examine this hypothesis, we apply our model to three distinct asset groups: (i) P-Tree portfolios constructed with varying numbers of leaves; (ii) individual stocks split by average market equity (Big 500 vs. Small 500);[6] and (iii) benchmark portfolios including the 25 ME/BM, Bi360, and Uni610 sets. Table 3 reports the posterior sparsity probabilities, cross-sectional R^2, and Sharpe ratios across these settings under varying prior specifications.

Table 3: Sparsity for Different Test Assets

This table reports performance statistics across various test assets, focusing on models with three latent factors to ensure comparability. The analysis encompasses portfolios with varying numbers of P-Tree assets (Panel A), two groups of individual stocks with equal sample sizes but differing average market equity value (Panel B), and three widely used portfolio sets: ME/BM 25, Bi360, and Uni610 (Panel C). The reported metrics include cross-sectional R^2, tangency portfolio Sharpe ratio, and estimates of q_α and q_β.

	prior mean: (0.5, 0.5)			prior mean: (0.9, 0.9)			prior mean: (0.1, 0.1)		
	CSR^2	TP. SR	(q_α, q_β)	CSR^2	TP. SR	(q_α, q_β)	CSR^2	TP. SR	(q_α, q_β)
Panel A: P-Tree									
100	42.4	1.00	0.69,0.43	43.7	0.76	0.83,0.57	45.6	1.38	0.54,0.31
200	50.9	1.09	0.60,0.37	52.3	1.13	0.75,0.53	50.2	1.06	0.46,0.21
400	45.2	0.49	0.54,0.32	43.7	0.60	0.68,0.48	43.3	0.80	0.39,0.04
Panel B: Ind. Stock									
Big500	31.4	0.80	0.61,0.29	31.5	0.63	0.75,0.49	32.9	0.65	0.43,0.09
Small500	3.9	3.64	0.49,0.38	5.7	10.75	0.62,0.55	2.4	2.26	0.35,0.23
Panel C: Others									
ME/BM25	33.6	0.25	0.80,0.50	31.0	0.31	0.93,0.64	42.0	0.29	0.64,0.36
Bi360	7.8	1.03	0.50,0.20	10.8	0.40	0.65,0.33	5.9	0.65	0.36,0.04
Uni610	48.0	0.61	0.44,0.20	47.8	0.60	0.58,0.33	47.9	0.71	0.30,0.05

We begin with the baseline prior mean of 0.5 for both q_α and q_β. As the num-

[6]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.

ber of P-Tree portfolios increases from 100 to 400, the posterior mean of the sparsity probability in factor loadings declines from 0.43 to 0.37 and then to 0.32. Similarly, the mispricing sparsity probability drops from 0.69 to 0.60 and then to 0.54. 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. Consistent with the diminishing marginal contribution of additional portfolios to the tangency Sharpe ratio, the decline in sparsity also 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.

Next, we examine whether pricing difficulty influences sparsity while keeping the asset class and number of test assets constant. Panel B of Table 3 compares two groups of individual stocks, each comprising 500 securities: the 500 Big includes the largest firms by market equity, while the 500 Small consists of stocks ranked 501st to 1000th. For individual stocks, the underlying factor structure is latent and potentially complex compared to P-Tree portfolios, where construction characteristics are known. Despite this complexity, we find notable differences in sparsity estimates. For factor loadings, the posterior mean of q_β is 0.29 for large-cap stocks and 0.38 for small-cap stocks, indicating a denser factor structure for larger firms. In contrast, the mispricing sparsity probability (q_α) is 0.61 for the Big 500 and decreases to 0.49 for the Small 500, suggesting greater pricing difficulty among smaller firms.

These findings suggest that mispricing becomes less prevalent when transitioning from large- to small-cap stocks, potentially due to the greater cross-sectional heterogeneity of smaller firms, which makes them more susceptible to idiosyncratic pricing errors. This is consistent with prior evidence that small-cap stocks exhibit higher idiosyncratic volatility (e.g., Campbell et al., 2001), lower liquidity (e.g., Antweiler and Frank, 2004), and greater limits to arbitrage (e.g., Baker and Wurgler, 2006). Interestingly, the opposite holds for factor loadings: the model selects fewer characteristics to explain $\boldsymbol{\beta}_1$ for small-cap stocks, possibly reflecting their lower exposure to systematic

risk or more concentrated pricing channels (e.g., Daniel and Titman, 1997).

Finally, this pattern extends to other commonly used portfolio-based test assets. Panel C of Table 3 shows that the posterior sparsity probability for factor loadings is 0.50 for the 25 ME/BM portfolios, compared to 0.20 for the Bi360 and Uni610 portfolios. The ME/BM portfolios are constructed using only two firm characteristics, whereas the Bi360 and Uni610 portfolios incorporate a significantly 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 structure closely tracks the complexity and informational richness of the test assets.

Overall, Table 3 supports our hypothesis, demonstrating that the structure and complexity 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 demands a careful balance between parsimony and fit, while appropriately reflecting the complexity of the underlying return-generating process.

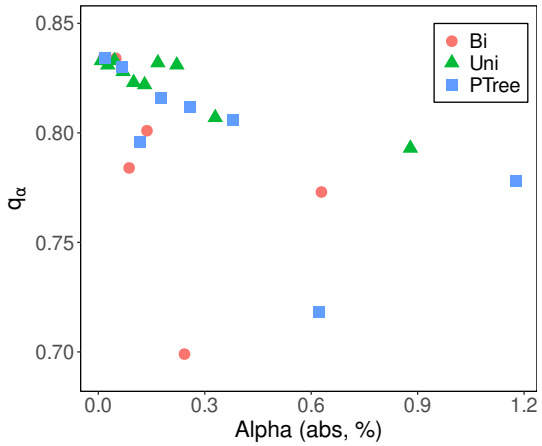
Alpha and Sharpe Ratio versus Sparsity. We have shown above that posterior sparsity probabilities differ across test assets. We now adopt a more granular and economically grounded approach to evaluating pricing difficulty and its implications for model sparsity. A natural question is, when a test-asset set is harder to price, does the model need to relax constraints in the mispricing or loadings, thereby exhibiting lower posterior sparsity?

Specifically, we use two empirical proxies: the absolute value of Jensen's alpha and the Sharpe ratio of the tangency portfolio. The former measures the magnitude of unconditional mispricing, corresponding to pricing difficulty in the mispricing itself, whereas the latter measures the strength of tradable risk compensation, corresponding to pricing difficulty in the factor components. To implement this, we sort the Bi360, Uni610, and P-Tree400 test assets into 6, 10, and 8 groups, respectively, according to the

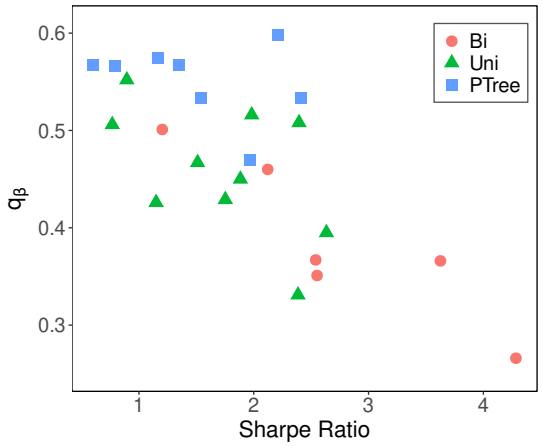
size of their CAPM alphas, resulting in group sizes of 60, 61, and 50 portfolios. For each group, we compute the tangency portfolio's Sharpe ratio and estimate our model with $K = 5$ latent factors to obtain posterior estimates of the sparsity probabilities q_α and q_β. Figure 3 presents the relationship between these learned sparsity probabilities — q_α for mispricing and q_β for factor loadings — and the associated measures of pricing difficulty.

Figure 3: Sparsity Levels and Pricing Difficulty of Test Assets

This figure illustrates the relationship between sparsity in alphas and betas and the pricing difficulty of test assets, as measured by absolute alpha and Sharpe ratio. Panel (a) depicts the relationship between q_α and the average absolute α of test assets. Panel (b) presents the relationship between q_β and the Sharpe ratio of the test assets. In both panels, circles, triangles, and squares denote bi-sorted, uni-sorted, and P-Tree portfolios, respectively.



(a) q_α and Alpha



(b) q_β and Sharpe Ratio

Both panels of Figure 3 reveal a consistent and economically intuitive pattern: test assets with higher absolute Jensen's alpha exhibit lower posterior sparsity in the mispricing component (q_α), while those with higher tangency portfolio Sharpe ratios exhibit lower sparsity in factor loadings (q_β).[7] Intuitively, when a test asset displays a large unconditional alpha, the model requires more flexibility in the alpha channel to account for persistent pricing errors, which leads to a denser $\boldsymbol{\alpha}_1$. Conversely, high Sharpe ratio portfolios reflect richer exposure to priced risks, prompting the model to draw on a broader set of characteristics to capture time-varying betas, thereby reducing sparsity in $\boldsymbol{\beta}_1$. The monotonic nature of these relationships — observed across

[7]Figure A.4 presents analogous results for the restricted model, demonstrating consistent relationships observed in the absence of mispricing.

bi-sorted, uni-sorted, and P-Tree portfolios — underscores that sparsity is not fixed but adapts systematically to pricing difficulty. These results provide direct visual evidence for our central thesis: the degree of sparsity should be treated as an object of inference, not imposed ex ante.

Characteristics Importance. We aim to identify the specific characteristics that drive mispricing and factor loadings. Figure 4 reports the posterior inclusion probabilities alongside the corresponding coefficient estimates — $\boldsymbol{\alpha}_{1,i}$ for mispricing and $\boldsymbol{\beta}_{1,i}$ for factor loadings — for each characteristic across the various test asset sets.

Several characteristics are consistently selected for factor loadings across all test asset sets, exhibiting posterior inclusion probabilities close to one. These characteristics include market beta (BETA), book-to-market ratio (BM), one-month momentum (MOM1M), twelve-month momentum (MOM12M), sales-to-price ratio (SP), and idiosyncratic volatility (SVAR). These variables are strongly linked to systematic risk, reflecting how asset returns respond to macroeconomic and market-wide fluctuations. Their consistently high inclusion probabilities across portfolios highlight their significance in explaining time-varying risk exposures.

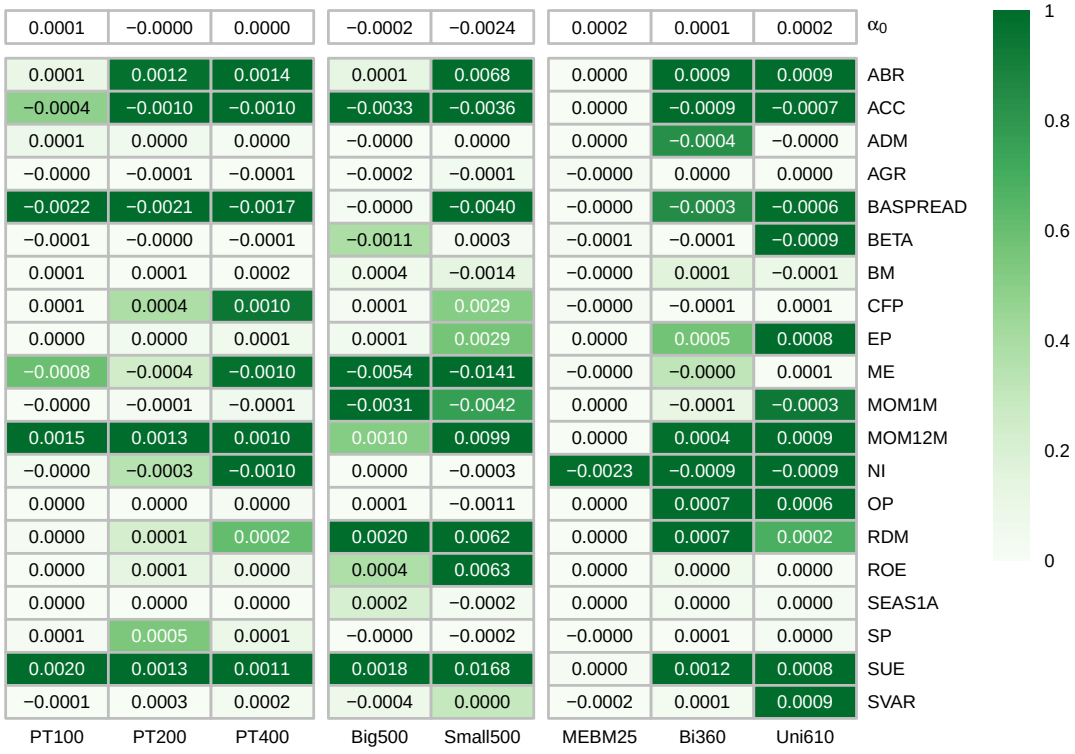
In contrast, characteristics such as accruals (ACC) and standardized unexpected earnings (SUE) are predominantly selected for mispricing, with high posterior inclusion probabilities across most test asset sets, except for the 25 ME/BM portfolios. These variables typically reflect short-term valuation errors or earnings-related anomalies (e.g., Sloan, 1996), making them more likely to influence the alpha component rather than factor loadings.

Notably, the characteristics driving mispricing and those driving factor exposures are largely disjoint, highlighting the model's ability to separate persistent pricing errors from systematic risk. This separation lends empirical support to the structural distinction between $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ in the model specification.

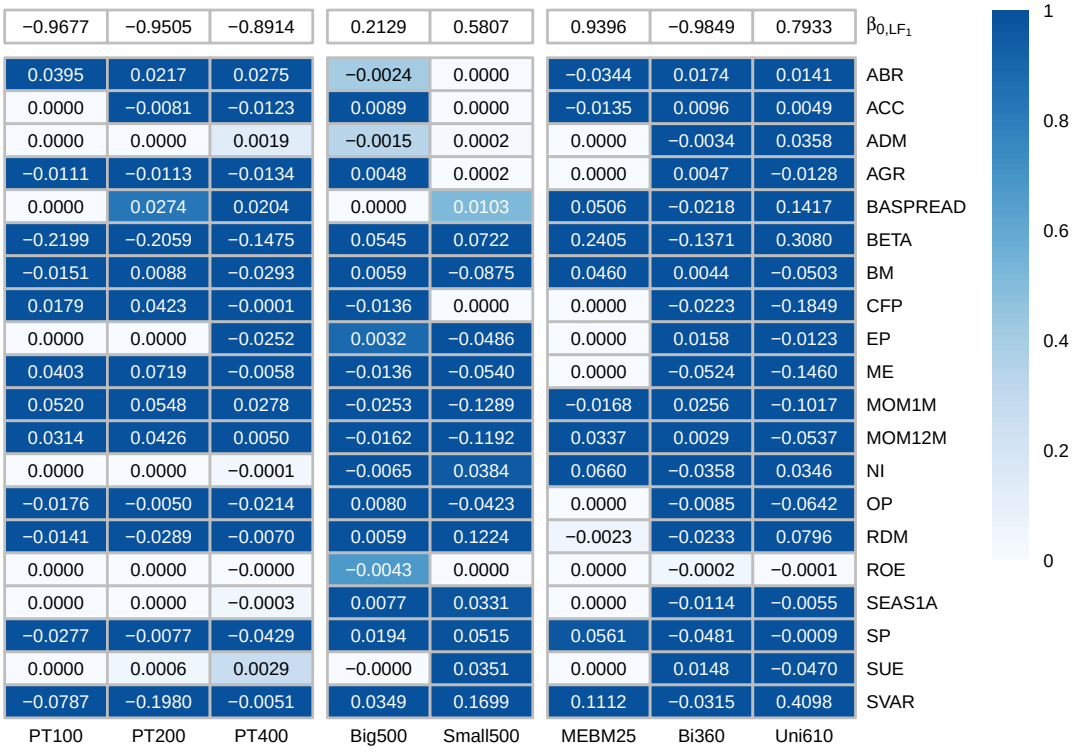
Several characteristics also exhibit context-dependent importance. For instance, asset growth (AGR), the cash flow-to-price ratio (CFP), and the R&D-to-market ratio (RDM) are not selected for the ME/BM 25 portfolios but contribute meaningfully to

Figure 4: Characteristics Importance in Different Test Assets

This figure depicts selection probabilities and coefficients of characteristics across test assets. Panel (a) reports the probability that each of the 20 characteristics is selected to explain mispricing, while Panel (b) shows the selection probability for explaining factor loadings. Results pertain to the first latent factor only. Each cell displays the selection probability, with color intensity reflecting its magnitude.



(a) α_1



(b) β_{1,LF_1}

return variation in Bi360, Uni610, and P-Tree portfolios.

This reinforces the idea that characteristic relevance depends on the informational complexity of the test assets. These movements are more pronounced on the loading side, indicating that as the test-asset universe becomes more complex and the number of priceable risk dimensions increases, the model must mobilize a broader set of characteristics on the loading, thereby lowering the posterior sparsity probability of beta. By contrast, the posterior sparsity probability of alpha need not decline mechanically; it only becomes necessary in settings where risk factors struggle to absorb the variation, such as Small500, to introduce a denser, characteristic-driven mispricing channel.

Overall, approximately 60% of the available characteristics are selected in at least one component, whether for mispricing or for factor loadings, across the test asset sets. The probabilistic nature of the Bayesian selection process allows the model to express uncertainty about inclusion while remaining interpretable. Furthermore, the estimated coefficients on the first latent factor reveal economically intuitive loadings — particularly for characteristics like momentum, value, and volatility — demonstrating that the latent factors capture meaningful structure in the cross section. [8]

To summarize, our analysis across multiple sets of test assets reveals that sparsity is not a fixed property, but rather it varies systematically with the complexity and pricing difficulty of the asset set. Specifically, we document a robust inverse relationship between posterior sparsity probabilities and pricing difficulty: test assets with higher absolute Jensen's alpha or higher tangency portfolio Sharpe ratios tend to exhibit lower levels of sparsity. In contrast, simpler and more homogeneous asset sets, such as the ME/BM 25 portfolios, exhibit higher sparsity levels, reflecting their limited need for complex explanatory structures.

These findings highlight that viewing asset pricing models as strictly sparse or

[8] Figure 2 presents the selection probabilities of characteristics driving factor loadings, along with their coefficients on the first factor, under a model specification without mispricing. While the selected characteristics remain fairly consistent, the estimated coefficients on the first factor vary considerably. This result indicates that imposing the mispricing restriction forces the model to absorb phenomena that should be attributed to mispricing through time-varying loadings, which renders beta denser and degrades its economic interpretability. Consequently, allowing for mispricing and estimating separate sparsity probabilities for the two channels is essential to prevent misattribution between risk and mispricing.

dense creates an overly simplistic dichotomy. Instead, sparsity emerges as a property to be inferred from the data. Our probabilistic framework provides a flexible and empirically grounded approach to modeling this variation, offering deeper insights into expected returns across diverse sets of test assets.

5.2 Macro Regimes and Sparsity Probability

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 firm 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 mispricing and factors, may itself be time-varying. In this section, we empirically examine whether sparsity changes across macroeconomic regimes and structural breaks, and assess whether different sets of firm characteristics contribute to return dynamics at different points in time.

To investigate this hypothesis, we apply our method across two regime classifications proposed in the literature. The first one follows the three structural regimes identified by Smith and Timmermann (2021). Panel A of Table 4 reports that the explanatory power of characteristics for both alphas and betas varies substantially across regimes. Notably, the number of characteristics driving mispricing declines over time: the posterior mean of q_α increases from 0.72 in the first regime to 0.77 in the third,

suggesting that fewer characteristics are relevant for capturing mispricing in recent decades. This trend may reflect increasing market efficiency and a diminished role for persistent pricing errors. Conversely, q_β falls from 0.56 to 0.46 over the same period, indicating that factor loadings have become less sparse and now depend on a richer set of characteristics. These findings suggest a growing complexity in risk exposures over time, even as mispricing becomes more concentrated.

Table 4: Time Variation Analysis: Sparsity in Structural Breaks / Business Cycles

The table displays performance statistics of P-Tree 100 test assets over different periods, focusing on models with three latent factors. Panel A considers structural breaks discovered in Smith and Timmermann (2021). Regime 1 covers Jan. 1990 to Jul. 1998, Regime 2 covers Sep. 1998 to Jun. 2010, and Regime 3 covers Jul. 2010 to Dec. 2024. Panel B consists of regimes by business cycles, in which the recession periods, totaling 88 months, are defined using Sahm Rule-designated phases: October 1990 - November 1992, May 2001 - October 2002, March 2008 - May 2010, March 2020 - March 2021, and June 2024 - September 2024. Panel C is the whole period from January 1990 to December 2024. The reported statistics include model performance metrics (cross-sectional R^2 and tangency Sharpe ratio), as well as estimates of q_α and q_β.

	CSR^2	TP. SR	(q_α, q_β)
Panel A: Sequential segmentation			
Regime1	48.5	1.92	0.72,0.56
Regime2	24.1	0.82	0.71,0.53
Regime3	59.7	0.72	0.77,0.46
Panel B: Macro-driven segmentation			
Normal	53.8	1.18	0.67,0.46
Recession	14.2	0.77	0.76,0.50
Panel C: Full period			
Whole	42.4	1.00	0.69,0.43

Second, while Smith and Timmermann (2021) identify structural regimes based on calendar time, we further examine how sparsity evolves across business cycles using the real-time Sahm Rule Recession Indicator. This measure classifies recessions as periods when the three-month average unemployment rate exceeds its 12-month minimum by at least 0.5 percentage points. Based on this classification, we estimate the model separately for recessionary and normal periods.

Panel B of Table 4 reveals a pronounced deterioration in model performance during recessions. The cross-sectional R^2 drops sharply from 53.8% in normal periods to

just 14.2%, and the Sharpe ratio declines from 1.18 to 0.77. These shifts reflect a contraction in return predictability and pricing efficiency under macroeconomic stress.

At the same time, posterior sparsity probabilities change significantly during recessions. Specifically, q_α rises from 0.67 to 0.76, indicating that fewer firm-level characteristics are needed to explain mispricing. Meanwhile, q_β increases modestly from 0.46 to 0.50, suggesting a slightly denser factor structure. These results suggest that recessions diminish the explanatory power of firm fundamentals for pricing errors, while increasing reliance on systematic exposures.

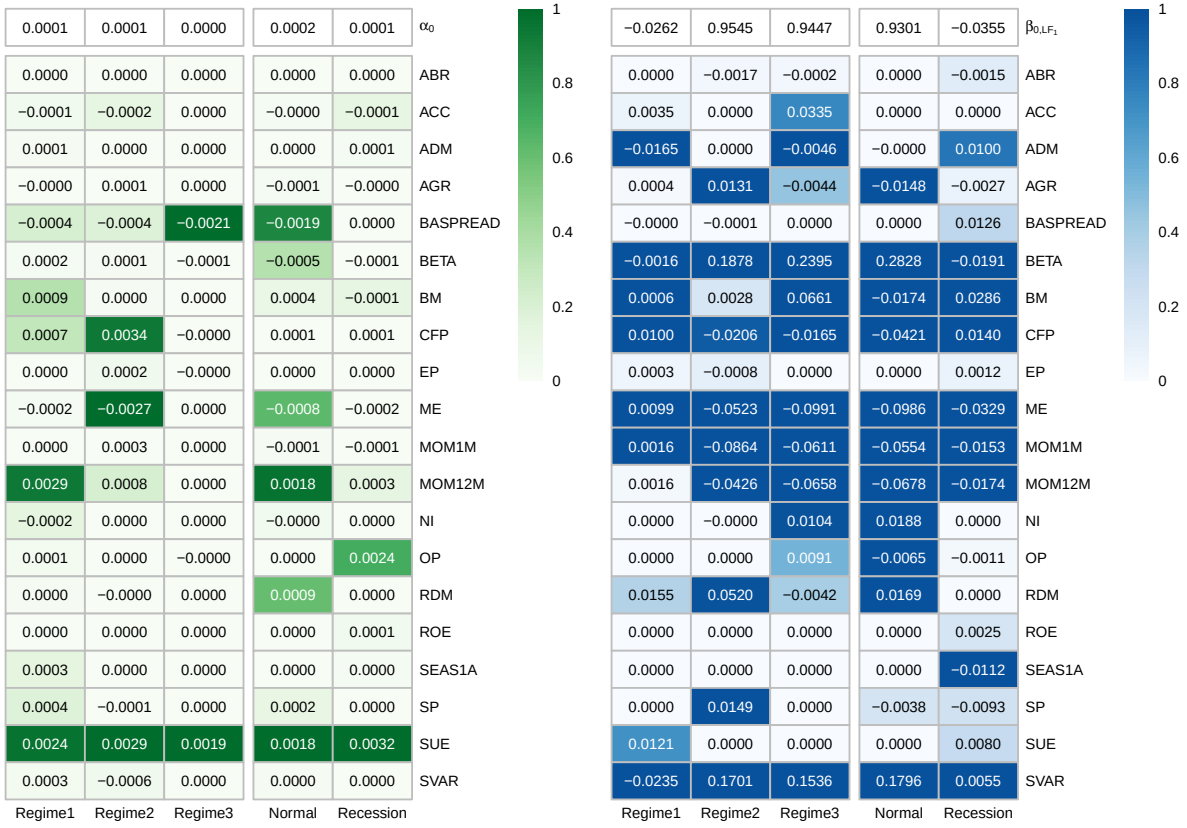
Several mechanisms may explain this pattern. (i) Heightened uncertainty can shift investor focus from firm-specific variables to broader macro-level risks, reducing the relevance of cross-sectional predictors (Avramov et al., 2007). (ii) During downturns, increased risk aversion may lead to a reallocation of capital toward safer assets (Baker and Wurgler, 2007) or amplify demand for downside protection (Avramov et al., 2022), thereby weakening firm-level predictors. (iii) Large macroeconomic shocks can disrupt the firm-level return-generation process (Avramov et al., 2013), undermining predictability during crises (Chordia and Shivakumar, 2002).

Regime-based characteristics importance. Having documented clear time dependence in sparsity probabilities, we now ask: across different macro regimes, which firm characteristics enter the mispricing and which enter the factor loadings? How are their weights reallocated as regimes switch? Thus, we further investigate which specific firm characteristics are relevant across different structural regimes. Figure 5 illustrates the evolution of the explanatory power of selected characteristics for mispricing from different regimes.

Notably, the influence of 12-month momentum as a driver of mispricing steadily declines across regimes, as evidenced by its decreasing posterior probability and coefficient. In contrast, during Regime 2, which encompasses both the 1997 Asian Financial Crisis and the 2008 Global Financial Crisis, additional characteristics emerge as significant contributors to mispricing. Specifically, the cash flow-to-price ratio (`CFP`) and market equity (`ME`) exhibit elevated posterior inclusion probabilities and relatively

Figure 5: Regime-based Characteristics Importance

This figure presents selection probabilities and coefficients of characteristics over regimes by structural breaks or business cycles. Panel (a) reports the selection probabilities for each of the 20 characteristics explaining mispricing, alongside the corresponding $\alpha_{1,i}$ coefficient estimates. Panel (b) shows selection probabilities for characteristics explaining factor loadings, together with the associated $\beta_{1,i}$ estimates. For brevity, results are shown for the first latent factor only. Each cell displays the coefficient for the characteristic, with color intensity reflecting the magnitude of the selection probability.



(a) α_1 (b) β_{1,LF_1}

large coefficients. This pattern suggests that in periods of elevated uncertainty and macroeconomic stress, investors may place greater weight on firms' cash-generating ability and size as signals of resilience. Firms with strong cash flow fundamentals likely offered more predictable earnings streams (Koijen and Van Nieuwerburgh, 2011), while large-cap stocks were perceived as safer investments (Ang et al., 2006), potentially contributing to their positive excess returns during this turbulent regime.

Furthermore, under recession and non-recession regimes as defined by the Sahm Rule Recession Indicator, we observe a significant shift in the drivers of mispricing. During recessionary periods, both the bid-ask spread (BASPREAD), a proxy for liq-

uidity costs, and 12-month momentum (MOM12M) exhibit substantially lower posterior inclusion probabilities. This decline suggests that in times of heightened macroeconomic stress, investors de-emphasize liquidity-related frictions and past return trends, instead prioritizing macroeconomic indicators and firm fundamentals. The reduced relevance of BASPREAD likely reflects the declining role of trading costs in pricing, as market participants shift focus to systemic risk and solvency (Pástor and Stambaugh, 2003). Similarly, the attenuated explanatory power of momentum during recessions is consistent with a shift in investor preferences toward more stable and defensive assets, reducing reliance on trend-following strategies that may underperform in volatile or risk-averse environments (Chordia and Shivakumar, 2002).

In addition, characteristics such as asset growth (AGR), net stock issues (NI), operating profitability (OP) and R&D-to-market equity (RDM) lose their ability to explain systematic risk. This likely reflects the fact that, during recessions, returns become increasingly driven by latent macroeconomic factors, thereby weakening the role of firm-specific characteristics in capturing variation in dynamic factor loadings. [9]

Taken together, the evidence indicates that sparsity is both regime- and cycle-dependent. Along a secular trend of efficiency improvement, alpha becomes more concentrated; under macroeconomic stress, beta becomes denser, and alpha is less systematically explained by firm-level characteristics. Treating sparsity as a learnable latent quantity and updating it over time in the posterior yields a more robust trade-off between statistical fit and economic interpretation.

6 Conditional Model for Observable Factors and Sparsity

We estimate separate models for each test asset set, deriving posterior probabilities of sparsity specific to each dataset. The latent factors vary across specifications, reflecting the distinct structure of each dataset. This section examines how factor specification affects the posterior probability of sparsity, with a focus on the role of

[9] Figure A.5 reports the selection probabilities of characteristics driving factor loadings and their corresponding coefficients on the first factor in the absence of mispricing. While the selected characteristics remain fairly consistent, the estimated coefficients on the first factor exhibit significant variability.

observable factors. Section 6.1 evaluates the impact of replacing latent factors with observable ones, while Section 6.2 examines the integration of both observable and latent factors in a unified framework. Our methodology flexibly accommodates models based on observable factors, latent factors, or a combination of both.

6.1 Observable Factors and Sparsity

To examine whether the sparsity patterns documented earlier persist under a common factor structure, we extend our framework by replacing the asset-specific latent factors with pre-specified factors shared across all models. This allows sparsity to emerge endogenously within the Bayesian framework, helping to separate differences in factor structure from the learning of sparsity probabilities and thereby highlighting the transferability and robustness of posterior sparsity.

We consider two categories of pre-specified factors. The first category includes widely used observable factors, such as the market factor and the Fama–French five factors. The second category consists of a five-factor specification extracted via IPCA, estimated from individual stock returns to maximize its ability to capture information embedded in the equity market. For the IPCA specification, we analyze two versions: one that incorporates mispricing and one that excludes it, corresponding to Eqs. (2) and (3), respectively. These factors are integrated into our Bayesian framework, where sparsity patterns are determined endogenously, yielding conditional factor models. Following Kelly et al. (2019), we report TotalR^2 in Table 5.[10]

We examine whether the sparsity probability patterns documented in Section 5.1 persist when the type of test assets is fixed but their number varies. Panel A of Table 5 reveals a robust trend: within the same type of test assets, the probability of sparsity declines as the number of assets increases. The magnitude of these probabilities varies across models, reflecting differences in the properties of the factors. Specifically, probabilities for CAPM and IPCA (WithM) are lower than those for FF5 and IPCA (WithoutM). For CAPM, the limited explanatory power of its single-factor structure

[10] $\text{Total}R^2 = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(r_{i,t}-\widehat{r}_{i,t})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(r_{i,t}-\hat{\beta}_i \text{MktRF}_t\right)^2}$.

likely necessitates capturing a larger portion of expected returns through mispricing, which, in turn, requires more characteristics to explain it. Similarly, for IPCA (WithM), incorporating mispricing explicitly during factor estimation produces factors that emphasize variation related to the factors themselves. Consequently, when applied in other models, additional characteristics are needed to account for the mispricing component. By contrast, FF5 and IPCA (WithoutM) yield greater sparsity in mirpricing. FF5 expands the span of the observed factor space. IPCA (WithoutM) disallows mispricing at the extraction stage, so the structure that would otherwise be attributed to mispricing is partially forced into the factor loadings. As a result, within our conditional model, the residual burden carried by alpha is systematically reduced.

As pricing difficulty increases, the sparsity of alpha and beta diverges in a complementary way; however, factor-extraction procedures that disallow mispricing can reverse this pattern. We analyze cases where the test assets share the same type and number but differ in pricing difficulty. Panel B of Table 5 reveals that for CAPM, FF5, and IPCA (WithM), small-stock portfolios exhibit a lower probability of sparsity in mispricing compared to the 500 big-stock portfolios. Conversely, the probability of sparsity in factor loadings is higher for small-stock portfolios. This pattern aligns with our earlier findings. In contrast, the IPCA (WithoutM) model shows a distinct pattern in mispricing. This difference may arise because the IPCA (WithoutM) factors incorporate mispricing information during estimation. As a result, when these factors are incorporated into other models, the corresponding mispricing component is reduced. Put differently, whether mispricing is accommodated at the factor-estimation stage determines the burden borne by alpha in the subsequent conditional model. This observation is consistent with the difference in q_α between IPCA (WithM) and IPCA (WithoutM), as noted in Panel A.

Third, we assess whether the estimated probabilities of sparsity differ across commonly used test asset sets. Panel C reveals substantial variation in sparsity probabilities across test assets, even for conditional pre-specified factor models. Notably, models estimated on ME/BM25 portfolios consistently display the highest probability

Table 5: Sparsity for Different Pre-specified Factors

This table reports performance statistics for models with the market factor (CAPM), the Fama–French five factors (FF5), the IPCA five factors estimated with mispricing (IPCA LF5 (WithM)), and the IPCA five factors estimated without mispricing (IPCA LF5 (WithoutM)). The analysis encompasses portfolios with varying numbers of P-Tree assets (Panel A), two groups of individual stocks with equal sample sizes but differing average market equity value (Panel B), and three widely used portfolio sets: ME/BM 25, Bi360, and Uni610 (Panel C). Reported metrics include the total R^2 and estimates of q_α and q_β.

	CAPM		FF5		IPCA LF5 (WithM)		IPCA LF5 (WithoutM)	
	TotalR^2	(q_α, q_β)	TotalR^2	(q_α, q_β)	TotalR^2	(q_α, q_β)	TotalR^2	(q_α, q_β)
Panel A: P-Tree								
100	1.9	0.55,0.37	31.2	0.74,0.21	35.6	0.54,0.22	34.5	0.75,0.20
200	1.7	0.48,0.31	31.5	0.69,0.17	35.8	0.47,0.21	35.0	0.62,0.19
400	2.0	0.43,0.32	30.5	0.65,0.17	30.8	0.38,0.17	29.9	0.58,0.17
Panel B: Ind. Stock								
500 big	0.0	0.56,0.31	6.0	0.58,0.17	5.9	0.63,0.17	5.7	0.62,0.16
500 small	-1.0	0.44,0.39	3.0	0.47,0.22	8.5	0.54,0.17	8.7	0.68,0.17
Panel C: Others								
ME/BM25	1.6	0.72,0.56	47.1	0.77,0.20	76.8	0.78,0.18	76.3	0.78,0.19
Bi360	1.6	0.32,0.25	38.4	0.53,0.17	63.6	0.5,0.17	62.8	0.43,0.17
Uni610	1.1	0.43,0.31	65.9	0.55,0.17	20.9	0.46,0.17	19.5	0.46,0.17

of being sparse.

A broader pattern emerges: models with a larger number of factors exhibit lower probabilities of sparsity in factor loadings. For instance, in the P-Tree100 specification, the posterior mean of q_β is 0.37 for CAPM but only about 0.2 for the other three models. Similarly, for ME/BM25 portfolios, CAPM's q_β is 0.56, compared with roughly 0.2 for the others. This suggests that simpler models rely more heavily on individual asset characteristics to capture price dynamics, whereas richer factor structures reduce this reliance by incorporating broader systematic components. In parallel, disallowing mispricing at the factor-extraction stage further reduces the complexity allocated to the mispricing within the conditional model.

Further evidence supports the coexistence of sparse and dense structures in the cross-section of asset prices. Using pre-specified factors consistent across models—rather than latent factors estimated for each test asset set—produces similar sparsity patterns. This consistency indicates that these patterns are not driven by dataset-specific latent factors but instead reflect fundamental characteristics of cross-sectional asset price dy-

namics.

6.2 Augmented Conditional Model for Observable Factors

A foundational question in asset pricing is whether traditional observable factors, such as those in the CAPM or Fama-French models, are sufficient to explain systematic return variation, or whether latent factors are necessary to account for the residual pricing structure. This section examines this question using a conditional Bayesian framework, which jointly models latent and observed risk exposures while allowing sparsity patterns to emerge endogenously. This design decouples the choice of factor specification from the choice of sparsity, enabling a direct assessment of their marginal complementarity and substitutability.

Rather than treating observable and latent factors as competing alternatives, we investigate their complementary roles. Observable factors capture well-understood risks, whereas latent factors address omitted structures or nonlinear pricing effects that traditional models cannot span. We consider three specifications: models with observed factors only, models with latent factors only, and models that jointly include both observed and latent factors. This design offers a nuanced perspective on how combining known and previously hidden sources of systematic variation affects model performance and the explanatory power of firm characteristics.

Table 6 presents results from models estimated on the P-Tree 100 portfolios. Panels A–C explore conditional models with varying inclusion of observed and latent factors, while Panel D provides unconditional benchmarks for comparison. Across these specifications, we evaluate how characteristic-based sparsity adapts to expanded information sets, shedding light on the interplay between factor exposures and systematic return variation.

First, we evaluate the baseline impact of observable and latent factors in isolation. Panel A demonstrates that conditioning on observable factors alone, particularly FF3 and FF5, substantially improves model fit relative to MKT. Specifically, CSR^2 increases from 14.9% to 50.4%, accompanied by a reduction in α RMSE. The sparsity patterns indicate that FF models induce relatively sparse α_1 and highly sparse β_1,

Table 6: Augmented Conditional Observable Factor Models

This table presents estimation metrics from Sparse Bayesian IPCA and asset-specific regressions across model specifications using P-Tree 100 portfolios. The " $\boldsymbol{\alpha}$ RMSE" metric captures the root mean squared error between mispricing ($\alpha_0 + \boldsymbol{\alpha}_1(\mathbf{Z}_{i,t-1})$) and zero. For brevity, " LF " denotes latent factors, with the number indicating their count. Panel A presents the results from a conditional model that uses only observable factors. Panel B shows the results using only latent factors. Panel C considers a model that incorporates both observable and latent factors. Panel D presents results from asset-specific regressions, with separate regressions estimated for each asset and aggregated for overall performance.

	CSR^2	TP.SR	(q_α, q_β)	$\boldsymbol{\alpha} RMSE$
Panel A: only obs				
MKT	14.9	0.57	0.55,0.37	0.0032
FF3	27.3	0.60	0.65,0.26	0.0026
FF5	50.4	1.13	0.74,0.39	0.0014
Panel B: only latent				
LF1	29.5	0.35	0.52,0.47	0.0036
LF3	45.0	1.00	0.68,0.58	0.0021
LF5	56.8	1.02	0.77,0.66	0.0011
Panel C: obs + latent				
MKT+LF1	53.9	0.87	0.69,0.35	0.0015
MKT+LF3	53.8	0.85	0.78,0.41	0.0006
MKT+LF5	56.5	1.06	0.79,0.48	0.0005
FF3+LF1	41.6	1.07	0.67,0.27	0.0014
FF3+LF3	53.5	1.03	0.82,0.34	0.0001
FF3+LF5	57.4	1.26	0.80,0.56	0.0003
FF5+LF1	50.6	1.21	0.67,0.35	0.0012
FF5+LF3	53.2	1.41	0.80,0.42	0.0004
FF5+LF5	55.8	1.25	0.79,0.58	0.0005
Panel D: uncond. model				
MKT	/	0.57	/	0.0060
FF3	11.5	0.60	/	0.0056
FF5	49.3	1.13	/	0.0042

suggesting that observed factors capture some risk premia while leaving considerable characteristic-level heterogeneity unexplained.

In contrast, Panel B highlights the superior explanatory power of models incorporating latent factors. For example, increasing latent factors from LF1 to LF5 increases CSR^2 from 29.5% to 56.8%. This improvement reflects the ability of latent factors to capture more systematic variation in returns. Moreover, as more return variation is explained by factors, the density of $\boldsymbol{\beta}_1$ sparsity decreases, reducing reliance on mis-

pricing or spurious factor loadings. This endogenous shift exemplifies the concept of *Schrödinger's Sparsity*, where the sparsity structure evolves dynamically as the model's factor span deepens.

Second, Panel C examines models that jointly incorporate observable and latent factors, and the results indicate pronounced complementarity. Adding latent factors to observable-only models consistently enhances pricing metrics. For instance, FF5 alone achieves CSR^2 of 50.4%, while FF5 + LF5 increases CSR^2 to 55.8% and reduces $\boldsymbol{\alpha}$ RMSE from 0.0014 to 0.0005. Furthermore, sparsity reallocates toward systematic components: across MKT+LF and FF+LF combinations, q_β increases while q_α stabilizes or rises modestly. Intuitively, latent factors absorb systematic variation not spanned by observed factors, thereby reducing the dependence of factor loadings on characteristics.

Third, we emphasize the informational value of conditioning, beyond raw performance. Comparing conditional and unconditional specifications in Panels A vs. D, the CSR^2 of FF3 rises from 11.5% to 27.3%, while $\boldsymbol{\alpha}$ RMSE drops by over 50%. The unconditional regressions based solely on observed factors tend to misattribute part of the time-varying exposure to the mispricing, whereas conditioning reallocates those components back to the factor channel. These gains are further amplified when latent factors are considered. The results suggest that observable-only models leave residual pricing structure in the cross section, which our framework recovers through latent factor learning and adaptive sparsity.

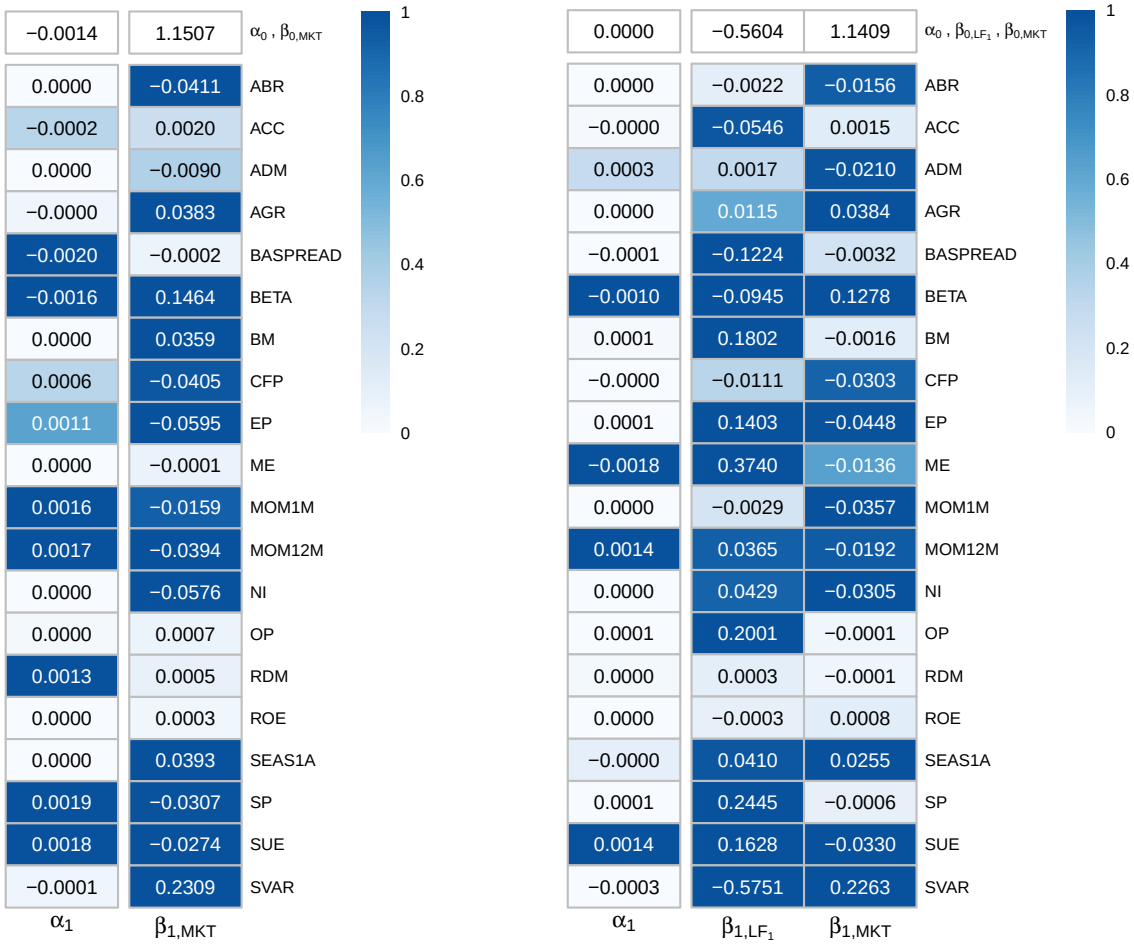
Figure 6 further illustrates how latent factors reconfigure sparsity across characteristics. Under the MKT-only model, variables such as `BM` and `MOM12M` dominate the market beta channel, while others (e.g., `ACC`) contribute minimally to it. Adding a latent factor shifts the explanatory weight: `ADM`, `SP`, and `SVAR` gain prominence, while previously dominant variables diminish in importance. This reflects a richer allocation of explanatory roles — an effect only visible through joint modeling of conditional factor structure and probabilistic sparsity.

Finally, this section outlines a promising direction for future research. Rather than

enforcing a unified sparsity prior on observed and latent components, future studies could estimate separate sparsity processes for each. This approach may yield sharper insights into the distinction between known and unknown sources of systematic risk.

Figure 6: Characteristics Importance in Conditional Factor Models

This figure presents coefficients and posterior selection probabilities for characteristics driving alphas (mispricing) and betas (factor loadings) under two specifications: (a) market factor only and (b) market plus latent factor. The posterior probability of a characteristic contributing to alphas is denoted by α_1. Columns $\beta_{1,\,\mathrm{MKT}}$ and $\beta_{1,\,\mathrm{LF}_1}$ capture posterior selection probabilities for the market factor and the first latent factor, respectively. Each cell displays the corresponding coefficient ($\boldsymbol{\alpha}_{1,l}$ or $\boldsymbol{\beta}_{1,l}$), with color intensity reflecting the magnitude of selection probability.



(a) MKT

(b) MKT + LF1

7 Implications for Out-of-sample Predictive Accuracy

Earlier empirical analysis shows that posterior sparsity adaptively reallocates model complexity between mispricing and factor loadings, materially affecting both pricing performance and interpretability. We now verify that this learned sparsity is not an

in-sample mirage but also operates out of sample while minimizing predictive losses arising from structural and parameter uncertainty. To this end, we conduct two out-of-sample exercises. Section 7.1, on the model side, fixes the test-asset universe and compares multiple specifications to assess whether endogenously estimated sparsity probabilities deliver generalization, prior robustness, and complexity matching. Section 7.2, on the data side, varies the type and complexity of the test assets to examine whether, as cross-sectional heterogeneity and dimensionality increase, sparsity learning continues to align model complexity with the asset set's information-carrying capacity and translates into stable OOS gains. To preserve identification and estimation stability, recession segments in the time-series dimension are too short; therefore, we do not implement a separate out-of-sample exercise.

7.1 Learning Sparsity, Prior Robustness, and Complexity Matching

The previous results in Section 4 indicate two key points: (i) the posterior distribution of q_β rarely concentrates at high values, and (ii) models with pre-specified sparsity levels exhibit sizable differences in pricing performance, implying substantial model uncertainty. A direct implication is that exogenous sparsity-based procedures, which fix the set of predictors and ignore model uncertainty, are prone to predictive losses. Extrapolation error in our cases primarily arises from two sources of uncertainty: structural uncertainty (how sparse alpha and beta should be) and parameter uncertainty (noise amplification under weak signals). Consequently, if sparsity is learned from the data rather than fixed in advance, we should observe improvements along three dimensions: (i) generalization, that is, whether learning sparsity consistently prevails on CSR^2; (ii) prior robustness, meaning that within a reasonable family of priors, OOS metrics vary only modestly across prior changes; and (iii) complexity matching, whereby as the number of factors K increases, information is reallocated between the mispricing and factor loading channels and the reallocation is reflected in OOS performance rather than being merely an in-sample tuning artifact. We show these by comparing the out-of-sample (OOS) performance across alternative model specifications.

To test the three predictions above, we implement an evaluation protocol with three modules. First, generalization: we estimate the Bayesian with learning / fixed sparsity, Bayesian model without sparsity, and IPCA model on a common training set and compare out-of-sample CSR^2 and TP.SR on a held-out window. Second, prior robustness: we vary the prior means for q_α and q_β over $0.1, 0.5, 0.9$ and assess the stability of OOS metrics and posterior sparsity. Third, complexity matching: we vary the factor dimension $K \in 1, 5$ and examine how information reallocation between the α and β channels affects OOS performance. Throughout, preprocessing and the test-asset universe are held fixed, and inference is summarized by posterior dominance probabilities and effect sizes. Table 7 reports both in-sample (INS) and OOS performance across four distinct model specifications (consistent with those in Table 1), using the first 15 years of the 35-year sample for estimation and the subsequent 20 years for evaluation.

Focusing on cross-sectional model fitness, when $K = 5$, the specification that learns the sparsity level delivers the strongest OOS performance ($CSR^2 = 68.7\%$). This model ((q_α prior mean, q_β prior mean) = (0.9,0.1)) demonstrates INS posterior means of 0.94 and 0.31, respectively. It also shows the highest alpha-sparsity probability and the lowest beta-sparsity probability among all $K = 5$ specifications in Panel A, aligning with the best-performing model in the full sample.

Beyond this result, we establish three interpretations. First, models that learn sparsity exhibit stronger prior robustness, thereby attenuating the influence of model uncertainty relative to exogenously fixed-sparsity alternatives. Panel A shows that the models perform similarly under different priors for the sparsity probabilities. While prior choices do matter, their impact is limited: when $K = 1$, the largest OOS CSR^2 difference is only $29.1 - 28.3 = 0.8(\%)$. By contrast, Panel B reveals sizable performance gaps across models that fix different sparsity levels: when $K = 1$, the largest OOS CSR^2 difference is $30.1 - 17.7 = 12.9(\%)$. These results highlight the cost of ignoring model uncertainty. Since the true sparsity probabilities are unknown, learning them endogenously using a neutral prior with means set to 0.5 for both q_α and q_β is

Table 7: Out-of-sample Model Performance under Different Priors

This table presents in-sample (INS) and out-of-sample (OOS) results under varying prior assumptions for sparsity on the P-Tree 100 test assets. Metrics include cross-sectional R^2, Sharpe ratio of the latent factor tangency portfolio, and the posterior mean of INS sparsity probabilities. K denotes the number of latent factors. In Panel A, the row labeled "(q_α prior mean, q_β prior mean)" reflects unconstrained sparsity priors, where q_α and q_β follow Beta distributions: Beta$(9,1)$, Beta$(5,5)$, and Beta$(1,9)$ with means of 0.9, 0.5, and 0.1, respectively. Panel B imposes restrictions on characteristics influencing $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$. Specifically, M_α limits characteristics impacting $\boldsymbol{\alpha}_1$, while M_β constrains those affecting each factor loading $\boldsymbol{\beta}_{1,k}$. Panel C reports Bayesian estimates without sparsity constraints, whereas Panel D provides results using the standard (dense) IPCA methodology.

		INS						OOS			
		CSR^2		TP. SR		(q_α, q_β)		CSR^2		TP. SR	
		$K=1$	$K=5$	$K=1$	$K=5$	$K=1$	$K=5$	$K=1$	$K=5$	$K=1$	$K=5$
Panel A: Unrestricted # selected chars.											
(q_α prior mean, q_β prior mean)	0.9,0.9	16.0	52.9	0.37	1.54	0.72,0.60	0.94,0.60	28.3	68.1	0.39	0.71
	0.5,0.9	16.4	53.1	0.36	1.46	0.56,0.60	0.79,0.60	28.9	67.5	0.38	0.76
	0.1,0.9	16.5	51.3	0.35	1.61	0.41,0.60	0.62,0.61	29.0	66.2	0.38	0.75
	0.9,0.5	16.2	53.0	0.36	1.40	0.72,0.44	0.94,0.47	28.6	67.9	0.39	0.64
	0.5,0.5	16.4	53.8	0.36	1.43	0.57,0.44	0.80,0.44	28.9	68.1	0.38	0.88
	0.1,0.5	16.6	55.4	0.35	1.41	0.41,0.44	0.65,0.44	29.0	67.3	0.38	0.71
	0.9,0.1	16.3	53.3	0.36	1.49	0.72,0.27	**0.94,0.31**	28.7	68.7	0.39	0.88
	0.5,0.1	16.5	52.7	0.36	1.51	0.57,0.27	0.78,0.33	29.0	67.8	0.38	0.72
	0.1,0.1	16.7	52.1	0.35	1.55	0.41,0.27	0.62,0.33	29.1	66.2	0.38	0.85
Panel B: Fixed # selected chars.											
(M_α, M_β)	2,2	11.1	30.9	0.42	0.58	/	/	19.9	62.0	0.47	0.07
	10,2	14.0	20.6	0.37	0.76	/	/	29.9	48.1	0.41	0.20
	18,2	14.6	8.5	0.35	0.48	/	/	30.6	38.1	0.40	0.05
	2,10	14.7	51.7	0.41	1.62	/	/	19.2	64.9	0.46	0.65
	10,10	15.6	35.2	0.35	2.11	/	/	28.4	51.9	0.38	1.38
	18,10	16.5	29.7	0.33	1.78	/	/	30.1	48.9	0.37	0.43
	2,18	14.7	47.6	0.41	1.31	/	/	17.7	66.3	0.46	0.58
	10,18	17.9	43.9	0.34	1.72	/	/	29.6	57.0	0.37	0.98
	18,18	17.5	28.4	0.32	4.13	/	/	29.5	45.0	0.36	0.04
Panel C: No sparsity											
(M_α, M_β)	20,20	16.9	39.6	0.28	0.95	/	/	27.8	56.5	0.32	0.66
Panel D: IPCA											
(M_α, M_β)	20,20	21.1	37.8	0.32	0.57	/	/	8.1	36.0	0.37	0.49

reasonable. Even if this prior is far from the truth, its influence on realized performance remains limited. In contrast, pre-specifying the sparsity level might yield poor model performance. Put differently, sparsity is a learnable latent quantity: the data pull the posterior toward an interior region favored by the evidence, and the prior is a starting point rather than a conclusion. By contrast, exogenously fixing sparsity is often misaligned with the evaluation-window regime and thereby translates directly into predictive loss.

Moreover, the functional division between mispricing and factor loadings is validated out of sample, implying that complexity matching is borne out by out-of-sample evidence. When K increases to 5, the best OOS specification corresponds to a posterior

with "sparser alpha and richer beta" (INS posterior means approximately $(0.94, 0.31)$), consistent with our mechanism: a higher factor dimension allows the factor loading channel to absorb a larger share of systematic variation, so the mispricing channel need not bear components that factor loading can explain. Posterior sparsity learning further implements spike selection and hard thresholding on alpha, while applying mild shrinkage to beta, resulting in a complexity allocation that is tilted toward beta. This reallocation of information across channels is clearly reflected in the OOS metrics: CSR^2 improves systematically as K increases.

In addition, indiscriminate densification and reliance on non-averaged point estimates both underperform out of sample. Comparing Panels C and D (without sparsity) with Panels A and B (with sparsity) demonstrates that incorporating sparsity is crucial for estimating asset pricing models. A fully dense Bayesian approach lacks disciplined exclusion of weak signals and tends to parameterize noise, leading to overfitting within the evaluation window. IPCA does not assign probabilistic weights over the model space and imposes no structured coefficient shrinkage. Thus, its point estimates can conflate uncertainty with predictive heterogeneity, thereby amplifying the error of extrapolation.

Overall, these findings highlight the pervasiveness of model uncertainty and the importance of probabilistically learning sparsity, support the central proposition of Schrödinger's sparsity. Now we have reported both full-sample and OOS performance across alternative model specifications. Our objective is to demonstrate that endogenously estimating the sparsity probability is a viable and coherent approach for estimating asset pricing models. Predictive gains stem from disciplined management of model uncertainty via shrinkage and averaging, rather than from merely stacking additional parameters or fixing sparsity ad hoc.

7.2 OOS Performance across Test Assets: Complexity, Sparsity Discipline and Predictability

Since the construction of test assets is closely tied to the sparsity probability and the predictability, we ask whether sparsity learning can, as cross-sectional complexity

varies, continue to align model complexity with the asset set's information-carrying capacity and maintain stable predictive performance out of sample.

Consistent with the design in Section 5.1, we cover three classes of test assets. [11] Table 8 reports, for the three prior-mean pairs $(0.5, 0.5)$, $(0.9, 0.9)$, and $(0.1, 0.1)$, the in-sample CSR^2 and the posterior means of (q_α, q_β), together with the out-of-sample CSR^2.

Table 8 conveys three stable findings. First, within a given test-asset class, OOS predictive power is clearly nonmonotonic in cross-sectional complexity. When the P-Tree portfolios increase from 100 to 200, OOS CSR^2 rises markedly, whereas expanding from 200 to 400 brings it back to roughly 60. The corresponding INS sparsity adjusts in parallel: as the number of assets grows, q_α and q_β decline overall (for a prior mean of (0.5, 0.5), from 0.68 to 0.43 and from 0.57 to 0.31), indicating that the model must mobilize a broader set of characteristics to carry information. Once the cross-section reaches 400, additional noise and redundancy offset the gains, and OOS performance falls. This pattern of initial improvement, saturation, and subsequent decline illustrates the concept of complexity matching: there exists an interior region where the information load and the sparsity constraint are most closely aligned. Importantly, this conclusion is robust to the prior; across the three prior-mean pairs, OOS differences typically do not exceed 2-3 percentage points.

Second, individual stocks are harder than portfolios to deliver stable OOS predictability, and this difficulty manifests as a clear division of labor in the sparsity structure. For large caps (Big500), INS values show higher q_α and lower q_β (under the prior mean (0.5, 0.5): 0.63 and 0.37), indicating that risk exposures require a denser set of characteristics, while the mispricing channel is relatively sparse. For small caps (Small500), the pattern reverses, with (q_α, q_β) = (0.55, 0.51), suggesting that greater cross-sectional heterogeneity is loaded into the alpha channel. Consistent with this, OOS CSR^2 is generally lower and nearly insensitive to priors, indicating that stock-

[11] When computing out-of-sample statistics, we still rely on in-sample information, yet the individual-stock panel is unbalanced. We therefore restrict the sample to assets observed in both the in-sample and out-of-sample periods. As a result, in this exercise, Big500 and Small500 each contain 399 stocks. We retain these labels solely for consistency with the full sample.

level idiosyncratic noise erodes transportable signals; sparsity learning primarily acts as regularization, guarding against overfitting rather than producing a high out-of-sample fit.

Table 8: Out-of-sample Model Performance for Different Test Assets

This table reports in-sample (INS) and out-of-sample (OOS) model results under different prior assumptions for sparsity on different test assets, focusing on models with three latent factors for simplicity. The analysis includes portfolios with varying numbers of P-Tree assets, individual stocks, and three additional portfolio sets: ME/BM 25, Bi360, and Uni610. The metrics include INS and OOS cross-sectional R^2, and the posterior mean of INS sparsity probabilities.

	prior mean: (0.5,0.5)			prior mean: (0.9,0.9)			prior mean: (0.1,0.1)		
	INS		OOS	INS		OOS	INS		OOS
	CSR^2	(q_α, q_β)	CSR^2	CSR^2	(q_α, q_β)	CSR^2	CSR^2	(q_α, q_β)	CSR^2
Panel A: P-Tree									
100	46.2	0.68,0.43	64.1	46.8	0.83,0.57	64.5	43.6	0.53,0.32	61.3
200	46.0	0.59,0.40	72.2	46.8	0.75,0.53	72.8	45.8	0.41,0.23	73.2
400	27.3	0.57,0.31	60.4	26.1	0.72,0.43	60.7	26.8	0.43,0.17	59.6
Panel B: Ind. Stock									
500 big	32.3	0.63,0.37	14.7	33.4	0.77,0.50	14.2	33.5	0.48,0.24	14.4
500 small	6.8	0.55,0.51	13.0	8.4	0.70,0.66	13.5	6.8	0.41,0.37	13.0
Panel C: Others									
ME/BM25	32.8	0.75,0.60	58.0	55.4	0.93,0.73	53.7	14.8	0.57,0.46	59.7
Bi360	-2.5	0.54,0.29	40.7	0.5	0.68,0.48	51.1	28.2	0.31,0.07	54.5
Uni610	36.6	0.56,0.23	67.2	37.2	0.70,0.37	67.5	35.7	0.42,0.10	66.8

Third, classic portfolio sets of different types also exhibit complexity that matches out-of-sample. For information-rich sets such as Bi360 and Uni610, OOS predictability is better unlocked only when the factor-loading channel is allowed to be denser. For Bi360, as the prior means move from (0.9,0.9) down to (0.1,0.1), the posterior qbeta declines from 0.48 to 0.07, while OOS CSR^2 rises from 51.1% to 54.5%. For Uni610, OOS remains near 67% and beta stays persistently dense, with q-beta ranging from 0.23 to 0.10, indicating that the model automatically allocates complexity to risk exposures. By contrast, for the simpler ME/BM25, OOS is slightly higher under dense priors (59.7% versus 58.0% and 53.7%), yet differences across priors are limited, and its INS results are extremely sensitive to priors, with CSR^2 jumping from 32.8% to 55.4% and then falling to 14.8%. This warns that in-sample fit alone can create a performance illusion.

In sum, Table 8 suggests the following: when the cross section is more complex, the best OOS performance arises from denser factor loadings combined with moderately sparse mispricing; when the cross section is simpler, the model is more robust to priors, but in-sample fit should not substitute for an out-of-sample test of complexity matching. Moreover, the simpler the cross section, the more prior-robust the results, with OOS metrics fluctuating only modestly across priors. Nevertheless, superior INS does not substitute for an out-of-sample test of complexity matching. If the factor dimension is underspecified or sparsity constraints are misaligned, OOS predictability deteriorates.

8 Conclusion

This paper establishes a new probabilistic foundation for understanding sparsity in asset pricing models. Departing from the conventional binary paradigm — where models are assumed to be either sparse or dense — we develop a Bayesian framework that models sparsity as a latent feature to be inferred from the data. Using hierarchical spike-and-slab priors within a conditional latent factor model, we learn the probability that each firm characteristic contributes to mispricing or factor exposures.

Our results challenge the notion of sparsity as a fixed property of the return-generating process. We demonstrate that the degree of sparsity varies systematically with the complexity of the test assets, the informativeness of observable factors, and the prevailing macroeconomic regimes. Simpler portfolios, such as ME/BM sorts, require only a few predictive characteristics and display high sparsity. In contrast, richer asset sets, such as Panel-Tree portfolios and small-cap stocks, demand denser representations. Recession periods also induce greater sparsity, reflecting the dominance of a narrower set of priced risks under stress.

Importantly, we find that mispricing terms are generally sparser than factor loadings, and the characteristics driving these two channels are largely distinct. Models that allow the data to determine the relevant level of sparsity outperform those that impose rigid structures, whether through hard constraints or fully dense specifica-

tions. Out-of-sample tests likewise show that treating sparsity as a learnable latent quantity materially improves OOS performance and is robust to priors. As asset and factor dimensionalities increase, information is progressively reallocated from the mispricing to the factor loadings. The better-performing specifications exhibit denser beta together with a moderately sparse alpha. Moreover, substituting latent factors with observable factors maintains the overall random pattern of alpha and beta sparsity. However, the specific values are subject to variation depending on the underlying factor models. Augmenting observable factor models with latent components improves pricing accuracy.

By reframing sparsity as a probabilistic object of inference, our approach provides a more flexible and empirically grounded perspective on model complexity. It bridges the divide between sparse and dense modeling philosophies and provides new insight into how, when, and why firm characteristics matter in the cross section of returns.

References

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*(1), 259–299.

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

Antweiler, W. and M. Z. Frank (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance 59*(3), 1259–1294.

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

Avramov, D., S. Cheng, L. Metzker, and S. Voigt (2023). Integrating factor models. *Journal of Finance 78*(3), 1593–1646.

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 (2007). Momentum and credit rating. *Journal of Finance 62*(5), 2503–2520.

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

Avramov, D., T. Chordia, G. Jostova, and A. Philipov (2022). The distress anomaly is deeper than you think: Evidence from stocks and bonds. *Review of Finance 26*(2), 355–405.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance 61*(4), 1645–1680.

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.

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.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *Journal of Finance 56*(1), 1–43.

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

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

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.

Cui, L., G. Feng, Y. Hong, and J. Yang (2025). Do asset pricing models change over time? Technical report, City University of Hong Kong.

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. B. Ke, B. T. Kelly, and S. Malamud (2023). Complexity in factor pricing 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.

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.

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.

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.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1), 429–450.

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.

Jagannathan, R. and Z. Wang (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance 51*(1), 3–53.

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.

Koijen, R. S. and S. Van Nieuwerburgh (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics 3*(1), 467–491.

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.

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

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

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

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

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review 71*(3), 289–315.

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

Appendix

A.I Single Testing

As mentioned in Section 4.1, we also examine whether a specific characteristic systematically explains the mispricing component. Specifically, we test whether the intercept term α_0 and each characteristic-driven coefficient $\alpha_{1,l}$ in Γ_α are statistically significant. For each model, we conduct $L+1$ separate tests:

$$H_0 : \alpha_0 = 0 \quad \text{and} \quad H_0 : \alpha_{1,l} = 0,\ l = 1, \ldots, L,$$

where rejection of H_0 indicates statistically significant mispricing.

Figure A.1 reports the posterior confidence intervals for the intercept coefficients under different prior specifications in the five-factor setting. Panels (a), (c), and (e) demonstrate that changes in the q_β prior have minimal impact on the confidence intervals of α, while panels (b), (d), and (f) reveal similarly limited sensitivity to variations in the q_α prior. Notably, $\alpha_{1,18}$ remains statistically different from zero across all specifications. This stability is consistent with the $M_\alpha = 1$ result in Table A.2, which confirms that at least one characteristic-driven intercept in $\boldsymbol{\alpha}_1$ is significant in models with five latent factors.

These findings underscore the robustness and economic relevance of mispricing components in explaining cross-sectional returns.

Figure A.1: Posterior Confidence Intervals for Γ_α

This figure presents the posterior confidence intervals of all coefficients in Γ_α under different prior specifications. The x-axis denotes α_0 and $\alpha_{1,l}$ ($l = 1, \ldots, 20$), while the y-axis reports their estimated values. Green bars indicate that the corresponding confidence interval does not include zero, whereas purple bars indicate that zero lies within the interval.

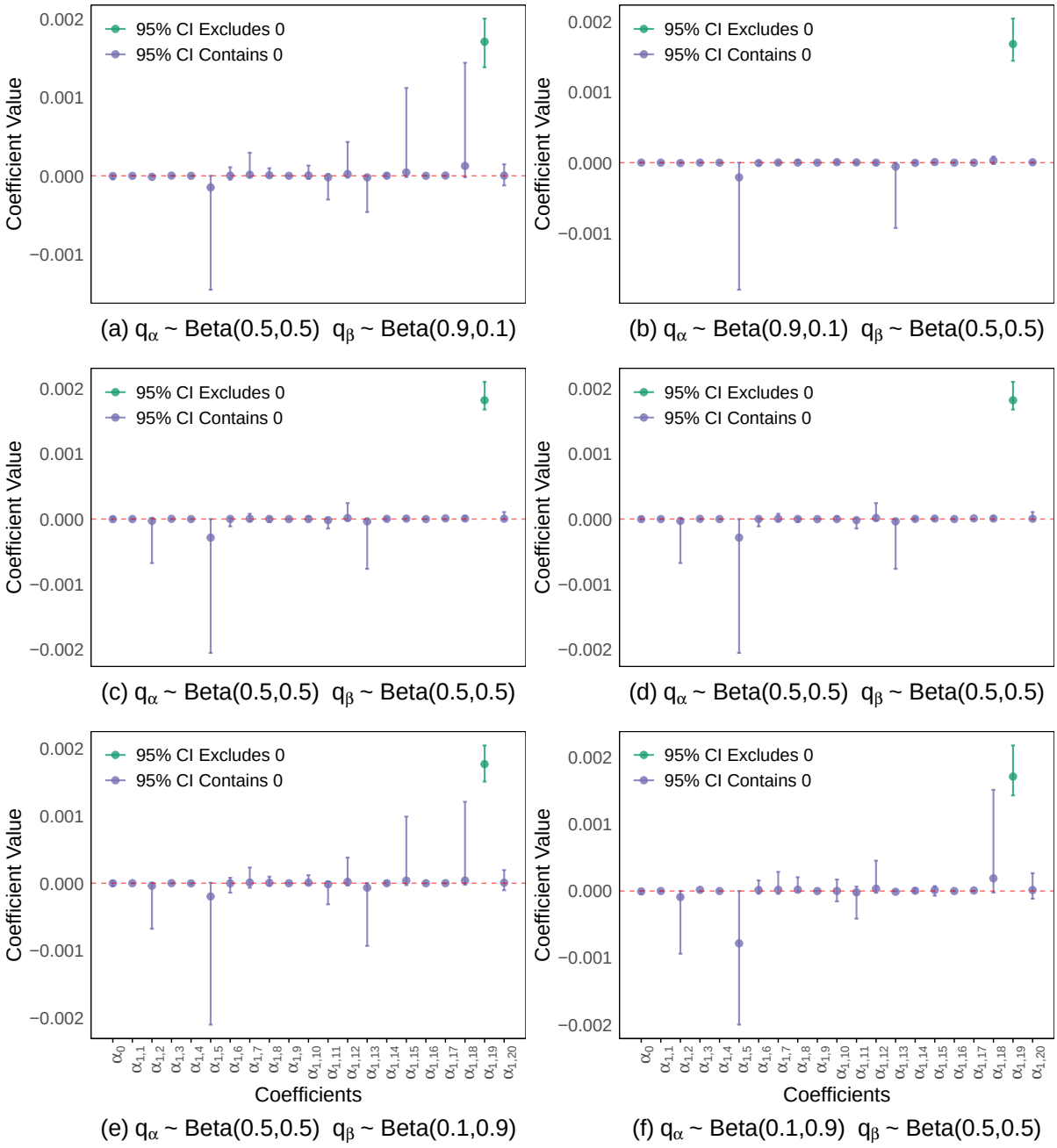
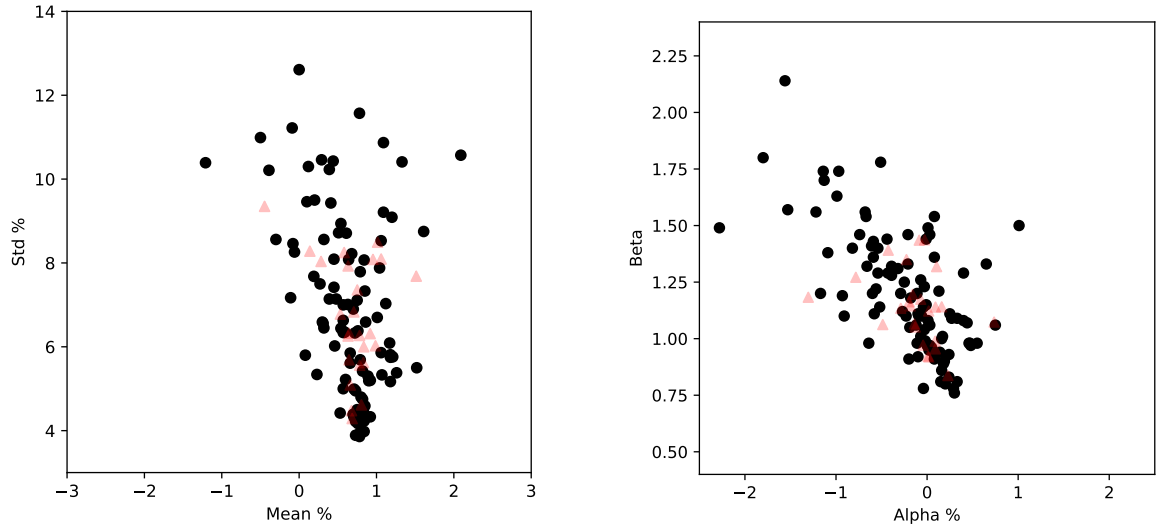


Figure A.2: Diversified P-Tree Test Assets

This figure compares the performance of P-Tree test assets with the ME/BM 25 portfolios. The 100 black circles represent P-Tree test assets, while the 25 light-red triangles denote ME/BM portfolios. Panel (a) plots the mean against the standard deviation of monthly returns (in percentages), and Panel (b) plots CAPM alpha (in percentages) against beta.



(a) Mean v.s. Standard Deviation

(b) CAPM Alpha-Beta

Figure A.3: Histograms of Some Selected Characteristics

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

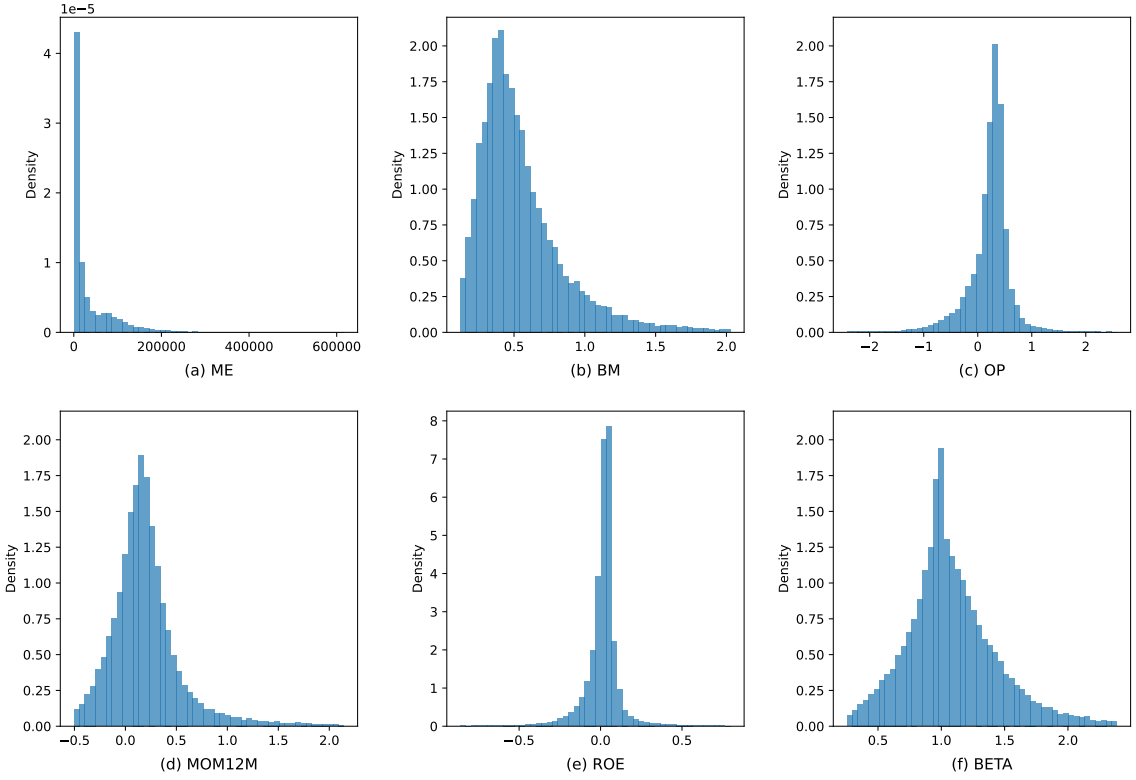


Figure A.4: Sparsity Levels and Pricing Difficulty of Test Assets (without Mispricing)

This figure illustrates the relationship between sparsity in factor loadings and the pricing difficulty of test assets. The plot presents the relationship between q_β and the Sharpe ratio of the test assets. Circles, triangles, and squares denote bi-sorted, uni-sorted, and P-Tree portfolios, respectively.

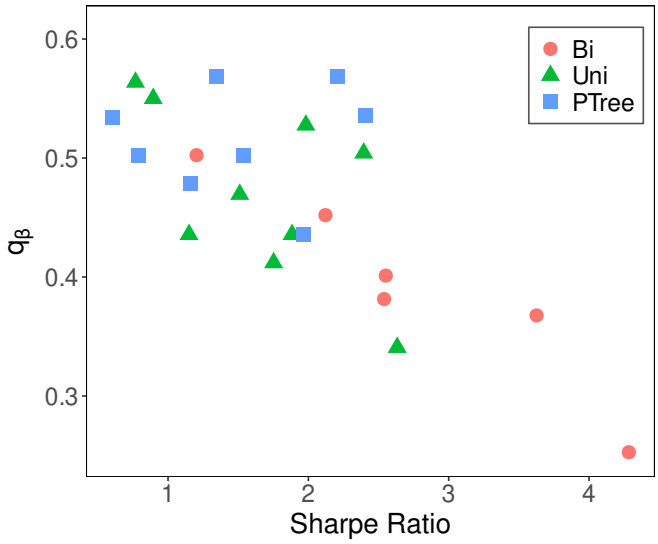
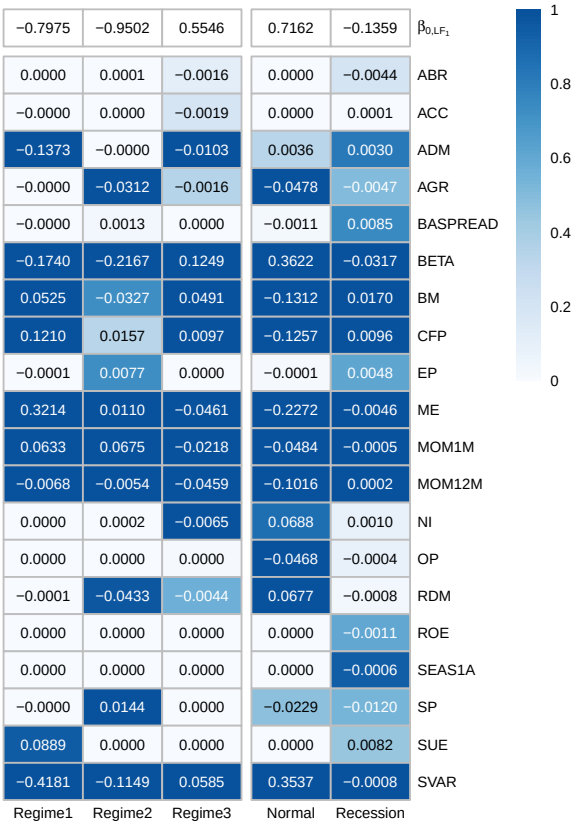


Figure A.5: Regime-based Characteristics Importance (without Mispricing)

This figure presents selection probabilities of characteristics over regimes by structural breaks or business cycles. Panel (a) shows selection probabilities for characteristics explaining factor loadings and the associated $\beta_{1,i}$ estimates. For brevity, results are shown for the first latent factor only. Each cell displays the coefficient for the characteristic, with color intensity reflecting the magnitude of selection probability.



(a) β_{1,LF_1}

Table A.1: Equity Characteristics

This table lists the descriptions of 61 characteristics used in the empirical study.

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

Table A.2: Number of Selected Characteristics in Different Models

This table reports the number of selected characteristics in different models. "M_α" represents the number of characteristics driving α, "M_β" denotes the number of characteristics driving the factor loadings, and K refers to the number of factors. Panel A shows the case where separate spike-and-slab priors are used for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, while Panel B corresponds to the models without mispricing. The number of selected characteristics is calculated by counting the number of characteristics that have a selection probability of 0.5 or greater. The selection probability for each characteristic is the posterior mean of its $p(d \mid \sim)$.

		M_α			M_β		
Panel A: With mispricing models		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
(q_α prior mean, q_β prior mean)	0.9,0.9	10	4	1	10	12	10
	0.5,0.9	10	5	1	10	11	10
	0.1,0.9	10	5	1	10	11	10
	0.9,0.5	10	5	1	10	11	10
	0.5,0.5	10	4	1	10	12	10
	0.1,0.5	10	5	1	10	11	10
	0.9,0.1	10	5	1	11	12	11
	0.5,0.1	10	4	1	11	12	11
	0.1,0.1	10	5	2	11	11	10
Panel B: Without mispricing models							
q prior mean	0.9	/	/	/	10	11	11
	0.5	/	/	/	10	12	10
	0.1	/	/	/	10	13	14

Table A.3: Out-of-sample Model Performance under Different Priors (without Mispricing)

This table reports in-sample (INS) and out-of-sample (OOS) model results under various prior assumptions for sparsity on P-Tree 100 test assets, including cross-sectional R^2, Sharpe ratio from the latent factor tangency portfolio, and the posterior mean of INS sparsity probabilities. K indicates the number of latent factors. Panel A's row labeled "q_β prior mean" corresponds to unconstrained sparsity settings, where priors on q_β indicate three Beta distributions: $\text{Beta}(9,1)$, $\text{Beta}(5,5)$, and $\text{Beta}(1,9)$, with means of 0.9, 0.5, and 0.1, respectively. Panel B restricts the number of characteristics driving $\boldsymbol{\beta}_1$. Specifically, M_β restricts the number of characteristics affecting each factor loading $\boldsymbol{\beta}_{1,k}$.

		INS						OOS			
		CSR^2		TP. SR		q_β		CSR^2		TP. SR	
		$K=1$	$K=5$	$K=1$	$K=5$	$K=1$	$K=5$	$K=1$	$K=5$	$K=1$	$K=5$
Panel A: Unrestricted # selected chars.											
q_β prior mean	0.9	15.0	53.2	0.43	1.44	0.59	0.60	14.7	68.3	0.49	0.92
	0.5	15.0	51.5	0.43	1.49	0.43	0.47	14.6	68.0	0.49	0.90
	0.1	15.1	53.2	0.43	1.60	0.24	0.32	14.5	68.9	0.49	0.98
Panel B: Fixed # selected chars.											
M_β	2	9.7	39.6	0.45	0.89	/	/	13.6	62.7	0.50	0.63
	10	14.9	48.0	0.43	1.38	/	/	13.8	64.9	0.49	0.66
	18	15.6	55.1	0.43	1.83	/	/	14.9	66.2	0.49	0.55

Table A.4: Sparsity for Different Test Assets (without Mispricing)

This table presents performance statistics across different test assets, focusing on models with three latent factors for simplicity. The analysis includes portfolios with varying numbers of P-Tree assets, individual stocks, and three additional portfolio sets: ME/BM 25, Bi360, and Uni610. Reported statistics comprise cross-sectional R^2, tangency portfolio Sharpe ratio, and estimates of q_β.

	CSR^2	TP. SR	q_β
Panel A: P-Tree			
100	53.2	0.73	0.43
200	64.9	0.76	0.40
400	60.1	0.73	0.19
Panel B: Ind. Stock			
500 big	43.1	0.23	0.37
500 small	22.9	3.48	0.28
Panel C: Others			
ME/BM 25	51.1	0.54	0.50
Bi360	61.4	0.46	0.18
Uni610	61.1	0.31	0.20

Table A.5: Time Variation Analysis: Sparsity in Structural Breaks / Business Cycles (without Mispricing)

The table displays performance statistics of P-Tree 100 test assets over different periods, focusing on models with three latent factors for simplicity. Panel A considers sequential segmentation, whose breakpoints follow Smith and Timmermann (2021). Regime 1 covers January 1990 to July 1998 (103 months), Regime 2 spans September 1998 to June 2010 (143 months), and Regime 3 extends from July 2010 to December 2024 (174 months). Panel B consists of macro-driven segmentations, in which the recession periods, totaling 88 months, are defined using Sahm Rule-designated phases: October 1990-November 1992, May 2001-October 2002, March 2008-May 2010, March 2020-March 2021, and June 2024-September 2024. Panel C is the whole period from January 1990 to December 2024. The reported statistics include model performance metrics (cross-sectional R^2 and tangency Sharpe ratio), as well as estimates of q_β.

	CSR^2	TP. SR	q_β
Panel A: Sequential segmentation			
Regime1	54.4	1.28	0.53
Regime2	34.4	0.69	0.51
Regime3	69.2	1.04	0.49
Panel B: Macro-driven segmentation			
Normal	62.4	0.89	0.45
Recession	16.6	0.70	0.42
Panel C: Full period			
Whole	53.2	0.73	0.43