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

Doron Avramov Guanhao Feng Jingyu He Shuhua Xiao

May 18, 2025

Abstract

This paper investigates whether predictive characteristics in asset pricing models are inherently sparse or dense—a dilemma we term Schrödinger's Sparsity. We introduce a Bayesian conditional latent factor model that endogenously identifies sparsity levels in both mispricing (alphas) and factor loadings (betas) from data. Empirical results reveal context-dependent sparsity: it varies across test assets, over time, and between model components. Specifically, alphas exhibit greater sparsity than betas, and sparsity intensifies during recessions. By opening Schrödinger's box empirically, our framework provides a flexible, data-driven resolution to the sparsity dilemma in high-dimensional asset pricing.

Key Words: Conditional CAPM; Factor models; Variable selection; Spike and slab Regression.

JEL Classification: C12, G11, G12

*We thank Semyon Malamud, Gustavo Schwenkler, the seminar and conference participants at CityUHK and the 2024 First Macau International Conference on Business Intelligence and Analytics for their valuable comments. Avramov (E-mail: doron.avramov@runi.ac.il) is at Reichman University; Feng (E-mail: gavin.feng@cityu.edu.hk), He (E-mail: jingyuhe@cityu.edu.hk) and Xiao (E-mail: shxiao3-c@my.cityu.edu.hk) are at the City University of Hong Kong.

1 Introduction

Sparse modeling has become a cornerstone in high-dimensional scientific research, valued for its parsimonious structure, interpretability, and robust out-of-sample performance. In asset pricing, the proliferation of risk factors and anomalies (e.g., Harvey et al., 2016; Green et al., 2017) has motivated parallel advancements in sparse modeling to identify economically meaningful characteristics or factors. For example, Feng et al. (2020) and Bryzgalova et al. (2023) provide evidence of sparse factor risk prices. Studies increasingly favor sparse representations of high-dimensional data to explain average returns or predict risk premia (e.g., Gu et al., 2020; Freyberger et al., 2020; Barillas and Shanken, 2017; Chinco et al., 2019; Evgeniou et al., 2023; Feng et al., 2023). Bayesian frameworks have been pivotal for factor selection (e.g., Bryzgalova et al., 2023), model comparison (e.g., Barillas and Shanken, 2018; Chib et al., 2020, 2024), and model averaging (e.g., Avramov, 2002; Avramov and Chao, 2006; Avramov et al., 2023).

However, the sparsity assumption has faced growing skepticism. Giannone et al. (2021) question its universal validity in economic contexts, demonstrating that Bayesian posteriors rarely concentrate on sparse models and highlighting a pervasive "illusion of sparsity". In asset pricing, Kozak et al. (2020) contend that modeling the cross-section of returns requires the stochastic discount factor (SDF) to incorporate numerous firm characteristics, advocating shrinkage over strict variable selection. He et al. (2024) introduce a formal test and find limited empirical support for sparsity in factor models, while Cong et al. (2023) propose a heterogeneous sparse factor model as an alternative to global sparsity assumptions. Concurrently, research has increasingly explored the merits of complex high-dimensional models: Kelly et al. (2024) and Didisheim et al. (2023) show that models with nonlinear transformations or deep latent structures can outperform traditional sparse approaches. Statistically, Shen and Xiu (2024) demonstrate that Ridge regression (dense) often surpasses Lasso (sparse) in weak-signal environments-common in return prediction settings.

The conflicting conclusions in the literature stem from fundamentally distinct modeling philosophies. Most methodologies require researchers to commit *ex ante* to either a sparse or dense specification, thereby choosing between variable selection (e.g., L_1 penalty) or shrinkage (e.g., L_2 or none) and enforcing that structure globally. For instance, applications of Instrumented Principal Component Analysis (IPCA) explicitly assume either a dense form (Kelly et al., 2019) or a sparse structure (Bybee et al., 2023). Yet such modeling outcomes often reflect prior assumptions rather than data-generating processes, raising a critical question: how can researchers determine the optimal model structure *before* observing the data?

We refer to this dilemma as "Schrödinger's Sparsity." Much like the thought experiment in quantum mechanics, the model's sparsity status remains unknowable until the data are fully examined. This paper addresses the issue by developing a flexible Bayesian framework that avoids imposing sparsity *a priori*. Instead, we allow the posterior distribution to "open the box" and reveal whether the data support a sparse or dense representation, and to what extent.

We investigate whether asset pricing models are sparse within the conditional latent factor structure of IPCA (Kelly et al., 2019), where both mispricing (alphas) and factor loadings (betas) are modeled as functions of lagged firm characteristics. We adopt a Bayesian implementation that extends the framework of Geweke and Zhou (1996), introducing hierarchical spike-and-slab priors on the characteristic coefficients. These priors probabilistically interpolate between variable selection and shrinkage, allowing the degree of sparsity to be estimated from the data. The posterior distribution captures both the magnitude and uncertainty of inclusion, producing soft and interpretable estimates of model complexity.

Our framework includes several innovations. First, we permit sparsity levels to be freely estimated or fixed exogenously, enabling systematic evaluation of model performance under different sparsity constraints. Second, we separate the sparsity of alphas from that of betas, reflecting their distinct economic roles. Third, we incorporate observable traded factors alongside latent ones, allowing for estimation of con-

ditional versions of well-known models such as the CAPM and Fama–French three-factor model, while also recovering unspanned components. This alleviates omitted variable bias and connects our method to the literature on conditional factor models (e.g., Lettau and Ludvigson, 2001; Giglio and Xiu, 2021).

In addition, our framework enables time variation and heterogeneity in sparsity patterns, allowing us to examine how the effective set of characteristics shifts across regimes or economic states. This extends prior work on time-varying parameter models (e.g., Bollerslev et al., 1988; Ferson and Harvey, 1999; Lewellen and Nagel, 2006; Ang and Kristensen, 2012; Connor et al., 2012; Adrian et al., 2015; Bali et al., 2017), as well as Bayesian treatments of time variation and structural breaks (e.g., Kalli and Griffin, 2014; Bitto and Frühwirth-Schnatter, 2019; Huber et al., 2021; Koop and Korobilis, 2023; Cui et al., 2023; Smith and Timmermann, 2021; Chib and Smith, 2024).

We apply our method to a panel of U.S. equities from January 1980 to December 2024. Test assets include Fama–French 25 portfolios, long–short characteristic portfolios, bivariate- and trivariate-sorted portfolios, P-Tree portfolios (Cong et al., 2025), and individual stocks. Our empirical results yield several key findings.

First, we find that the best-performing models are neither extremely sparse nor fully dense. Across asset classes, models using a substantial yet selective set of characteristics achieve superior explanatory power, providing strong evidence for the "illusion of sparsity" in characteristic-based pricing models.

Second, when sparsity is imposed exogenously, model performance is highest when the imposed level aligns with the endogenous level selected by the posterior. Misaligned sparsity levels—either too sparse or too dense—lead to poorer statistical and economic performance. This emphasizes the need to estimate sparsity rather than assume it.

Third, sparsity varies across test asset sets. Sparse models perform well for Fama–French 25 portfolios, which exhibit strong factor structure and limited heterogeneity. In contrast, models for P-Tree portfolios and individual stocks require many more characteristics, reflecting greater idiosyncratic variation and weaker factor dominance.

Fourth, mispricing is typically sparser than factor loadings. The two components exhibit a complementary relationship: when factor loadings are dense, mispricing becomes more concentrated, and vice versa. Moreover, the sets of characteristics that drive each component are generally distinct.

Fifth, sparsity is time-varying. We show that models become more sparse during recessions and following structural breaks (Smith and Timmermann, 2021), consistent with a contraction in the set of useful predictors under macroeconomic stress.

Sixth, models that combine observable and latent factors outperform those that use either component alone. Latent factors capture unspanned variation; observable factors aid interpretation. Their combination yields improved fit and robustness, especially when characteristic sparsity is properly estimated.

Our paper contributes to multiple strands of the literature. First, it extends the class of conditional factor models in which alphas and betas depend on firm characteristics (e.g., Jagannathan and Wang, 1996; Lettau and Ludvigson, 2001; Lettau and Pelger, 2020; Kim et al., 2021; Fan et al., 2024) by providing a unified Bayesian framework that flexibly estimates time-varying mispricing and exposures without prespecifying the sparsity structure of the characteristics. Second, it advances the literature on Bayesian model selection, averaging, and shrinkage in finance (e.g., Avramov, 2002; Avramov and Chao, 2006; Barillas and Shanken, 2018; Chib et al., 2020, 2024; Avramov et al., 2023; Bryzgalova et al., 2023) by developing a conditional spike-and-slab prior structure that allows both inclusion probabilities and sparsity levels to be inferred from the data.

Third, it responds to the ongoing debate over sparsity versus complexity in asset pricing (e.g., Kozak et al., 2020; Giannone et al., 2021; He et al., 2024; Kelly et al., 2024; Didisheim et al., 2023; Shen and Xiu, 2024) by empirically demonstrating that asset pricing models are neither strictly sparse nor fully dense, and by formalizing this ambiguity as a "Schrödinger's Sparsity" dilemma that must be resolved through posterior inference. Fourth, it complements machine learning–based approaches that extract latent factor structures from high-dimensional data (e.g., Gu et al., 2021; Chen

et al., 2024; Feng et al., 2024) by offering a more interpretable and probabilistically grounded alternative that can incorporate domain knowledge through priors and decompose characteristic effects across model components. Finally, it extends the literature on time-varying and regime-dependent models of expected returns and factor loadings (e.g., Bollerslev et al., 1988; Ferson and Harvey, 1999; Lewellen and Nagel, 2006; Ang and Kristensen, 2012; Connor et al., 2012; Adrian et al., 2015; Bali et al., 2017; Kalli and Griffin, 2014; Bitto and Frühwirth-Schnatter, 2019; Huber et al., 2021; Koop and Korobilis, 2023; Cui et al., 2023; Smith and Timmermann, 2021; Chib and Smith, 2024) by documenting substantial variation in estimated sparsity levels across macroeconomic regimes and test asset structures, and by providing a flexible Bayesian tool to model such dynamics.

By framing the sparse-versus-dense dilemma as a case of "Schrödinger's Sparsity", we provide a unified and flexible Bayesian methodology that allows the data—not prior beliefs—to resolve the structure of characteristic-driven pricing models.

The remainder of the paper is structured as follows. Section 2 introduces the Sparse BayesIPCA model. Section 3 details the dataset. Section 4 reports empirical findings, emphasizing sparsity in asset pricing models. Section 5 concludes.

2 Methodology

Let $r_{i,t}$ denote the return of asset i at time t, and let $\mathbf{r}_t = (r_{1,t}, \ldots, r_{N_t,t})^\top$ denote the vector of returns for the N_t assets available for investment at time t. The dataset spans T time periods and may be unbalanced, meaning the number of available assets N_t can vary over time. For each asset i, let $\mathbf{Z}_{i,t-1} = (z_{i,1,t-1}, \ldots, z_{i,L,t-1})^\top$ denote the L-dimensional vector of lagged characteristics observed at time $t-1$, where L is the number of characteristics. Finally, $\mathbf{f}_t$ denotes a K-dimensional vector of latent factors at time t that are common across all assets.

Our conditional factor model allows for time-varying mispricing and factor loadings that depend on firm characteristics, and incorporates varying degrees of prior belief regarding the sparsity of the characteristics that drive variation in both alphas

and betas. Specifically, the conditional model is defined as:[1]

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

In this specification, $\boldsymbol{\alpha}_1$ is an $L \times 1$ vector of characteristic loadings for the intercept (mispricing) component, $\boldsymbol{\beta}_0$ is a $K \times 1$ vector of baseline factor loadings, and $\boldsymbol{\beta}_1$ is a $K \times KL$ matrix capturing the interaction between K factors and L lagged firm characteristics. The Kronecker product $(\mathbb{I}_K \otimes \mathbf{Z}_{i,t-1})$ ensures that each factor loading varies flexibly with the characteristics.

For each period t, we assume that the idiosyncratic return vector satisfies $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_{N_t}^2)$. The latent factors follow standard assumptions: $\mathbb{E}[\mathbf{f}_t] = 0$, $\mathbb{E}[\mathbf{f}_t \mathbf{f}_t^\top] = \mathbb{I}$, $\mathbb{E}[\boldsymbol{\epsilon}_t \mid \mathbf{f}_t] = 0$, and $\mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top \mid \mathbf{f}_t] = \boldsymbol{\Sigma}$. We place heterogeneous inverse-gamma priors on the residual variances of each asset: $\sigma_i^2 \sim \text{IG}(v_{0,i}/2, S_{0,i}/2)$.

Substituting the characteristic-based specifications of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ into the return equation yields

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

We are now ready to address the role of sparsity in asset pricing. When sparsity is not imposed, we assign a conjugate normal prior, $\mathcal{N}(0, \xi^2)$, to all regression coefficients—this specification defines the baseline BayesIPCA model. In contrast, the Sparse BayesIPCA formulation explicitly incorporates sparsity by imposing priors that allow for variable selection, and is developed as follows.

[1] Our approach builds on the Bayesian literature for drawing latent factors, such as Geweke and Zhou (1996), and takes inspiration from the conditional structure in Instrumented Principal Component Analysis (IPCA) proposed by Kelly et al. (2019). However, it differs substantially by jointly modeling mispricing and factor exposures using a flexible sparsity-based prior.

2.1 Full Factor Model with Sparsity

We focus on the sparsity of characteristics, particularly in the components associated with $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$. To capture global sparsity, we adopt the hierarchical prior structure of Giannone et al. (2021), which builds on the standard spike-and-slab prior of Mitchell and Beauchamp (1988) and George and McCulloch (1993). Specifically, we impose the following prior on the elements of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$, governed by a *global* probability of inclusion q:

$$[\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1] \overset{iid}{\sim} \begin{cases} \mathcal{N}(0, \gamma^2) & \text{with probability } q, \\ 0 & \text{with probability } 1-q, \end{cases} \tag{3}$$

where γ is a hyperparameter controlling the degree of shrinkage in the slab component, with a prior distribution $\gamma^2 \sim \text{IG}(A_\gamma/2, B_\gamma/2)$.

The spike-and-slab prior includes two components: the slab component assigns a normal prior, allowing for nonzero coefficients with mild shrinkage, while the spike component fixes the coefficient at zero, effectively excluding the corresponding variable from the model. Although one could assess sparsity ex-post by counting the number of nonzero coefficients—for example, He et al. (2024) define a sparse model as one with fewer than ten active factors—we instead adopt a fully Bayesian approach that estimates the inclusion probability q directly.

In this framework, q serves as a continuous summary of global sparsity. When $q < 0.5$, the model is classified as sparse, as a majority of the characteristic-based components in $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1})$ and $\boldsymbol{\beta}(\mathbf{Z}_{i,t-1})$ are inactive. Conversely, when $q > 0.5$, the model is considered dense. Thus, the closer q is to zero, the greater the sparsity; the closer it is to one, the richer the characteristic-based structure.

We consider three specifications for the key parameter q.

In the *first setting*, we fix q to a predetermined value that reflects the investor's prior belief about model sparsity. A small q implies a belief in a sparse structure, while a large q reflects a preference for a denser model. Importantly, fixing q does not

constrain the fraction of variables that can be selected—it merely encodes prior beliefs and can be overturned by strong data-driven likelihood.

In the *second setting*, following Giannone et al. (2021), we treat q as an unknown parameter with a Beta prior, $q \sim \text{Beta}(a_q, b_q)$, and estimate its posterior distribution. This specification captures a *global* inclusion level, allowing data to inform the degree of sparsity in both the mispricing and factor loading components.

The *third setting* allows for differentiated sparsity in the characteristics that drive mispricing versus those that drive factor loadings. Specifically, we impose *local* sparsity levels by assigning separate spike-and-slab priors:

$$\boldsymbol{\alpha}_1 \overset{iid}{\sim} \begin{cases} \mathcal{N}(0, \gamma_\alpha^2) & \text{with probability } q_\alpha \\ 0 & \text{with probability } 1 - q_\alpha \end{cases}, \tag{4}$$

$$\boldsymbol{\beta}_1 \overset{iid}{\sim} \begin{cases} \mathcal{N}(0, \gamma_\beta^2) & \text{with probability } q_\beta \\ 0 & \text{with probability } 1 - q_\beta \end{cases}. \tag{5}$$

We assign Beta priors $q_\alpha \sim \text{Beta}(a_{q_\alpha}, b_{q_\alpha})$ and $q_\beta \sim \text{Beta}(a_{q_\beta}, b_{q_\beta})$, and distinguish the shrinkage hyperparameters as γ_α and γ_β, respectively. The intercepts α_0 and $\boldsymbol{\beta}_0$ are assigned standard conjugate Gaussian priors, $\mathcal{N}(0, \xi^2)$.

Next, define $\Gamma_\alpha = [\alpha_0, \boldsymbol{\alpha}_1]$ and $\Gamma_\beta = [\boldsymbol{\beta}_0, \boldsymbol{\beta}_1]$. We impose the normalization $\Gamma_\beta^\top \Gamma_\beta = \mathbb{I}_K$, 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. For identification, we also impose $\Gamma_\alpha^\top \Gamma_\beta = \mathbf{0}_{1\times K}$, which we implement by regressing Γ_α on Γ_β and defining the orthogonal residual as $\widehat{\Gamma}_\alpha$.

Thus far, we have formulated a conditional latent factor model with prior beliefs about the degree of sparsity in the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ coefficients. We now extend the model to allow for the coexistence of latent and observable factors.

The augmented factor model is given by:

$$r_{i,t} = \underbrace{\alpha_0 + \boldsymbol{\alpha}_1^\top \mathbf{Z}_{i,t-1}}_{\text{mispricing}} + \underbrace{\boldsymbol{\beta}_0^\top \tilde{\mathbf{f}}_t + \boldsymbol{\beta}_1^\top \left[\tilde{\mathbf{f}}_t \otimes \mathbf{Z}_{i,t-1}\right]}_{\text{observable factors, conditional loadings}} + \underbrace{\boldsymbol{\beta}_0^\top \mathbf{f}_t + \boldsymbol{\beta}_1^\top \left[\mathbf{f}_t \otimes \mathbf{Z}_{i,t-1}\right]}_{\text{latent factors, dynamic loadings}} + \epsilon_{i,t}, \tag{6}$$

where $\tilde{\mathbf{f}}_t$ denotes observable factors—such as the market factor (MKT), Fama-French three-factor (FF3), or five-factor (FF5) models—and $\mathbf{F}_t$ represents the full set of factors, including both observable and latent components.

Introducing latent factors into the conditional observable factor model helps mitigate misspecification arising from omitted or unobserved risk sources, thereby enhancing the model's ability to explain the cross-section of returns. Conversely, incorporating observable factors improves the interpretability of the latent structure. This extension also introduces a variable selection mechanism within the conditional factor framework, alleviating the challenges of choosing conditioning variables in traditional time-varying approaches and reducing the risk of overfitting. As a result, it improves the estimation accuracy of conditional betas.

We validate our Sparse BayesIPCA methodology through a simulation study. Specifically, we generate coefficients, characteristics, and factors, which are then combined to compute the conditional expected return $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1}) + \boldsymbol{\beta}(\mathbf{Z}_{i,t-1})\mathbf{f}_t$. Given a predefined signal-to-noise ratio (SNR), we add normally distributed noise to obtain the final simulated return, as specified in Equation 2.

Appendix A.I details the simulation design and shows that our method successfully identifies the "useful" characteristics and accurately recovers the coefficients specified by the data-generating process (DGP).

2.2 Precise Control of Model Sparsity

In our model, the selection of characteristics that drive mispricing and factor loadings is governed jointly by the prior and the likelihood, with the latter often exerting a dominant influence. Characteristic inclusion is determined probabilistically via spike-and-slab priors, enabling flexible, data-driven selection. While the prior on the inclu-

sion probability q shapes the overall sparsity tendency, it does not provide precise control over the number of selected characteristics. As a result, even strong prior beliefs may be overridden by the data, making it difficult to enforce exact sparsity through prior specification or hyperparameter tuning alone. To address this limitation and facilitate a transparent comparison across sparsity regimes, we introduce an alternative formulation that explicitly constrains the number of selected characteristics for both mispricing and factor loadings.

In this framework, controlling sparsity effectively amounts to controlling the number of selected characteristics. To enable precise comparisons between sparse and dense models, we modify the prior on the inclusion indicators z_i by imposing a constraint on their sum. Specifically, we adopt the following joint prior:

$$(z_1, z_2, \cdots, z_L) \sim \prod_{i=1}^{L} \text{Bernoulli}(L) \times \mathbb{I}\left(\sum_{i=1}^{L} z_i = M\right), \tag{7}$$

where $M \leq L$ denotes the desired number of selected characteristics.

Equation (7) resembles a product of independent Bernoulli distributions, but introduces a hard constraint: the total number of selected variables must equal M, a fixed threshold specified ex ante. When the model includes mispricing, there are $(1 + K)$ distinct sets of characteristics, each of length L. The prior in Equation (7) is applied separately to each set, ensuring that M characteristics are selected independently within each block.

This formulation provides a direct and transparent mechanism to control sparsity through the prior. The corresponding inclusion level is given by M/L, where a larger M corresponds to a denser model specification. To allow for independent control over the sparsity of characteristics driving mispricing and factor loadings, we further ex-

tend the prior as follows:

$$(z_1^{\alpha}, z_2^{\alpha}, \cdots, z_L^{\alpha}) \sim \prod_{i=1}^{L} \text{Bernoulli}(L) \times \mathbb{I}\left(\sum_{i=1}^{L} z_i^{\alpha} = M_{\alpha}\right), \tag{8}$$

$$(z_1^{\beta_k}, z_2^{\beta_k}, \cdots, z_L^{\beta_k}) \sim \prod_{i=1}^{L} \text{Bernoulli}(L) \times \mathbb{I}\left(\sum_{i=1}^{L} z_i^{\beta_k} = M_{\beta}\right), \quad \text{for } k = 1, \ldots, K. \tag{9}$$

Here, M_{α} constrains the number of characteristics selected to drive mispricing, while M_{β} controls the number of characteristics driving each of the K latent factor loadings.

We implement the constrained joint priors in Equations (8) and (9) using a Metropolis–Hastings algorithm.

The full Gibbs sampling algorithm used for the specifications with unconstrained sparsity, along with the Metropolis–Hastings procedure employed to implement the constrained prior formulation, is provided in Appendix A.II.

3 Data

Test assets. We apply our approach to both U.S. individual stocks and portfolios.

For individual stock returns, we use monthly data from January 1980 through December 2024, spanning 540 months and covering 21,165 stocks.[2]

All portfolio-level test assets are derived from the above individual stock data. We construct three sets of portfolios: (i) the standard 25 ME/BM portfolios, (ii) 61 long-short portfolios based on individual characteristics, and (iii) 357 bivariate-sorted portfolios.[3]

In addition, we employ test assets generated using the Panel-Tree (P-Tree) method proposed by Cong et al. (2025), which produces 400 test portfolios by applying a customized machine learning tree algorithm with 10 "leaves" across 40 "trees." The P-Tree framework extends traditional multi-way sorting by enabling non-linear interactions among high-dimensional characteristics, aiming to maximize the Sharpe ratio of

[2]Standard filters (see, e.g., Fama and French, 1992) are applied, 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 and 11; and (3) excluding stocks with negative book equity or lagged market equity. The sample begins in 1980 due to fewer missing observations in the IBES dataset after that year.

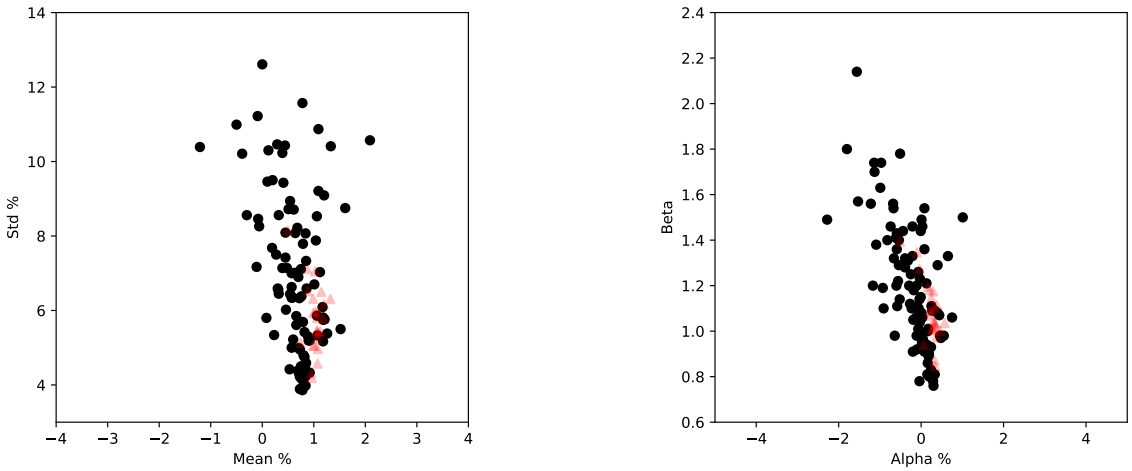
[3]A $2 \times 3 \times 60$ sorting scheme should yield 360 portfolios, but three groups are empty due to lack of stock assignments.

the efficient portfolio spanned by the resulting test assets. The sequential nature of the algorithm implies that each additional test asset contributes decreasing marginal explanatory power.

To mitigate look-ahead bias, we follow Cong et al. (2025) in constructing the P-Tree using data from January 1980 to December 1989. The resulting tree structure is then applied to generate test assets for the period January 1990 through December 2024, forming the basis of our main empirical analysis. Figure 2 compares the cross-sectional dispersion of the P-Tree 100 test assets with that of the 25 ME/BM portfolios, showing greater dispersion among P-Tree portfolios.

Figure 1: Diversified P-Tree Test Assets

This figure compares the performance of P-Tree test assets with the 5×5 ME/BM portfolios. The 100 black circles represent P-Tree test assets, and the 25 light-red triangles represent ME/BM portfolios. Subfigure (a) plots mean versus standard deviation of monthly returns (in percent), and subfigure (b) plots CAPM alpha (in percent) versus beta.



(a) Mean v.s. Standard Deviation

(b) CAPM Alpha-Beta

Characteristics. Our dataset includes 61 firm-level characteristics, categorized into eight thematic groups: size, value, investment, momentum, profitability, liquidity, volatility, and intangibles. The characteristics are detailed in Table A.4. Each variable is standardized cross-sectionally and scaled monthly to the range $[-1, 1]$.

For portfolio-level characteristics, we apply a return-based weighting scheme to aggregate raw firm-level characteristics, followed by cross-sectional standardization. This procedure is also used to compute characteristics associated with the P-Tree test assets. Figure A.3 displays raw values for selected characteristics of the P-Tree portfo-

lios.

Regime data. To analyze time-varying sparsity, we rely on macroeconomic regimes identified by Smith and Timmermann (2021), who detect structural breaks in the U.S. stock market. In addition, we incorporate the real-time Sahm Rule Recession Indicator from the Federal Reserve Economic Data (FRED). Recession periods are grouped into a single regime, enabling us to compare model behavior across recessionary and normal economic conditions.

4 Empirical Results

This section presents our key empirical findings.[4] Section 4.1 presents asset pricing and investment results under alternative sparse prior specifications, holding the test assets fixed. Section 4.2 explores time variation in sparsity across economic regimes. Section 4.3 examines how sparsity responds to changes in test asset composition and dimensionality. Both sections also analyze how alpha and beta dynamics vary across assets and periods. Finally, Section 4.5 assesses the performance of the conditional model that combines observable and latent factors.

4.1 Sparsity for P-Tree 100 Test Assets

We begin by evaluating model performance using the full sample of P-Tree100 test assets (Cong et al., 2025), which are highly diversified and span efficient portfolios with high Sharpe ratios. An important advantage of the P-Tree framework is its full transparency: we can explicitly trace which characteristics were used in constructing each test asset. This level of interpretability is rarely available in traditional sorted portfolios or latent factor constructions, and it serves as a key motivation for adopting the P-Tree test assets in our analysis.

Table 1 reports results across different sparsity specifications for the global inclusion probability q. Lower posterior means of q imply greater sparsity, as more coeffi-

[4] Incorporating sparsity into the Gibbs sampler increases computational complexity, as it requires iteratively deciding whether to include each characteristic. To mitigate this, we restrict attention to the 20 bolded characteristics listed in Table A.4. For similar reasons, we use P-Tree100 as our baseline test asset set—constructed via the P-Tree method with 10 trees and 10 leaves—to represent a parsimonious yet representative asset universe.

cients in $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are likely to be zero.

We consider three modeling approaches:

(i) A *data-driven model* with a Beta prior on q, allowing the sparsity level to be inferred from the data (Panel A);

(ii) A model with a *fixed number of selected characteristics* (Panel B);

(iii) A *fully dense model* using all 20 characteristics (Panel C).

In Panel A, we vary the prior mean of q across 0.1, 0.5, and 0.9. Despite the differing priors, the model's performance is remarkably stable. For example, with $K = 3$, the cross-sectional R^2 (CSR^2) ranges narrowly between 43.62 and 43.66 across the different priors, and the tangency portfolio Sharpe ratio (TP.Sp) improves as the prior becomes more diffuse. This stability suggests that the posterior is driven more by the data than by the prior. Moreover, increasing the number of latent factors enhances CSR^2, rising from about 29% with $K = 1$ to over 55% with $K = 5$. The number of selected characteristics for $\boldsymbol{\beta}_1$ stabilizes around ten, indicating moderate sparsity rather than extreme parsimony.

Panel B examines fixed-sparsity models where the same number M of characteristics drives both mispricing and factor loadings. When $M = 2$, the model achieves the highest CSR^2 at $K = 3$ (52.49%), but performance declines when $K = 5$, as two characteristics are insufficient to span a more complex factor structure. Conversely, larger values of M (10 or 18) do not consistently improve model fit, likely due to overfitting. This illustrates the importance of matching model complexity (i.e., the number of factors) with an appropriate degree of sparsity.

Comparing Panels A and B, we find that *flexible sparsity* generally leads to better explanatory power when the number of factors increases. For instance, when $K = 5$ and the prior mean of q is 0.1, the effective number of selected characteristics is approximately $M_\alpha = 1$ and $M_\beta = 9$, highlighting that mispricing and factor loadings benefit from *differential sparsity*, a point we further investigate in Table 2.

Finally, Panel C reports the performance of the fully dense model using all 20 characteristics. While it performs well with $K = 1$, its CSR^2 and TP.Sp deteriorate as K increases, suggesting overfitting in high-dimensional settings.

Overall, Table 1 supports the view that *moderate and data-driven sparsity* enhances model performance. Rigid assumptions of either sparsity or density may hinder generalization. Motivated by this, we adopt a Beta$(5, 5)$ prior on the global inclusion probability q—or on q_α and q_β separately—throughout the subsequent analysis, allowing the data to flexibly guide the selection of characteristics.

Table 1: Model Performance under Global Sparse Priors

This table reports model results under varying priors for global sparsity, including cross-sectional R^2 and investment performance metrics. "TP.Sp" denotes the annualized Sharpe ratio of the tangency portfolio constructed from latent factors. All models are estimated on P-Tree100 test assets, with K representing the number of latent factors. Panel A presents results under unconstrained sparsity with priors on q specified as Beta$(1, 9)$, Beta$(5, 5)$, and Beta$(9, 1)$, corresponding to prior means of 0.1, 0.5, and 0.9, respectively. These specifications do not restrict the number of selected characteristics. Panel B imposes a fixed number of characteristics driving both alpha and beta, where M explicitly limits the number of characteristics influencing $\boldsymbol{\alpha}_1$ and each factor loading $\boldsymbol{\beta}_{1,k}$.

		CSR^2			TP.Sp		
		$K = 1$	$K = 3$	$K = 5$	$K = 1$	$K = 3$	$K = 5$
Panel A: Unrestricted # selected chars.							
q prior mean	0.1	29.37	**43.66**	**55.57**	0.35	1.36	0.92
	0.5	29.54	43.63	54.79	0.35	1.44	0.92
	0.9	**29.71**	43.62	53.89	**0.35**	**1.50**	**0.95**
Panel B: Fixed # selected chars.							
M	2	25.44	**52.49**	**51.02**	**0.44**	**1.11**	0.48
	10	29.53	38.32	41.51	0.35	0.87	**1.12**
	18	27.48	39.31	42.02	0.33	0.55	0.95
Panel C: No sparsity							
M	20	**29.92**	36.88	45.23	0.35	0.57	0.95

Table 2 evaluates model performance under separate sparsity specifications for the characteristics driving mispricing ($\boldsymbol{\alpha}_1$) and factor loadings ($\boldsymbol{\beta}_1$). The results show that optimal sparsity allocations depend on the number of latent factors. When the factor structure is limited (e.g., small K), better performance arises from assigning more characteristics to mispricing and fewer to factor loadings. Conversely, as the number of latent factors increases, the model benefits from shifting characteristic relevance toward factor loadings and away from mispricing.

This shift reflects a change in how characteristics contribute to pricing information. With only a few factors, the model's ability to capture systematic variation is limited, so mispricing components (via $\widehat{\boldsymbol{\alpha}}(\mathbf{Z}_{i,t-1})$) absorb some of that variation. A small number of characteristics is often sufficient in this case. As K increases, however, a richer factor structure allows more systematic return variation to be captured directly through $\widehat{\boldsymbol{\beta}}(\mathbf{Z}_{i,t-1})\mathbf{f}_t$. Allocating more characteristics to factor loadings thus improves overall model fit by reducing the need for mispricing terms to absorb latent variation.

Comparing Table 2 with Table 1 further highlights the advantage of allowing independent sparsity levels for mispricing and loadings. This flexibility leads to improvements in both cross-sectional explanatory power and Sharpe ratios of the constructed portfolios, suggesting that rigidly imposing uniform sparsity across components may be suboptimal.

Notably, the highest cross-sectional R^2 is achieved when approximately half the characteristics (10 out of 20) are active, consistent with the information used in constructing the P-Tree test assets. The P-Tree method builds portfolios by sequentially partitioning the data using a subset of characteristics, typically relying on around 10 informative variables. Figure 2 illustrates this structure, showing that the first tree—which plays a central role in the partitioning process—draws on nine characteristics. This aligns closely with the optimal sparsity levels inferred in Tables 1 and 2.

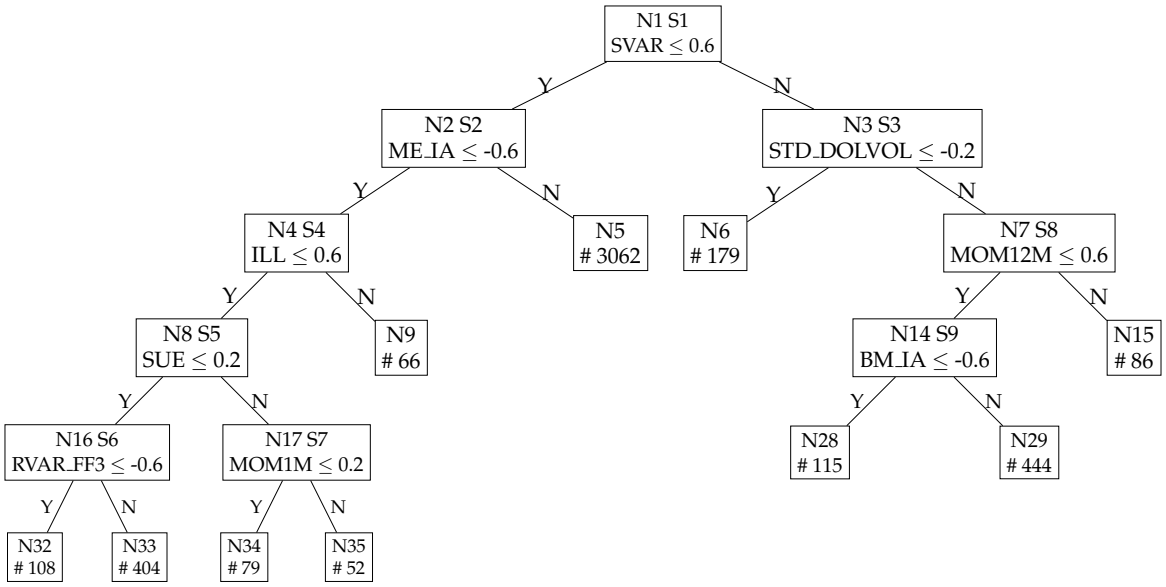
Table 2: Model Performance under Separate Sparse Priors on Alphas and Betas

This table reports model results under various priors for sparsity, including cross-sectional R^2 and investment performance metrics. "TP.Sp" denotes the annualized Sharpe ratio of the tangency portfolio constructed with latent factors. All models use P-Tree100 test assets, and K indicates the number of latent factors. Panel A's row labeled "(q_α prior mean, q_β prior mean)" represents unconstrained sparsity settings, where priors on q_α and q_β combine three distributions: Beta$(1, 9)$, Beta$(5, 5)$, and Beta$(9, 1)$, with means of 0.1, 0.5, and 0.9, respectively. These specifications impose no constraints on the number of selected characteristics. In contrast, Panel B restricts the number of characteristics driving alpha and beta. M_α explicitly limits characteristics influencing $\boldsymbol{\alpha}_1$, while M_β restricts characteristics affecting each factor loading $\boldsymbol{\beta}_{1,k}$.

		CSR^2			TP. Sp		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Unrestricted # selected chars.							
(q_α prior mean, q_β prior mean)	0.1,0.1	29.17	44.09	**59.20**	0.34	0.75	0.71
	0.5,0.1	29.37	43.27	58.47	0.35	0.77	0.79
	0.9,0.1	29.41	43.54	58.00	0.35	1.14	0.68
	0.1,0.5	29.29	43.53	57.82	0.34	0.75	1.00
	0.5,0.5	29.48	42.49	56.84	0.35	1.01	1.14
	0.9,0.5	29.53	43.65	54.94	0.35	1.17	0.92
	0.1,0.9	29.48	**45.11**	58.72	0.34	0.99	0.77
	0.5,0.9	29.64	42.48	56.84	0.35	1.00	**1.14**
	0.9,0.9	**29.73**	44.13	56.69	**0.35**	**1.27**	0.90
Panel B: Fixed # selected chars.							
(M_α, M_β)	2,2	25.44	49.34	48.39	**0.44**	**1.10**	0.95
	10,2	27.98	51.07	50.10	0.37	0.57	0.87
	18,2	25.17	47.01	38.00	0.32	0.79	0.68
	2,10	28.85	51.17	56.83	0.42	0.60	0.87
	10,10	29.59	37.87	41.20	0.35	0.89	0.97
	18,10	27.19	40.97	39.03	0.32	0.47	0.88
	2,18	29.81	**54.91**	**56.99**	0.43	0.65	1.13
	10,18	**29.88**	34.24	51.26	0.36	1.01	**1.22**
	18,18	27.46	39.30	42.11	0.33	0.53	0.94

Figure 2: Panel Tree from 1980 to 1989

We provide splitting characteristics, cutpoint values for each parent node, and their respective node and splitting indexes. For example, node N1 is split by SVAR at 0.4 as the first split S1, and the second split S2 is on node N2 by ME_IA at 0.2. The median monthly number of assets in the terminal leaf basis portfolios is also included. For example, node N6 has 179 stocks by monthly median.



4.2 Time-varying Sparsity

While many studies in empirical asset pricing impose sparsity—particularly through explicit variable selection—this assumption may not hold uniformly over time. Economic environments evolve, and the relevance of firm characteristics may shift with macroeconomic conditions, investor preferences, and financial innovation. For example, characteristics related to distress, profitability, or momentum may play different roles during expansions versus recessions. These considerations suggest that the degree of sparsity itself may vary across time. In this section, we investigate whether sparsity in pricing models changes across structural regimes and business cycles, and whether different sets of characteristics drive returns in different periods.

Time-varying sparsity levels. To address this question, we apply our method to the three structural regimes identified by Smith and Timmermann (2021). Table 3 shows that the explanatory power of characteristics—both for mispricing and factor loadings—varies meaningfully across regimes. In particular, the number of characteristics

driving mispricing declines steadily over time. For example, Panel B reports a drop in q_α from 0.30 in the first regime to 0.29 in the second, and further down to 0.23 in the third. This trend may reflect greater market efficiency in recent decades, reducing the role of persistent mispricing. In contrast, q_β increases from 0.42 to 0.56 over the same periods, suggesting that a growing number of characteristics are needed to capture time-varying factor exposures. The global sparsity measure q follows a similar upward trend, driven largely by the increase in q_β.

Taken together, these patterns indicate that while mispricing components have become more concentrated over time, explaining factor loadings increasingly requires a broader set of characteristics. These dynamics emphasize the importance of identifying which characteristics are relevant in each subperiod—a topic we revisit later in this subsection.

Next, we examine whether sparsity differs between recessionary and non-recessionary periods. We use the real-time Sahm Rule, which signals the onset of a recession when the three-month average unemployment rate rises by at least 0.5 percentage points above its 12-month minimum. Based on this criterion, we partition the sample into recession and normal periods and estimate the model separately for each. Table 3 shows that during recessions, the explanatory contribution of characteristics declines for both mispricing and factor loadings. Specifically, q, q_α, and q_β each fall by approximately 10% relative to normal periods, indicating that the asset pricing model becomes more sparse during economic downturns.

Several mechanisms may explain this recession-induced sparsity. First, in times of heightened market uncertainty, investors tend to focus on macroeconomic and systematic risks, reducing the relevance of firm-specific information (Fama and French, 1993). Second, elevated risk aversion during recessions may shift investor preferences toward safer stocks (Baker and Wurgler, 2007) or increase compensation for downside risks (Farago and Tédongap, 2018), diminishing the effectiveness of some characteristics. Third, large macroeconomic shocks during recessions may weaken the ability of firm-level traits to capture variation in returns (Campbell and Cochrane, 1999).

Overall, our findings highlight the presence of time-varying sparsity and underscore the importance of accounting for regime dynamics in the estimation of asset pricing models. In particular, the model tends to be denser during normal periods and sparser during recessions, suggesting that characteristic relevance is conditional on broader economic conditions.

Table 3: Time Variation Analysis: Sparsity in Regimes

This table reports estimates of q_α, q_β, and q over different periods based on P-Tree100 data, focusing on a model with three latent factors for simplicity. As q_α, q_β, and q decline, the model becomes sparser. 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). These breakpoints follow Smith and Timmermann (2021). 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.

	Different periods					
	Regime1	Regime2	Regime3	Normal	Recession	Full
Panel A: Global prior						
q	0.37	0.41	0.42	0.47	0.42	0.48
Panel B: Separate priors						
q_α	0.30	0.29	0.23	0.27	0.24	0.31
q_β	0.42	0.46	0.56	0.54	0.53	0.59

Dynamic Roles of Characteristics Thus far, we have observed that the sparsity of the asset pricing model varies across time series. Next, we examine which specific characteristics have greater importance for the time-series dimension.

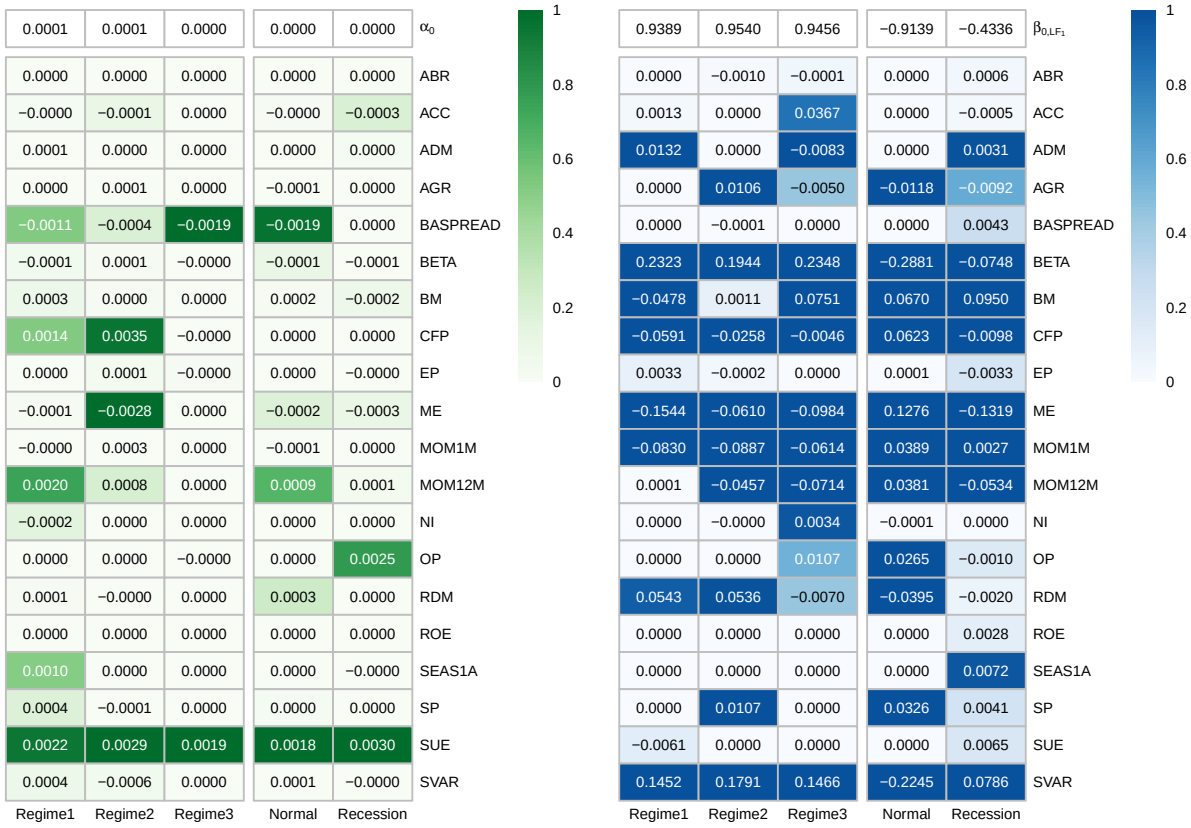
Figure 3 illustrates that from Regime 1 to Regime 3, certain characteristics—such as 12-month momentum (past return over the prior year) and seasonality in January returns—progressively lose their explanatory power for mispricing. In Regime 2, additional drivers of alpha emerge, including cash flow-to-price ratio and market equity (firm size), both exhibiting relatively large coefficients. This regime, which spans the 1997 Asian Financial Crisis and the 2008 Global Financial Crisis, was marked by heightened uncertainty and volatility. During this period, investors appeared to prioritize firms' cash flow performance. Firms with high cash flow-to-price ratios demonstrated more stable fundamentals, while large-cap firms generally delivered higher excess returns.

During economic recessions, both the bid-ask spread (a measure of liquidity costs) and 12-month momentum (past return over the prior year) become less relevant for explaining mispricing. This shift reflects investors' increased focus on macroeconomic risks, leading them to rely more on macro indicators and fundamental data rather than liquidity-related signals—thereby diminishing the explanatory power of the bid-ask spread. Similarly, momentum-based strategies lose significance as investors shift toward safer, more stable assets rather than pursuing riskier, trend-driven positions in times of heightened uncertainty.

In addition, characteristics such as operating profitability, research and development-to-market equity, and sales-to-price ratio lose their ability to explain systematic risk. This likely reflects that, during recessions, returns become increasingly driven by latent macroeconomic factors, weakening the role of firm-specific characteristics in capturing variation in dynamic factor loadings.

Figure 3: Changing Roles of Characteristics in Recessions

This figure presents selection probabilities of characteristics over time. The x-axis denotes time periods, and the y-axis lists characteristics. Panel (a) reports the selection probabilities for each of the 20 characteristics explaining mispricing, alongside the corresponding $\mu_{1,i}$ coefficient estimates. Panel (b) 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) α_1 (b) β_{1,LF_1}

4.3 Large Sets of Test Assets

The previous section examined the sparsity of characteristics in explaining the average returns of the P-Tree100 test assets. It is well known that the choice of test assets is critical and can materially influence the conclusions of asset pricing tests—a point that likely extends to models with latent factors. For example, Kozak et al. (2020) show that estimating the stochastic discount factor (SDF) using Fama-French 25 (FF25) portfolios versus 50 anomaly portfolios yields markedly different results. While the anomaly portfolios exhibit almost no sparsity—due to limited redundancy among the underlying strategies—the FF25 portfolios display substantial redundancy, making it possible to span the SDF with only a small subset of test assets.

Intuitively, fewer characteristics are required to explain the returns of test assets with simple structure, such as the FF25 ME/BM portfolios. In contrast, more complex asset sets, such as those constructed via the P-Tree method, necessitate a richer set of explanatory variables. P-Tree (Cong et al., 2025) extends multi-way sorting by allowing nonlinear interactions among high-dimensional characteristics to construct a diverse and challenging test asset set. Its objective is to maximize the Sharpe ratio of the efficient portfolio spanned by the constructed assets.

Moreover, P-Tree builds assets sequentially: each new asset is designed to hedge the exposures of the previously generated ones, thereby improving the overall Sharpe ratio of the portfolio. As a result, the marginal contribution of each new asset diminishes with sequence order. Consequently, P-Trees with 100, 200, or 400 test assets pose different levels of difficulty for pricing models. In this section, we analyze how sparsity in characteristic-based models varies across these alternative test asset constructions.[5]

Table 4 reports sparsity estimates across various test asset sets, including both global sparsity and separate measures for mispricing (q_α) and factor loadings (q_β).

Panels A and B compare the P-Tree test assets and individual stocks, respectively.[6]

[5]For simplicity, we adopt a prior mean of 0.5 for q_α, q_β, or q in this table and the subsequent analyses.

[6]We begin by selecting stocks with at least ten years of return data. From this group, we identify the 500 stocks with the highest and lowest average market equity (ME), labeled as "Big 500" and "Small

The results show that explaining the average returns of the P-Tree 400 test assets requires a sparsity level similar to that of individual stocks.

In general, most models exhibit little sparsity in global inclusion probability (q) and in the sparsity level for factor loadings (q_β). This finding reinforces the view that the P-Tree assets are well diversified and that the efficient portfolio spanned by these test assets approximates the one generated directly from individual stocks.

Table 4: Sensitivity Analysis: Sparsity for Different Test Assets

This table reports estimates of q_α, q_β, and q across various test asset types and sample sizes. Lower values of q_α, q_β, and q indicate increased sparsity. The analysis covers portfolios with varying numbers of P-Tree assets, individual stocks, and three additional portfolio sets: FF25, LS61, and Bi357.

	Global prior			Separate priors			
	q	M_α	M_β	q_α	q_β	M_α	M_β
Panel A: P-Tree							
100	0.48	5	11	0.31	0.59	4	12
200	0.60	7	14	0.40	0.67	5	14
400	0.70	9	15	0.47	0.85	9	18
Panel B: Ind. Stock							
Small 500	0.62	11	13	0.51	0.65	9	13
Big 500	0.68	8	16	0.41	0.82	6	18
Panel C: Others							
FF25	0.41	1	10	0.20	0.50	1	10
LS61	0.67	4	17	0.24	0.83	2	17
Bi357	0.81	11	19	0.50	0.90	10	19

Second, Panel C reports results for three widely used test asset sets: the Fama-French 25 ME/BM portfolios (FF25), a long-short portfolio constructed from 61 characteristics (LS61), and a bivariate-sorted portfolio based on the same set of characteristics (Bi357).[7] The differences in sparsity across these portfolios highlight the critical role that the choice of test assets plays in shaping the structure of asset pricing models. In general, more diversified or complex test assets tend to yield denser models. For instance, when $K = 3$, the global sparsity values for FF25, LS61, and Bi357 are 0.41, 0.67, and 0.81, respectively. Notably, Bi357 relies on nearly all available characteristics to explain factor loadings—a natural consequence of its construction, which integrates

500." Small-cap stocks are typically more difficult to price due to lower liquidity and greater idiosyncratic risk.

[7]The $2 \times 3 \times 60$ classification is expected to yield 360 portfolios. However, due to three empty cells with no stock assignments, the final number of portfolios is 357.

information from each characteristic into the test asset formation process. Among the three, the FF25 portfolios yield the sparsest model.

Third, the number of test assets also affects the degree of sparsity. Panel A shows that increasing the number of P-Tree portfolios—from 100 to 200 to 400—results in progressively denser models. This is intuitive: as more test assets are introduced, each adds incremental pricing information that requires more characteristics to explain. Larger sets of test assets capture richer and more diverse patterns in the cross-section of returns, thereby increasing the complexity of the pricing problem. This aligns with our earlier observation that the marginal contribution of each additional P-Tree portfolio declines over time. The first 100 portfolios constructed by P-Tree are the most representative of the underlying return space; later portfolios provide diminishing incremental information.

Fourth, the pricing difficulty of the test assets also influences the number of selected characteristics, and this effect differs between characteristics driving mispricing and those driving factor loadings. To isolate the impact of difficulty while holding the number of assets constant, Panel B compares two groups of 500 stocks: one composed of large-cap stocks ("Big 500") and the other of small-cap stocks ("Small 500"). Small-cap stocks are generally harder to price. Accordingly, the model selects more characteristics to explain mispricing for the Small 500 group, under both prior specifications. Interestingly, the pattern is reversed for factor loadings: fewer characteristics are selected to drive beta for small-cap stocks, consistent with the idea that these stocks are less exposed to systematic risk.

Fifth, the characteristics that explain mispricing often differ from those that explain factor exposures, and the latter tend to be denser. For example, when using the P-Tree 400 as test assets, $q_\beta = 0.85$ while $q_\alpha = 0.47$. This pattern is consistent across most cases, suggesting that the characteristics affecting betas are largely distinct from those affecting alphas. A higher q_β implies that many characteristics are needed to explain systematic variation in returns. While q_β remains relatively stable across asset sets, q_α varies more substantially, reflecting heterogeneity in the presence and magni-

tude of mispricing.

Finally, the number of latent factors in the model also affects sparsity. Figure A.2 plots the posterior means of q, q_α, and q_β for different values of K when using P-Tree100 test assets. As K increases, q_α tends to decline, consistent with the idea that additional latent factors absorb more systematic risk, leaving less residual mispricing. In contrast, q_β and the global q do not exhibit a similar monotonic pattern. This is because q_β reflects the total number of effective characteristics across all factor loadings, which are influenced

4.4 Heterogeneous Roles of Characteristics

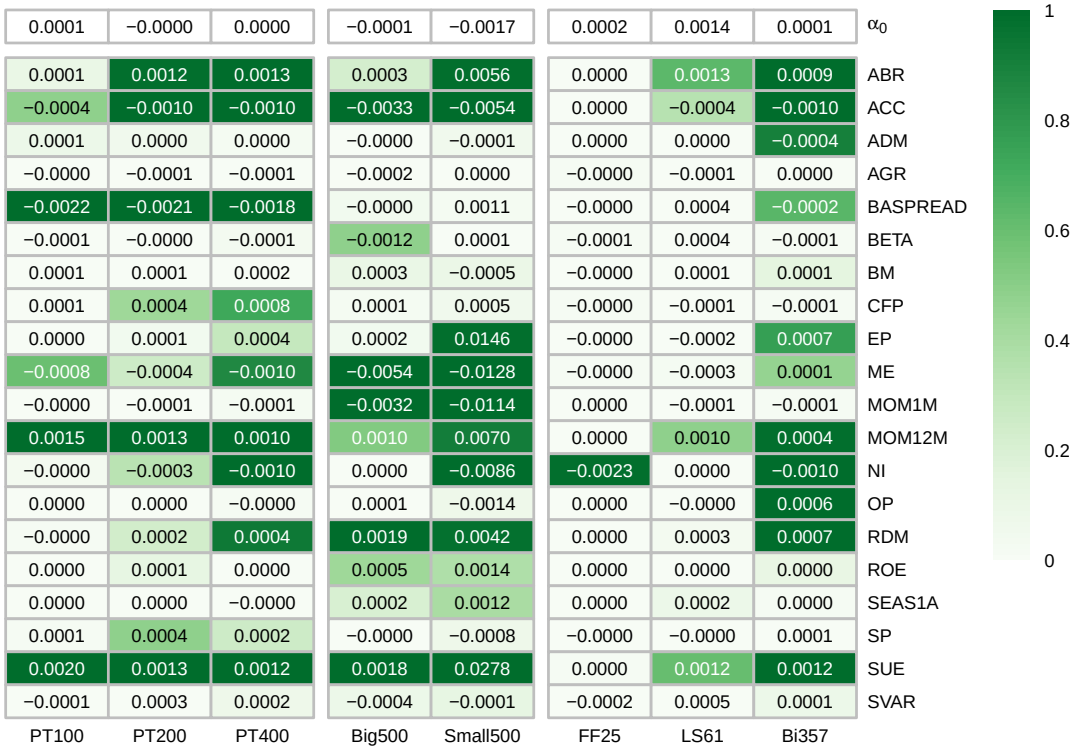
Up to this point, we have examined both global and component-specific sparsity in mispricing and factor loadings, documenting that the degree of sparsity varies meaningfully across the cross section. We now turn to identifying which characteristics are most influential in the model—specifically, those with nonzero coefficients in the mispricing component, $\boldsymbol{\alpha}(\mathbf{Z}_{i,t-1})$, and in the time-varying factor loadings, $\boldsymbol{\beta}(\mathbf{Z}_{i,t-1})$.

Figure 4 displays the posterior inclusion probabilities and corresponding coefficients ($\boldsymbol{\alpha}_{1,l}$ or $\boldsymbol{\beta}_{1,l}$) for each characteristic across a variety of test asset sets, along with the constant components not driven by characteristics (α_0 and β_0). A visual comparison between Panel (a) and Panel (b) reveals that while some characteristics influence both mispricing and factor loadings, others are specific to one component. This reflects both overlap and heterogeneity in characteristic relevance across the cross-section. The selection of characteristics is asset-dependent, highlighting the conditional and flexible nature of the model.

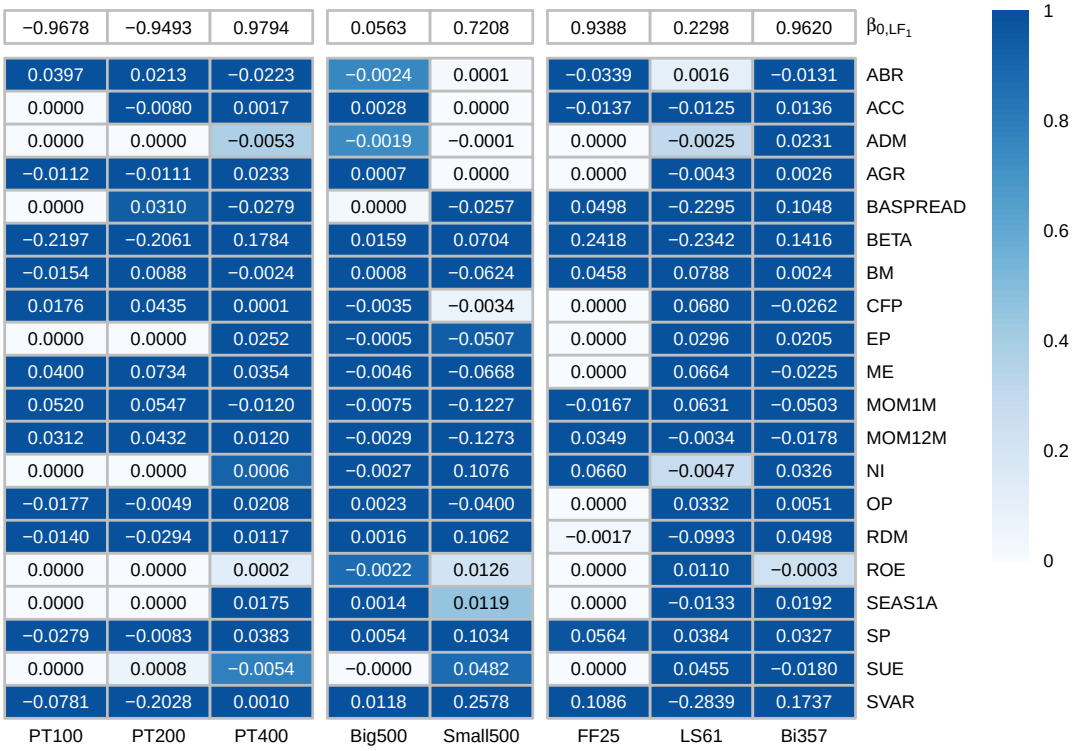
Several characteristics are consistently selected for factor loadings across all test asset sets, with posterior inclusion probabilities close to one. These include market beta (BETA), book-to-market ratio (BM), one-month momentum (MOM1M), twelve-month momentum (MOM12M), sales-to-price ratio (SP), and idiosyncratic volatility (SVAR). These characteristics are closely tied to systematic risk and capture how asset returns respond to macroeconomic and market-wide fluctuations. For example, mar-

Figure 4: Sensitivity Analysis: Heterogeneous Characteristics in Test Assets

This figure depicts selection probabilities of characteristics across test assets. The x-axis denotes test assets, and the y-axis lists characteristics. Panel (a) reports the probability that each of the 20 characteristics is selected to explain mispricing, while Panel (b) shows the probability of selection for explaining factor loadings. Results are shown for 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}

ket beta measures sensitivity to aggregate market returns, the book-to-market ratio reflects valuation-based exposures, and idiosyncratic volatility captures firm-specific risk. Their high relevance for factor loadings is consistent with their role in shaping time-varying exposures to systematic risk. In contrast, characteristics such as accruals (ACC) and standardized unexpected earnings (SUE) are selected for mispricing in nearly all test asset sets, except for the Fama-French 25 portfolios. These variables tend to reflect short-term pricing anomalies or temporary valuation errors, making them more likely to influence the alpha component rather than systematic exposure.

Other characteristics exhibit more context-dependent relevance. For example, asset turnover (ADM), asset growth (AGR), and the cash flow-to-price ratio (CFP) are not selected for the Fama-French 25 test assets but play a meaningful role in explaining returns in other portfolios. This reinforces the idea that no characteristic is universally irrelevant. Overall, approximately 60% of the available characteristics are selected in at least one component (mispricing or factor loadings) across the various test asset classes. Despite the heterogeneity in inclusion patterns, the Bayesian framework successfully detects and prioritizes asset-specific predictors, allowing the model to adapt to differing pricing environments and return-generating processes.

4.5 Resurrecting Conditional Observable Factors Model

Given our objective of assessing the explanatory power of factors in pricing the cross-section of returns, our primary metric of interest is the cross-sectional R^2 (CSR^2) excluding the mispricing component. Table 5 presents estimation results under various model specifications using our Bayesian method. To simplify the analysis, we impose a $\text{Beta}(5,5)$ prior on both q_α and q_β, which correspond to the inclusion probabilities for mispricing and factor loadings, respectively. In the table, $\beta_{0,\text{MKT}}$ denotes the intercept component of the market beta—that is, the portion not driven by characteristics. It reflects the common component of market exposure, while the heterogeneous, asset-specific part is captured by $\boldsymbol{\beta}_1(\mathbb{I}_K \otimes \mathbf{Z}_{i,t-1})$.

A comparison between Panel A and Panel C shows that jointly modeling observable and latent factors mitigates misspecification. When relying solely on observ-

able factors, the model is vulnerable to omitted variable bias due to unaccounted-for sources of systematic risk. This reflects the broader challenge in empirical asset pricing: the "true" model is never known, which has led to the proliferation of candidate factors in the so-called "factor zoo." Introducing latent factors provides a partial remedy by uncovering missing dimensions embedded in the high-dimensional characteristic space. In doing so, latent factors help recover factor structure hidden in the residuals of the observable-factor-only model, which can in turn improve the estimation of loadings on observable factors. As shown in the $\beta_{0,\text{MKT}}$ column, the inclusion of latent factors materially affects the estimate of the market beta, highlighting that both observable and latent sources influence asset exposures. That said, the use of latent factors does not eliminate omitted variable bias altogether.

The results also highlight the substantial gains in explanatory power when latent factors are introduced. As shown in the CSR^2 column, adding latent structure significantly enhances model fit. For example, the increase in CSR^2 between the single-factor market model and the two-factor model that combines the market and one latent factor reaches 38.94%. Although a single latent factor on its own yields relatively modest explanatory power, its interaction with the market factor generates performance comparable to more complex models, such as the Fama-French five-factor model (FF5) or a five-latent-factor model (LF5).

This improved fit is also reflected in reductions in pricing error. Specifically, models that incorporate latent factors achieve lower root mean squared errors (RMSE) of the mispricing term $\boldsymbol{\alpha}$, as shown in the final column of Table 5. For instance, comparing the market-only model (MKT) with the model including both market and five latent factors (MKT, LF5), the $\boldsymbol{\alpha}$ RMSE declines from 0.0032 to 0.0007. The latent-factor-only model (LF5) yields an intermediate RMSE of 0.0011, reinforcing that combining observable and latent factors provides superior pricing accuracy by better aligning the model with the data-generating process.

In addition, as the number of factors increases, we observe a progressive decline in the posterior means of q_α and q_β. This pattern indicates that as more variation

in returns is attributed to factors, the portion previously absorbed by mispricing or spurious factor loadings diminishes. Consequently, fewer characteristics are needed to explain these components, reflecting a shift in explanatory burden from idiosyncratic mispricing to systematic sources of variation.

Table 5: Augmented Observable Factor Models

This table reports estimation metrics from Sparse BayesIPCA and asset-specific regressions across various model specifications using P-Tree 100 portfolios. $\beta_{0,\text{MKT}}$ denotes the market factor coefficient not driven by characteristics; in Panel C, it represents the average market factor coefficient from asset-specific regressions. "$\boldsymbol{\alpha}$ RMSE" is the root mean squared error between mispricing ($\alpha_0 + \boldsymbol{\alpha}_1(\mathbf{Z}_{i,t-1})$) and zero, averaged over all samples. For brevity, "LF" denotes latent factors, with the following number indicating their count.

	CSR^2	TP.Sp	(q_α, q_β)	$\beta_{0,\text{MKT}}$	$\boldsymbol{\alpha}$ RMSE
Panel A: only obs					
MKT	14.93	0.57	0.45,0.63	1.15	0.0032
FF3	27.34	0.60	0.35,0.74	1.09	0.0026
FF5	50.38	1.13	0.26,0.61	1.07	0.0014
Panel B: only latent					
LF1	29.48	0.35	0.49,0.53	/	0.0036
LF3	44.96	1.50	0.33,0.43	/	0.0021
LF5	56.81	1.13	0.23,0.34	/	0.0011
Panel C: obs + latent					
MKT+LF1	53.87	0.87	0.31,0.65	1.14	0.0015
MKT+LF3	53.33	1.00	0.22,0.57	1.14	0.0007
MKT+LF5	56.45	1.39	0.24,0.46	0.98	0.0007
FF3+LF1	41.58	0.74	0.33,0.73	1.10	0.0014
FF3+LF3	53.49	1.14	0.19,0.66	1.03	0.0002
FF3+LF5	59.48	1.42	0.18,0.43	1.03	0.0000
FF5+LF1	50.55	1.23	0.33,0.65	1.06	0.0012
FF5+LF3	53.43	1.70	0.18,0.55	1.04	0.0001
FF5+LF5	60.33	1.53	0.18,0.42	0.95	0.0001
Panel D: uncond. model					
MKT	/	0.57	/	1.19	0.0060
FF3	11.53	0.60	/	1.12	0.0056
FF5	49.25	1.13	/	1.09	0.0042

From another perspective, comparing Panel D with Panel A in Table 5 reveals that the conditional factor model outperforms the unconditional model in terms of cross-sectional explanatory power, underscoring the value of our framework. By identifying

which characteristics drive mispricing and factor loadings, our method offers direct insight into the underlying structure of return variation. These insights are captured by the posterior means of q_α and q_β, which indicate the relevance of each characteristic in explaining the alpha and beta components, respectively. Figure 5 illustrates the posterior inclusion probabilities and corresponding coefficients for characteristics under two model specifications, with results presented separately for mispricing and factor loadings.

First, we observe substantial heterogeneity between the sets of characteristics driving mispricing and those influencing factor loadings. Second, the latent factor captures pricing dimensions not captured by the market factor. For instance, asset turnover (ADM) is not selected as significant in the market-only model (MKT) or even in the two-factor model that includes both the market and latent factors (MKT + LF1), but it is strongly associated with the latent factor loading. This suggests that latent factors extract information orthogonal to the market, enriching the model's explanatory structure.

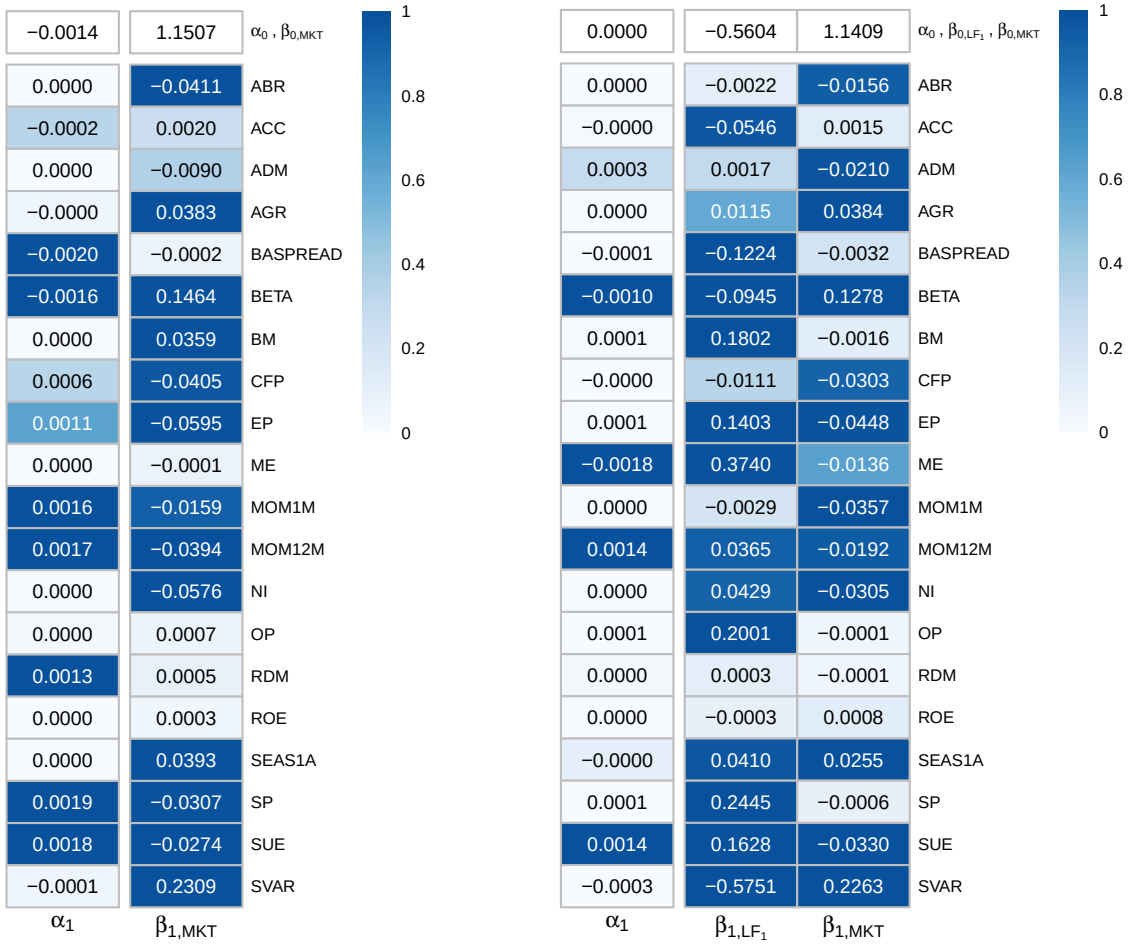
Moreover, including the latent factor also alters which characteristics drive the market factor. For example, analyst forecast revisions (ABR) and the cash flow-to-price ratio (CFP) are key drivers of market beta loadings in the single-factor model, but their importance diminishes when the latent factor is included. In contrast, characteristics such as accruals (ACC), market equity (ME), operating profitability (OP), and the bid-ask spread (BASPREAD) exhibit limited influence on the market factor in the single-factor model but become important once the latent factor is incorporated. This shift highlights the reallocation of characteristic-based explanatory power when latent factors are used to absorb omitted variation.

A comparison of Panels (a) and (b) in Figure 5 shows that the latent factor substantially reshapes the characteristics associated with market factor loadings, while its impact on the selection of mispricing characteristics is more limited. In addition, the estimated coefficients for market equity (ME) and idiosyncratic volatility (SVAR) are relatively large among the market beta drivers, suggesting that these variables have

strong and persistent associations with systematic risk as captured by the market factor.

Figure 5: Characteristics Importance in Alphas and Betas

This figure presents coefficients and selection probabilities for 20 characteristics explaining mispricing and factor loadings under two model specifications: market factor only (Panel a) and market plus latent factor (Panel b). α_1 denotes the posterior probability that a characteristic contributes to mispricing. The columns $\beta_{1,\text{MKT}}$ and β_{1,LF_1} report posterior inclusion probabilities for the market and first latent factors, respectively. Each cell shows the corresponding coefficient ($\alpha_{1,l}$ or $\beta_{1,l}$), with color intensity reflecting the magnitude of the selection probability.



(a) MKT

(b) MKT + LF1

5 Conclusion

Researchers often impose sparsity assumptions in asset pricing for parsimony and interpretability, yet the validity of this assumption is rarely tested. This paper provides a comprehensive examination of sparsity in conditional latent factor models, where both mispricing and factor loadings are driven by firm characteristics. We develop a

flexible Bayesian framework that does not impose sparsity ex ante but allows the data to determine the degree of sparsity in a probabilistic and interpretable manner.

Our approach estimates the global sparsity level via a parameter q, and also allows for separate sparsity levels in mispricing and factor loadings. This design facilitates a rigorous assessment of whether the asset pricing model is sparse or dense, and how sparsity varies with the choice of test assets, number of factors, and macroeconomic regimes. We find that while Fama–French 25 portfolios often support sparse representations, richer test assets—such as P-Tree portfolios and individual stocks—reveal much denser structures. Mispricing components tend to be sparser than factor loadings, and the two exhibit complementary sparsity dynamics. Moreover, sparsity varies over time, increasing during recessions when fewer characteristics remain relevant.

These findings highlight a key insight: the degree of sparsity is not fixed but conditional. Models that perform best typically strike a balance between sparsity and richness, depending on the complexity of the asset universe. Rigid assumptions—whether sparse or dense—can lead to misspecification or overfitting.

At the heart of our contribution is a resolution to what we call *Schrödinger's Sparsity*: the fundamental modeling dilemma of not knowing whether the pricing model is sparse or dense until the data are fully examined. Much like the paradox in quantum mechanics, the sparsity state remains indeterminate until we "open the box." Our framework opens that box in a statistically disciplined manner, allowing the data to determine the effective model complexity. This perspective not only bridges two opposing modeling philosophies but also offers a data-driven path forward in high-dimensional asset pricing.

References

Adrian, T., R. K. Crump, and E. Moench (2015). Regression-based estimation of dynamic asset pricing models. *Journal of Financial Economics 118*(2), 211–244.

Ang, A. and D. Kristensen (2012). Testing conditional factor models. *Journal of Financial*

Economics 106(1), 132–156.

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

Avramov, D. and J. C. Chao (2006). An exact Bayes test of asset pricing models with application to international markets. *Journal of Business 79*(1), 293–324.

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

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

Bali, T. G., R. F. Engle, and Y. Tang (2017). Dynamic conditional beta is alive and well in the cross section of daily stock returns. *Management Science 63*(11), 3760–3779.

Barillas, F. and J. Shanken (2017). Which alpha? *Review of Financial Studies 30*(4), 1316–1338.

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

Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*(1), 75–97.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*(1), 116–131.

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

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

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(2), 205–251.

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

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90*(432), 1313–1321.

Chib, S. and S. C. Smith (2024). Factor selection and structural breaks. Technical report, FEDS Working Paper.

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

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

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

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

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

Connor, G., M. Hagmann, and O. Linton (2012). Efficient semiparametric estimation of the Fama–French model and extensions. *Econometrica 80*(2), 713–754.

Cui, L., G. Feng, Y. Hong, and J. Yang (2023). Time-varying factor selection: A sparse fused GMM approach. Technical report, City University of Hong Kong.

Didisheim, A., S. B. Ke, B. T. Kelly, and S. Malamud (2023). Complexity in factor pricing models. Technical report, National Bureau of Economic Research.

Evgeniou, T., A. Gueciouer, and R. Prieto (2023). Uncovering sparsity and heterogeneity in firm-level return predictability using machine learning. *Journal of Financial and Quantitative Analysis 58*(8), 3384–3419.

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

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fan, J., Z. T. Ke, Y. Liao, and A. Neuhierl (2024). Structural deep learning in conditional asset pricing. Technical report, Princeton University.

Fan, J., Y. Liao, and W. Wang (2016). Projected principal component analysis in factor models. *Annals of Statistics 44*(1), 219.

Farago, A. and R. Tédongap (2018). Downside risks and the cross-section of asset returns. *Journal of Financial Economics 129*(1), 69–86.

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.

Feng, G., W. Lan, H. Wang, and J. Zhang (2023). Anomaly or risk factor? A stepwise evaluation. Technical report, City University of Hong Kong.

Ferson, W. E. and C. R. Harvey (1999). Conditioning variables and the cross section of stock returns. *Journal of Finance 54*(4), 1325–1360.

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

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

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

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

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

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.

Huber, F., G. Koop, and L. Onorante (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics 39*(3), 669–683.

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

Kalli, M. and J. E. Griffin (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics 178*(2), 779–793.

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.

Kim, S., R. A. Korajczyk, and A. Neuhierl (2021). Arbitrage portfolios. *Review of Financial Studies 34*(6), 2813–2856.

Koop, G. and D. Korobilis (2023). Bayesian dynamic variable selection in high dimensions. *International Economic Review 64*(3), 1047–1074.

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.

Lewellen, J. and S. Nagel (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics 82*(2), 289–314.

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

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

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

Appendix

A.I Simulation Evidence

We validate our Sparse BayesIPCA methodology through simulation. In a conditional factor model where factor loadings are determined by characteristics, the primary challenge lies in jointly modeling the cross-sectional and time-series relationships among characteristics while providing a plausible mechanism for generating dynamic betas. Fan et al. (2016) address this by calibrating nonlinear dynamics using cubic splines and sampling characteristics from a multivariate normal distribution. We take an alternative approach to this challenge.

We first generate characteristics, coefficients, and factors, then compute $(\boldsymbol{\alpha}_{i,t} + \boldsymbol{\beta}_{i,t}\mathbf{F}_t)$ and determine the size of $\epsilon_{i,t}$ based on a target signal-to-noise ratio (SNR), which in turn yields $r_{i,t}$. Specifically, the characteristics are sampled from a stationary VAR(1) process,

$$\mathbf{Z}_{i,t} = \mathbf{A} + \mathbf{B}\mathbf{Z}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}, \tag{A.1}$$

which ensures that the characteristics maintain both cross-sectional and time-series relationships across multiple quantities as they evolve over time.

Next, we sample coefficients $\Gamma = [\alpha_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1]$ from a normal distribution. Since both mispricing and factor loadings are assumed to be linearly related to characteristics, their values are computed from the affine functions $\boldsymbol{\alpha}_{i,t} = \alpha_0 + \boldsymbol{\alpha}_1\mathbf{Z}_{i,t-1}$ and $\boldsymbol{\beta}_{i,t} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1(\mathbb{I}_K \otimes \mathbf{Z}_{i,t-1})$. The factors $\mathbf{F}_t$ are independently generated from a multivariate normal distribution. Finally, we specify the SNR and determine the variance of $\epsilon_{i,t}$ conditional on the known value of $(\boldsymbol{\alpha}_{i,t} + \boldsymbol{\beta}_{i,t}\mathbf{F}_t)$, draw $\epsilon_{i,t}$ from a normal distribution, and obtain $r_{i,t} = \boldsymbol{\alpha}_{i,t} + \boldsymbol{\beta}_{i,t}\mathbf{F}_t + \epsilon_{i,t}$.

The process above defines a flexible conditional factor model for the data-generating process (DGP). It nests the static asset pricing model $\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{F}_t + \boldsymbol{\varepsilon}_t$ as a special case, and can be extended to accommodate dynamics in $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$. For example, macroeconomic variables could be used as drivers instead of characteristics, or the specification

could capture nonlinearities, such as $\boldsymbol{\alpha}_{i,t} = \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{Z}_{i,t-1} + \boldsymbol{\alpha}_2 \mathbf{Z}_{i,t-1}^2$.

We consider two DGP scenarios: DGP1 demonstrates the method's ability to identify the "useful" characteristics, while DGP2 evaluates its effectiveness in parameter estimation.

A.I.1 Identifying the "Useful" Characteristics

In DGP1, we generate 20 characteristics, of which only 5 are used to generate returns-these 5 are the "useful" characteristics. We consider both cases where the characteristics are cross-sectionally correlated and uncorrelated. Clearly, identifying the "useful" characteristics is more challenging when cross-sectional correlation is present.

We calibrate the model using monthly individual stock data from 1980 to 2023. The specific steps are as follows:

1. **Useful characteristics ($\tilde{\mathbf{Z}}$):** We first generate data for the five useful characteristics. Based on realized data, we regress each characteristic $z_{j,i,t}$ at time t on all available predictors at time $t-1$ using the following specification:

$$z_{j,i,t} = a_j + \sum_{j=1}^{20} b_j z_{j,i,t-1} + \varepsilon_{j,i,t}, \tag{A.2}$$

   to estimate the coefficient matrix $\mathbf{\Pi}$ and the residual covariance matrix $\mathbf{\Sigma}$. For the initial month, observations are drawn from a uniform distribution. Subsequent observations are generated recursively using the VAR(1) structure, where the coefficient matrix $\mathbf{\Pi}$ is applied to lagged characteristics, and shocks are sampled from a multivariate normal distribution with calibrated covariance matrix $\mathbf{\Sigma}$. Finally, we standardize each characteristic cross-sectionally to the range $[-1, 1]$.

2. **Other characteristics:** We consider two scenarios: (i) characteristics are cross-sectionally uncorrelated, and (ii) characteristics are cross-sectionally correlated. In the uncorrelated case, we use the VAR model in Equation A.2 while setting off-diagonal elements of $\mathbf{\Pi}$ and $\mathbf{\Sigma}$ to zero. In the correlated case, we construct 15

"useless" characteristics by injecting noise into the useful ones to target specific cross-sectional correlations. At each time t, we define the j-th "useless" characteristic as:

$$\widehat{\mathbf{z}}_j = \rho(\tilde{\mathbf{z}}_j - \bar{\tilde{\mathbf{z}}}_j) + \sqrt{\text{Var}(\tilde{\mathbf{z}}_j)(1-\rho^2)} \times C + \bar{\tilde{\mathbf{z}}}_j, \tag{A.3}$$

where C is an independent standard normal vector. This construction ensures that $\text{corr}(\tilde{\mathbf{z}}_j, \widehat{\mathbf{z}}_j) = \rho$. For each useful characteristic, we generate three noisy variants with $\rho \in \{0.5, 0.7, 0.9\}$, resulting in 15 additional characteristics.

3. **Coefficients (Γ):** We specify the values of α_0, $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_0$, and $\boldsymbol{\beta}_1$. The full coefficient vector Γ has dimension $(1 + \tilde{L} + \tilde{K} + \tilde{K} \times \tilde{L})$, where $\tilde{K}$ is the number of factors in the data-generating process. To draw the coefficients, we first sample $\gamma^2 \sim \mathcal{IG}\left(A_\gamma/2, B_\gamma/2\right)$ with $A_\gamma = 20$ and $B_\gamma = 1$. Then, each entry of Γ is drawn from $\mathcal{N}\left(0, \gamma^2\right)$.

4. **Factors (f):** Each factor $f_{k,t}$ ($k = 1, \ldots, \tilde{K}$) is independently drawn from a normal distribution:

$$f_{k,t} \sim \mathcal{N}\left(0, \sigma_f^2\right), \quad \text{where } \sigma_f^2 = 0.2. \tag{A.4}$$

5. **Excess return ($r_{i,t}$):** We generate the return signal by combining five calibrated characteristics $\tilde{\mathbf{Z}}$ and k factors according to:

$$\alpha_0 + \boldsymbol{\alpha}_1 \tilde{\mathbf{Z}}_{i,t-1} + \boldsymbol{\beta}_0 \mathbf{f}_t + \boldsymbol{\beta}_1 [\mathbf{f}_t \otimes \tilde{\mathbf{Z}}_{i,t-1}]. \tag{A.5}$$

To obtain $r_{i,t}$, we set the signal-to-noise ratio (SNR) to 1. The noise variance is computed as:

$$\epsilon^2_{\text{noise}} = \frac{\text{Var(signal)}}{\text{SNR}}. \tag{A.6}$$

We then sample $\epsilon_{i,t} \sim \mathcal{N}\left(0, \epsilon^2_{\text{noise}}\right)$ and compute the excess return as:

$$r_{i,t} = \alpha_0 + \boldsymbol{\alpha}_1 \tilde{\mathbf{Z}}_{i,t-1} + \boldsymbol{\beta}_0 \mathbf{f}_t + \boldsymbol{\beta}_1 [\mathbf{f}_t \otimes \tilde{\mathbf{Z}}_{i,t-1}] + \epsilon_{i,t}. \tag{A.7}$$

We consider between one and five useful factors in the simulations. For each

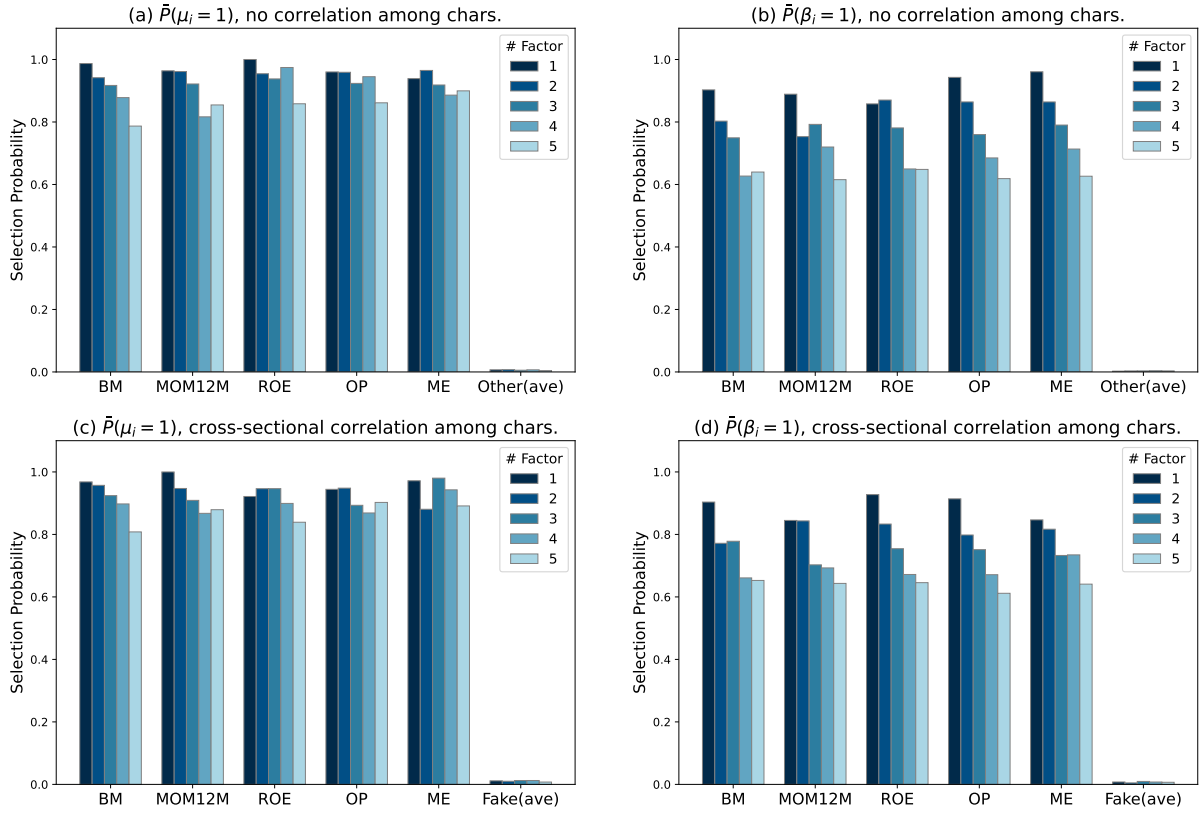
factor specification and scenario, we generate 50 datasets. For each dataset, we compute the selection probability of characteristics associated with $\boldsymbol{\alpha}_1$ as $P(\alpha_{1,i} = 1)$ for $i = 1, \ldots, 20$. Likewise, we compute the selection probability for each characteristic associated with $\boldsymbol{\beta}_1$ as $P(\beta_{1,i} = 1)$. When multiple latent factors are present, we compute $P(\beta_{1,i,f} = 1)$ for each factor f and take the average across factors:

$$P(\beta_{1,i} = 1) = \frac{1}{\tilde{K}} \sum_{f=1}^{\tilde{K}} P(\beta_{1,i,f} = 1). \tag{A.8}$$

Figure A.1 presents the results of our methodology in identifying useful characteristics under both cross-sectionally correlated and uncorrelated DGPs. The findings demonstrate that our method reliably recovers the informative characteristics in both settings.

Figure A.1: Selection Probability Estimates for Two Scenarios of DGP1

The four bar charts show the average selection probabilities of characteristics related to α_1 and β_1 under two scenarios. In panels (a) and (b), the first five labels on the x-axis correspond to the five "useful" characteristics in the DGP1, while the sixth, "Other(ave)," represents the average selection probability of the remaining 15 unrelated characteristics. In panels (c) and (d), "Fake(ave)" denotes the average selection probability of the remaining 15 correlated characteristics.



A.I.2 DGP2

In DGP2, we evaluate the accuracy of our method in estimating coefficients. As in Section A.I.1, we generate characteristic data, coefficients, and factor realizations, combine them, and add noise based on a target signal-to-noise ratio (SNR) to obtain returns. The key difference is that to facilitate comparison between estimated and true parameters, we directly specify the coefficients and limit the analysis to five characteristics, among which three influence mispricing and factor loadings. We consider three different factor settings: one with a single latent factor, one with a single observable factor, and one with both a latent and an observable factor. For each setting, we generate 200 simulated datasets.

We first sample the raw characteristics from a stationary vector autoregression

(VAR)(1) process: $\mathbf{Z}_{i,t} = \mathbf{A} + \mathbf{B}\mathbf{Z}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}$, where $\mathbf{A}$ is obtained from calibration, $\mathbf{B} = diag(0.8)$ and $\Sigma^z = diag(0.01)$. Then the latent factor is sample from $\mathcal{N}(\mathbf{0}, 0.02^2)$, and the observable factor(`MktRF`) is obtained from : $\texttt{MktRF}_t = e_0 + \eta_t$ where $e_0 = 0.6\%$ and $\eta_t \sim \mathcal{N}(0, 0.02^2)$. As for the coefficient ($\Gamma = [\alpha_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1]$), we set them as follows: $\alpha_0 = 0.1\% = 0.001$, $\alpha_{1,z} = 1\% = 0.01$, $\beta_{0,\text{mkt f}} = 1$, $\beta_{0,\text{latent f}} = 0.2$, and $\beta_{1,f,z} = 0.1$. We then generate the return signal by using calibrated characteristics and k factors as follows:

$$\alpha_0 + \boldsymbol{\alpha}_1 \tilde{\mathbf{Z}}^{\alpha}_{i,t-1} + \boldsymbol{\beta}_0 \mathbf{f}_t + \boldsymbol{\beta}_1 [\mathbf{f}_t \otimes \tilde{\mathbf{Z}}^{\beta}_{i,t-1}]$$

where $\tilde{\mathbf{Z}}^{\alpha}$ represents the `BM` and `ME` (size and value), while $\tilde{\mathbf{Z}}^{\beta}$ represents the `ME` and `MOM12M` (value and mom). To obtain $r_{i,t}$, we employ an SNR set to 10 and follow the same process in Appendix A.I.1.

Table A.1 presents the results of our methodology in estimation accuracy under the three different DGP2 settings. The results demonstrate that our method not only successfully selects the "useful" characteristics (with parameters unrelated to the characteristics close to zero and those related to the characteristics nonzero) but also ensures precise parameter estimation.

Table A.1: Estimation Accuracy for DGP2

This table presents the parameter estimates obtained from sparse BayesIPCA under three different model settings (each generating 200 simulated datasets) and their deviations from the true parameters. To illustrate the results more intuitively, the DGP2 specifies three characteristics for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, while the model estimation is conducted on data containing five characteristics. For simplicity, "LF" denotes latent factors, and "OF" represents observable factors. α_{1,z_i} represents the relationship between mispricing and the ith characteristic , while $\beta_{1,\text{LF(or OF)},z_i}$ represents the relationship between dynamic factor loading and the ith characteristic.

Coef.	# LF = 0, # OF =1			# LF = 1, # OF =0			# LF = 1, # OF =1		
	True paras.	Ave mse	Ave est.	True paras.	Ave mse	Ave est.	True paras.	Ave mse	Ave est.
α_0	0.001	0.000	0.001	-0.004	0.000	-0.003	-0.004	0.000	-0.003
α_{1,z_1}	0.010	0.000	0.010	0.010	0.000	0.010	0.010	0.000	0.010
α_{1,z_2}	0.010	0.000	0.010	0.008	0.000	0.008	0.008	0.000	0.009
α_{1,z_3}	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.000	-0.001
α_{1,z_4}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
α_{1,z_5}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\beta_{0,\text{LF}}$	/	/	/	0.816	0.000	0.827	0.816	0.003	0.858
$\beta_{0,\text{OF}}$	1.000	0.000	1.003	/	/	/	1.000	0.036	0.942
β_{1,LF,z_1}	/	/	/	0.000	0.000	0.000	0.000	0.001	-0.015
β_{1,LF,z_2}	/	/	/	0.408	0.000	0.407	0.408	0.009	0.331
β_{1,LF,z_3}	/	/	/	0.408	0.001	0.388	0.408	0.008	0.368
β_{1,LF,z_4}	/	/	/	0.000	0.000	0.000	0.000	0.000	0.001
β_{1,LF,z_5}	/	/	/	0.000	0.000	0.000	0.000	0.000	-0.003
β_{1,OF,z_1}	0.000	0.000	-0.001	/	/	/	0.000	0.000	0.000
β_{1,OF,z_2}	0.100	0.000	0.088	/	/	/	0.100	0.009	0.069
β_{1,OF,z_3}	0.100	0.000	0.095	/	/	/	0.100	0.006	0.071
β_{1,OF,z_4}	0.000	0.000	-0.002	/	/	/	0.000	0.000	0.000
β_{1,OF,z_5}	0.000	0.000	-0.002	/	/	/	0.000	0.000	0.000

Table A.2: Alpha Tests in Different Models

This table reports the results of univariate and multivariate tests on the estimated alpha vector $(\alpha_0, \boldsymbol{\alpha}_1)$, under different prior means (0.1, 0.5, and 0.9) imposed on the global q, as well as for models with different numbers of selected characteristics. The first merged column reports the number of nonzero elements in α_0 and $\boldsymbol{\alpha}_1$, whose 95% confidence intervals do not include zero, and the second merged column presents the p-values from Hotelling's t^2 test conducted on the joint alpha vector. xxx

		# α_0 and $\alpha_{1,i} \neq 0$			p-value		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Unrestricted # selected chars.							
q prior mean	0.1	10	5	1	0	0	0
	0.5	10	5	1	0	0	0
	0.9	10	5	1	0	0	0
Panel B: Fixed # selected chars.							
M	2	4	2	2	0	0	0
	10	14	4	3	0	0	0
	18	14	12	9	0	0	0
	20	21	18	16	0	0	0

Figure A.2: Sparsity Levels Across Varying Numbers of Latent Factors

The figure examines how the number of latent factors interacts with both local probabilities of inclusion (i.e., q_α and q_β) and the global probability of inclusion q.

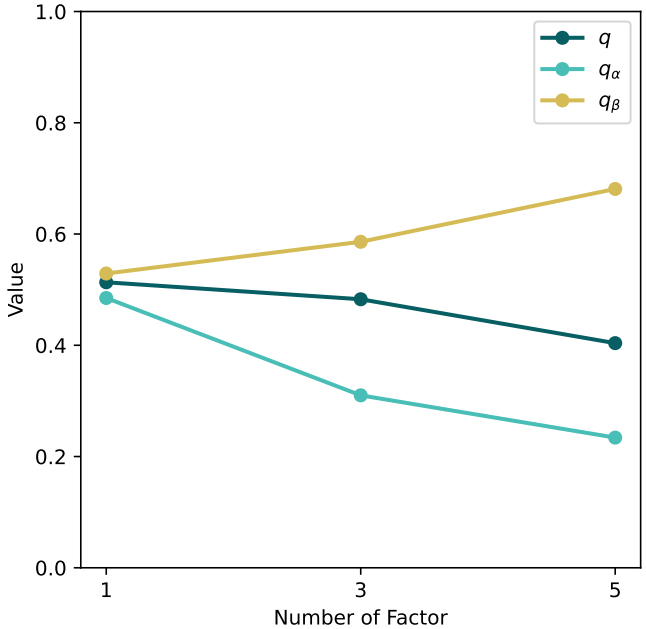


Table A.3: Number of Selected Characteristics in Different Models

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 the feature selection for both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ shares a common spike-and-slab prior, while Panel B presents the case where separate spike-and-slab priors are used for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The number of selected characteristics is calculated by counting how many characteristics have a selection probability greater than or equal to 0.5. The selection probability for each characteristic is the posterior mean of its $p(z\,|\sim)$.

		M_α			M_β		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
Panel A: Global prior							
q prior mean	0.1	10	5	1	10	11	9
	0.5	10	5	2	10	11	9
	0.9	10	5	1	11	11	9
Panel B: Separate priors							
(q_α prior mean, q_β prior mean)	0.1,0.1	10	5	1	10	11	10
	0.5,0.1	10	5	1	10	11	10
	0.9,0.1	10	5	1	10	11	10
	0.1,0.5	10	4	1	10	12	10
	0.5,0.5	10	4	2	10	12	14
	0.9,0.5	10	5	2	10	11	10
	0.1,0.9	10	5	1	11	11	11
	0.5,0.9	10	4	2	11	12	14
	0.9,0.9	10	5	2	11	11	14

Figure A.3: Histograms of Some Selected Characteristics

This figure illustrates a subset of characteristics from the P-Tree100 dataset. Specifically, `BM`, `ME`, `MOM12M`, `ROE`, and `OP` correspond to the factors in the Fama-French five-factor model.

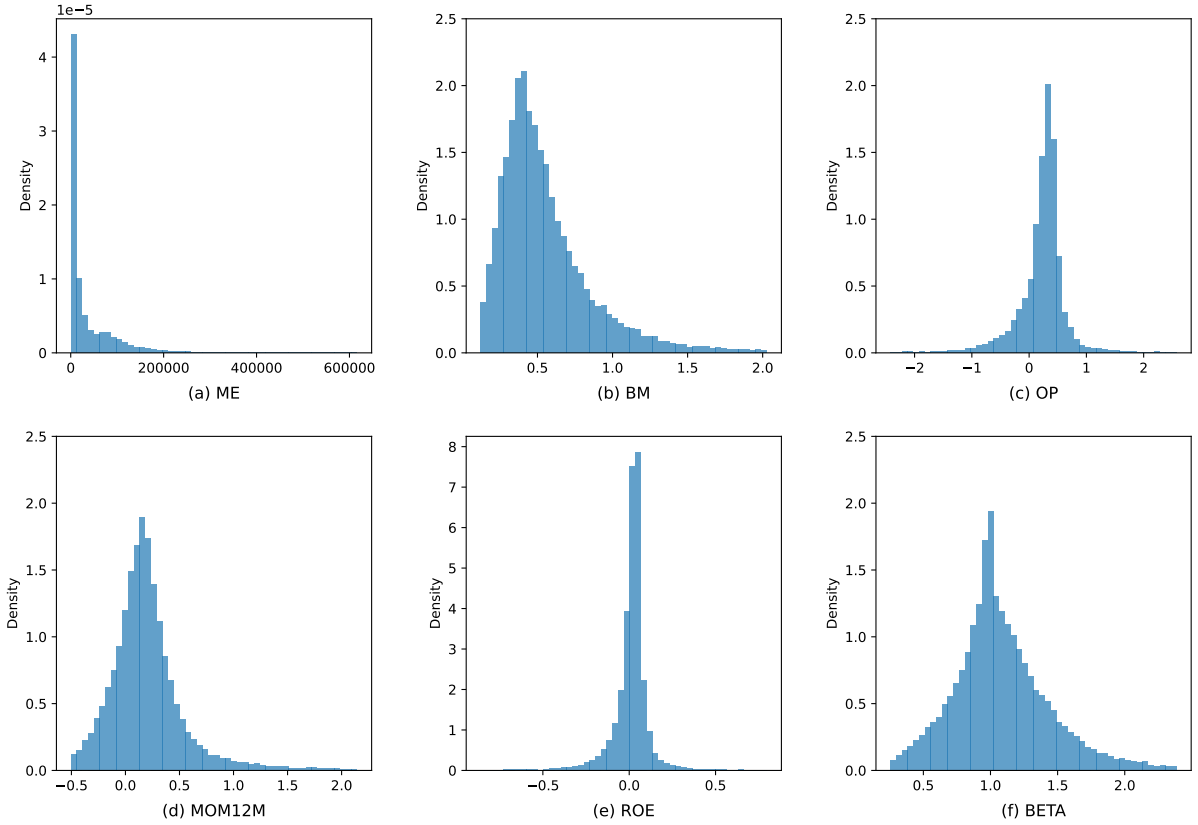


Table A.4: Equity Characteristics

This table lists the description 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 equity	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	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

A.II Prior setting and Gibbs sampler

The posterior inference of parameters for the (Sparse) BayesIPCA model is conducted using Markov Chain Monte Carlo (MCMC) methods with a Gibbs sampler. For simplicity, when conditional on $\mathbf{f}_t$ and $\mathbf{z}_{i,t-1}$, we express the model in Equation 2 as

$$r_{i,t} = \mathcal{W}_{i,t}\Gamma + \epsilon_{i,t} \tag{A.9}$$

where $\mathcal{W}_{i,t} = [1, \mathbf{z}_{i,t-1}, \mathbf{f}_t, \mathbf{f}_t \otimes \mathbf{z}_{i,t-1}]$ and $\Gamma = [\alpha_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1]$.

Suppose $i = 1, \cdots, N$ assets, each has T_i observations, then $\mathbf{R}_i = \mathcal{W}_i\Gamma + \boldsymbol{\epsilon}_i$, and $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma_i^2)$. Stacking all the asset returns together, we have

$$\mathbf{R} = \mathcal{W}\Gamma + \boldsymbol{\epsilon}, \tag{A.10}$$

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

The likelihood is

$$p(\mathbf{R} \mid \mathbf{Z}, \mathbf{F}, \Theta) = \prod_{i=1}^{N} (2\pi\sigma_i^2)^{-\frac{T_i}{2}} \exp\left[-\frac{1}{2\sigma_i^2}(\mathbf{R}_i - \mathcal{W}_i\Gamma)^\top(\mathbf{R}_i - \mathcal{W}_i\Gamma)\right] \tag{A.11}$$

where $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_L)_{\sum_{i=1}^N T_i \times L}$, $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_K)_{T \times K}$, $\Theta = [z, \gamma^2, q, \Gamma, \sigma_i^2, \xi^2]$.

For the Sparse BayesIPCA model, the priors for the parameters are:

$$\begin{aligned}
\sigma_i^2 &\overset{iid}{\sim} \mathcal{IG}(v_{0,i}/2, S_{0,i}/2) \\
\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1 &\overset{iid}{\sim} \begin{cases} \mathcal{N}(0, \gamma^2) \text{ with prob } q \\ 0 \text{ with prob } 1-q \end{cases} \qquad \gamma^2 \sim \mathcal{IG}(A_\gamma/2, B_\gamma/2) \\
\alpha_0, \boldsymbol{\beta}_0 &\overset{iid}{\sim} \mathcal{N}\left(0, \xi^2\right), \qquad \xi^2 \sim \mathcal{IG}(C/2, D/2)
\end{aligned}$$

If we wish to explore a specific perspective of investors, we set $q = q_{\text{fixed}}$. Regarding the prior for q, if we draw q globally, then $q \sim \text{Beta}(a_q, b_q)$.

If we draw q separately, then $q_\alpha \sim \text{Beta}(a_{q_\alpha}, b_{q_\alpha})$ and $q_\beta \sim \text{Beta}(a_{q_\beta}, b_{q_\beta})$. Note that

in this case, the spike-and-slab prior and the corresponding γ^2 should be separated:

$$\boldsymbol{\alpha}_1 \overset{iid}{\sim} \begin{cases} \mathcal{N}\left(0, \gamma_\alpha^2\right) \text{ with prob } q_\alpha \\ 0 \text{ with prob } 1 - q_\alpha \end{cases} \qquad \gamma_\alpha^2 \sim \mathcal{IG}\left(A_{\gamma_\alpha}/2, B_{\gamma_\alpha}/2\right)$$

$$\boldsymbol{\beta}_1 \overset{iid}{\sim} \begin{cases} \mathcal{N}\left(0, \gamma_\beta^2\right) \text{ with prob } q_\beta \\ 0 \text{ with prob } 1 - q_\beta \end{cases} \qquad \gamma_\beta^2 \sim \mathcal{IG}\left(A_{\gamma_\beta}/2, B_{\gamma_\beta}/2\right)$$

The priors related to the variances of coefficients are set as follows: $A_\gamma = A_{\gamma_\alpha} = A_{\gamma_\beta} = 10$, $B_\gamma = B_{\gamma_\alpha} = B_{\gamma_\beta} = 1$, $C = 4$ and $D = 1$. For the priors related to the probability that a characteristic is selected, the default setting (unless otherwise specified) is $a_q = a_{q_\alpha} = a_{q_\beta} = b_q = b_{q_\alpha} = b_{q_\beta} = 5$. When we target prior means of 0.1 or 0.9 for q, q_α or q_β, we set $a_q = a_{q_\alpha} = a_{q_\beta} = 1$ and $b_q = b_{q_\alpha} = b_{q_\beta} = 9$ for a prior mean of 0.1, or $a_q = a_{q_\alpha} = a_{q_\beta} = 9$ and $b_q = b_{q_\alpha} = b_{q_\beta} = 1$ for a prior mean of 0.9. As for the priors on σ_i^2, we run asset-specific regressions of each asset's return on the initial values of $\mathbf{F}$, calculate the residual variances from these regressions, and use them to determine the corresponding values of $v_{0,i}$ and $S_{0,i}$.

Define latent variable $z_l^\alpha, 1 \leq l \leq L$ as the selection indicator for $\boldsymbol{\alpha}_1$ and $z_{l,k}^\beta, 1 \leq l \leq L, 1 \leq k \leq K$ as the selection indicator for $\boldsymbol{\beta}_1$. $\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1$ contains $(L + K \times L)$ coefficients in total. Naturally, let $s(z^\alpha)$ and $s(z^\beta)$ represent number of selected variables among $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ in the model, i.e., $s(z^\alpha) = \sum_l^L z_l^\alpha$ and $s(z^\beta) = \sum_{l,k}^{LK} z_{l,k}^\beta$. Let $s(z)$ represent the number of selected variables in the whole model, where $s(z) = s(z^\alpha) + s(z^\beta)$.

Then the full conditionals are given as follows:

1. Sampling z_l^α and $z_{l,k}^\beta$.

   z_l^α and $z_{l,k}^\beta$ jointly indicate variables used in the regression (A.9). Thus for simplicity, we use the notation $z = (z^\alpha, z^\beta)$, and $\tilde{\mathcal{W}}_i$, $\tilde{\Gamma}$ indicate corresponding selected variables and their coefficients, which are subsets of $\mathcal{W}_i$ and Γ.

$$p(z \mid -) \propto q^{s(z)}(1-q)^{L+KL-s(z)} \left(\frac{1}{\gamma^2}\right)^{\frac{s(z)}{2}} \times |V_1|^{\frac{1}{2}} \left[\prod_{i=1}^{N} \left(\frac{1}{\sigma_i^2}\right)^{\frac{T_i+v_{0,i}}{2}+1}\right] \times \exp\left(-\sum_{i=1}^{N} \frac{S_i + S_{0,i}}{2\sigma_i^2} - \frac{1}{2}\sum_{i=1}^{N}\left(\left(\sigma_i^{-2}\tilde{\mathcal{W}}_i^\top \mathbf{R}_i\right)^\top \hat{\tilde{\Gamma}}_i\right) + \frac{1}{2}\tilde{\Gamma}_1^\top V_1^{-1}\tilde{\Gamma}_1\right) \tag{A.12}$$

where

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

When we need to decide whether $z_i = 1$ or $z_i = 0$, we first use Equation A.12 to compute the respective posterior probabilities, $\pi(z_i = 1 \mid -)$ and $\pi(z_i = 0 \mid -)$. Next, we calculate the relevant probability, compare it with a random draw from a uniform distribution, and follow Equation A.13 for the comparison logic.

$$z_i = \begin{cases} 1 & \text{if } U \leq P(z_i = 1) = \frac{\pi(z_i=1|-)}{\pi(z_i=0|-)+\pi(z_i=1|-)} \\ 0 & \text{Otherwise.} \end{cases} \tag{A.13}$$

Notably, when considering q_α and q_β instead of a global q, the posterior of z is

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

and $A = \text{diag}\left(1/\xi^2, \mathbb{I}_{s(z^\alpha)}/\gamma_\alpha^2, \mathbb{I}_K/\xi^2, \mathbb{I}_{s(z^\beta)}/\gamma_\beta^2\right)$.

2. Sampling q

$$q \mid z \sim \text{Beta}\left(s(z^\alpha) + s(z^\beta) + a_q, L + KL - s(z^\alpha) - s(z^\beta) + b_q\right).$$

or

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

3. Sampling γ^2

$$\gamma^2 \mid z, \tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\beta}}_1 \sim \mathcal{IG}\left(\frac{A_\gamma + s(z^\alpha) + s(z^\beta)}{2}, \frac{B_\gamma + \tilde{\boldsymbol{\beta}}_1^\top \tilde{\boldsymbol{\beta}}_1 + \tilde{\boldsymbol{\alpha}}_1^\top \tilde{\boldsymbol{\alpha}}_1}{2}\right).$$

or

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

where $\tilde{\boldsymbol{\alpha}}_1$ and $\tilde{\boldsymbol{\beta}}_1$ indicate subvector of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ for the selected coefficients only, with dimension $s(z^\alpha)$ and $s(z^\beta)$ respectively.

4. Sampling ξ^2

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

5. Sampling σ_i^2

$$\sigma_i^2 \mid z, \tilde{\Gamma} \sim \mathcal{IG}\left(\frac{v_{0,i} + T_i}{2}, \frac{S_{0,i} + S_i + \left(\tilde{\Gamma} - \hat{\tilde{\Gamma}}_i\right)^\top \tilde{\mathcal{W}}_i^\top \tilde{\mathcal{W}}_i \left(\tilde{\Gamma} - \hat{\tilde{\Gamma}}_i\right)}{2}\right).$$

6. Sampling $\tilde{\Gamma}$, i.e., the coefficients of non-zero terms:

$$\tilde{\Gamma} \mid z, \sigma_i^2, \gamma^2, \xi^2 \sim \mathcal{MVN}\left(\tilde{\Gamma}_1, V_1\right).$$

7. Sapming $\mathbf{f}_t$. Noted that sampling latent factor $\mathbf{F}$ conditional on coefficients and data is standard in the Bayesian literature. Basically, we evaluate the joint distribution of $(\mathbf{R}, \mathbf{F})$ and sample $\mathbf{F}$ from the conditional distribution $\mathbf{F} \mid \mathbf{R}$. For each time t, once conditional on $\boldsymbol{\Theta}$, the dynamic of alpha ($\boldsymbol{\alpha}(\mathbf{z}_{i,t-1})$) and beta ($\boldsymbol{\beta}(\mathbf{z}_{i,t-1})$) are fixed, and

the joint distribution of $\mathbf{f}_t$ and $\mathbf{r}_t$ is

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{r}_t \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} \mathbb{I} & \boldsymbol{\beta}' \\ \boldsymbol{\beta} & \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma} \end{pmatrix} \right] \tag{A.14}$$

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

$$\begin{aligned} \mathbb{E}(\mathbf{f}_t \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{r}_t) &= \boldsymbol{\beta}' \left(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}\right)^{-1} (\mathbf{r}_t - \boldsymbol{\alpha}) \\ \text{Cov}(\mathbf{f}_t \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{r}_t) &= \mathbb{I} - \boldsymbol{\beta}' \left(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}\right)^{-1} \boldsymbol{\beta} \end{aligned} \tag{A.15}$$

To simplify computation, we apply Woodbury's matrix identity,

$$\left(\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}\right)^{-1} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}\left(\mathbb{I} + \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}\right)^{-1}\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1} \tag{A.16}$$

For BayesIPCA without considering sparsity, the prior is specified as follows:

$$\sigma_i^2 \overset{iid}{\sim} \mathcal{IG}\left(v_{0,i}/2, S_{0,i}/2\right), \quad \alpha_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_1 \overset{iid}{\sim} \mathcal{N}\left(0, \xi^2\right),$$

and the Gibbs sampler is

1. Sampling ξ^2

$$\xi^2 \mid \alpha_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_1 \sim \mathcal{IG}\left(\frac{C + (1 + L + K + K \times L)}{2}, \frac{D + \alpha_0^2 + \boldsymbol{\beta}_0^\top\boldsymbol{\beta}_0 + \boldsymbol{\alpha}_1^\top\boldsymbol{\alpha}_1 + \boldsymbol{\beta}_1^\top\boldsymbol{\beta}_1}{2}\right).$$

2. Sampling Γ

$$\Gamma \mid \sigma_i^2, \xi^2 \sim \mathcal{MVN}\left(\Gamma_1, V_1\right).$$

- $\Gamma_1 = V_1\left(\sum_{i=1}^N \sigma_i^{-2}\mathcal{W}_i^\top\mathbf{R}_i\right)$ is the posterior mean for Γ
- $\widehat{\Gamma}_i = (\mathcal{W}_i^\top\mathcal{W}_i)^{-1}\mathcal{W}_i^\top\mathbf{R}_i$ is the OLS estimator for the i-th asset
- $S_i = \left(\mathbf{R}_i - \mathcal{W}_i\widehat{\Gamma}_i\right)^\top\left(\mathbf{R}_i - \mathcal{W}_i\widehat{\Gamma}_i\right)$
- $V_1 = \left(\sum_{i=1}^N \sigma_i^{-2}\mathcal{W}_i^T\mathcal{W}_i + A\right)^{-1}$ and $A = \text{diag}\left(\mathbb{I}_{(1+L+K+K\times L)}/\xi^2\right)$

3. Sampling σ_i^2

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

A.II.1 Metropolis-Hastings algorithm

Algorithm Metropolis-Hastings for Variable Selection

procedure DRAWZ(state)
  Input State with current variable selection $\boldsymbol{z}_x$, data matrices for all regressor $\boldsymbol{X}$, asset returns $\boldsymbol{R}$, parameters M_α, M_β, q_α, q_β, γ_α, γ_β
  Outcome Updated variable selection $\boldsymbol{z}_x$ with exactly M_α alpha and M_β beta elements per factor
  Validate initial $\boldsymbol{z}_x$ has M_α alpha and M_β beta elements per factor
  if validation fails **then**
    return
  end if
  Compute current state log-likelihood $\log p(\boldsymbol{z}_x|-)$
  for each stock i from 1 to N **do** ▷ Parallelized loop
    Select variables according to $\boldsymbol{z}_x$ indices $\mathcal{I}_0$
    Update $\log p(\boldsymbol{z}_x|-) \leftarrow \log p(\boldsymbol{z}_x|-) - \frac{S_0+s}{2\sigma_i^2} - \frac{(\boldsymbol{X}_i^T \boldsymbol{R}_i)^T \hat{\boldsymbol{\beta}}_0}{2\sigma_i^2}$ by Equation A.12
  end for
  Generate proposal $\boldsymbol{z}_x^{\text{proposal}}$ via randomly removing one selected variable and randomly adding one previously unselected variable
  Validate $\boldsymbol{z}_x^{\text{proposal}}$ has M_α chars. and M_β chars. elements per factor
  if validation fails **then**
    return
  end if
  Compute proposal state log-likelihood $\log p(\boldsymbol{z}_x^{\text{proposal}}|-)$ similarly
  Compute Metropolis-Hastings ratio:

$$\log \text{MH} = \log p(\boldsymbol{z}_x^{\text{proposal}}|-) - \log p(\boldsymbol{z}_x|-)$$

  Sample $u \sim \text{Uniform}(0,1)$
  if $u \leq \text{MH}$ **then**
    $\boldsymbol{z}_x \leftarrow \boldsymbol{z}_x^{\text{proposal}}$
  end if
  Update $\boldsymbol{z}_x$
  return $\boldsymbol{z}_x$
end procedure

A.III Derivation of Marginal Likelihood

Since our model lacks a closed-form expression for the marginal likelihood, we use posterior means of obtaining from the Gibbs sampler as the selection of $\Theta^* = \left(\gamma^{2*}, q^*, z_l^{\alpha*}, z_{l,k}^{\beta*}, \Gamma^*, \sigma_i^{2*}, \xi^{2*}\right)$. Note that when calculating $z^* = (z^{\alpha*}, z^{\beta*})$, we first determine the posterior mean of each z_l^{α} (or, $z_{l,k}^{\beta}$). If the posterior mean of z_l^{α} (or, $z_{l,k}^{\beta}$) is greater than 0.5, the corresponding $z_l^{\alpha*}$ (or, $z_{l,k}^{\beta*}$) is set to 1; otherwise, it is set to 0.

For given parameters Ω_ω^*, following Chib (1995), the marginal likelihood is

$$m(\mathbf{R} \mid M_\omega) = \frac{p(\mathbf{R} \mid \Omega_\omega^*, M_\omega) p(\Omega_\omega^* \mid M_\omega)}{p(\Omega_\omega^* \mid \mathbf{R}, M_\omega)} \tag{A.17}$$

where $p(\Omega_\omega^* \mid M_\omega)$ is the density function of the data $\mathbf{R} = (R_1, \ldots, R_N)^\top$ under model M_ω given the model-specific parameter matrix Ω_ω.

With a given Ω_ω^*, the first component of numerator Equation A.17 is

$$p(\mathbf{R} \mid \Omega_\omega^*, M_\omega) = \prod_{i=1}^{N} (2\pi\sigma_i^{2*})^{-\frac{T_i}{2}} \exp\left[-\frac{1}{2\sigma_i^{2*}} (\mathbf{R}_i - \mathcal{W}_i\Gamma^*)^\top (R_i - \mathcal{W}_i\Gamma^*)\right], \tag{A.18}$$

and the second component of the numerator in Equation A.17 is derived from:

$$\begin{aligned} p(\Omega_\omega^* \mid M_\omega) = p(z^{\alpha*}) \times p(z^{\beta*}) \times p(\boldsymbol{\alpha}_1^* \mid \gamma^{2*}, q^*, z^{\alpha*}) \times p(\boldsymbol{\beta}_1^* \mid \gamma^{2*}, q^*, z^{\beta*}) \\ \times p(\alpha_0^*) \times p(\boldsymbol{\beta}_0^*) \times p(\sigma^{2*}) \times p(\xi^{2*}) \times p(\gamma^{2*}) \times p(q^*). \end{aligned} \tag{A.19}$$

The calculation of the denominator in Equation A.17 is as follows. According to Chib (1995), decompose the posterior density into the product of conditional densities and utilize the subsequent reduced Gibbs run. Denote $\mathbf{Y} = (\mathbf{R}, M_\omega)$ and $\tilde{\Gamma} = (\alpha_0, \tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\beta}_0, \tilde{\boldsymbol{\beta}}_1)$. Thus, the decomposition of the posterior is:

$$\begin{aligned} p(\Omega_\omega^* \mid \mathbf{Y}) &= p(z^*, \tilde{\Gamma}^*, \sigma^{2*}, \gamma^{2*}, \xi^{2*}, q^* \mid \mathbf{Y}) \\ &= p(z^* \mid \mathbf{Y}) \times p(\tilde{\Gamma}^* \mid z^*, \mathbf{Y}) \times p(\sigma^{2*} \mid z^*, \tilde{\Gamma}^*, \mathbf{Y}) \times p(\gamma^{2*} \mid z^*, \tilde{\Gamma}^*, \sigma^{2*}, \mathbf{Y}) \\ &\times p(\xi^{2*} \mid z^*, \tilde{\Gamma}^*, \sigma^{2*}, \gamma^{2*}, \mathbf{Y}) \times p(q^* \mid z^*, \tilde{\Gamma}^*, \sigma^{2*}, \gamma^{2*}, \xi^{2*}, \mathbf{Y}) \end{aligned} \tag{A.20}$$

Using the following formula, we can compute each component of the posterior distribution based on the specific values of each MCMC sample.

$$
\begin{aligned}
\widehat{p}(z^* \mid \mathbf{Y}) &= \frac{1}{G}\sum_{g=1}^{G} p(z^* \mid \sigma^{2(g)}, \gamma^{2(g)}, \xi^{2(g)}, q^{(g)}, \mathbf{Y}) \\
\widehat{p}(\tilde{\Gamma}^* \mid z^*, \mathbf{Y}) &= \frac{1}{G}\sum_{g=1}^{G} p(\tilde{\Gamma}^* \mid z^*, \sigma^{2(g)}, \gamma^{2(g)}, \xi^{2(g)}, \mathbf{Y}) \\
\widehat{p}(\sigma_i^{2*} \mid z^*, \tilde{\Gamma}^*, \mathbf{Y}) &= p(\sigma_i^{2*} \mid z^*, \tilde{\Gamma}^*, \mathbf{Y}) \\
\widehat{p}(\gamma^{2*} \mid z^*, \tilde{\Gamma}^*, \sigma^{2*}, \mathbf{Y}) &= p(\gamma^{2*} \mid z^*, \tilde{\boldsymbol{\alpha}}_1^*, \tilde{\boldsymbol{\beta}}_1^*, \mathbf{Y}) \\
\widehat{p}(\xi^{2*} \mid z^*, \tilde{\Gamma}^*, \sigma^{2*}, \gamma^{2*}, \mathbf{Y}) &= p(\xi^{2*} \mid \alpha_0^*, \boldsymbol{\beta}_0^*, \mathbf{Y}) \\
\widehat{p}(q^* \mid z^*, \tilde{\Gamma}^*, \sigma^{2*}, \gamma^{2*}, \xi^{2*}, \mathbf{Y}) &= p(q^* \mid z^*, \mathbf{Y})
\end{aligned}
\tag{A.21}
$$

To calculate the marginal likelihood, we need to write down the full posterior densities for the parameters. Here, we only present the posterior components that need to be included in the posterior distribution provided by Appendix A.II.

- z^*

$$
\begin{aligned}
p(z \mid -) = \prod_{i=1}^{N} (2\pi)^{\frac{-T_i}{2}} \times \prod_{i=1}^{N} \frac{S_0/2^{v_0/2}}{\Gamma(v_0/2)} \times q^{s(z)}(1-q)^{L+KL-s(z)} \left(\frac{1}{\gamma^2}\right)^{\frac{s(z)}{2}} \times |V_1|^{\frac{1}{2}} \left[\prod_{i=1}^{N} \left(\frac{1}{\sigma_i^2}\right)^{\frac{T_i+v_0}{2}+1}\right] \\
\times \exp\left(-\sum_{i=1}^{N} \frac{S_i+S_0}{2\sigma_i^2} - \frac{1}{2}\sum_{i=1}^{N}\left(\left(\sigma_i^{-2}\tilde{\mathcal{W}}_i^\top R_i\right)^\top \hat{\tilde{\Gamma}}_i\right) + \frac{1}{2}\tilde{\Gamma}_1^\top V_1^{-1}\tilde{\Gamma}_1\right)
\end{aligned}
$$

- $\tilde{\Gamma}^*$

$$
p(\tilde{\Gamma}^* \mid z^*, \sigma^2, \gamma^2, \xi^2) = (2\pi)^{-\frac{1+s(z^\alpha)+K+s(z^\beta)}{2}} |V_1|^{-1/2} \exp\left(-\frac{1}{2}(\tilde{\Gamma}-\tilde{\Gamma}_1)^\top V_1^{-1}(\tilde{\Gamma}-\tilde{\Gamma}_1)\right)
$$

The previous description pertains to the case of a global q. When q is separated, the calculation process for the marginal likelihood should be modified. Equation A.19 becomes

$$
\begin{aligned}
p(\Omega_\omega^* \mid M_\omega) = p(z^{\alpha*}) \times p(z^{\beta*}) \times p(\boldsymbol{\alpha}_1^* \mid \gamma_\alpha^{2*}, q^*, z^{\alpha*}) \times p(\boldsymbol{\beta}_1^* \mid \gamma_\beta^{2*}, q^*, z^{\beta*}) \\
\times p(\alpha_0^*) \times p(\boldsymbol{\beta}_0^*) \times p(\sigma^{2*}) \times p(\xi^{2*}) \times p(\gamma_\alpha^{2*}) \times p(\gamma_\beta^{2*}) \times p(q_\alpha^*) \times p(q_\beta^*),
\end{aligned}
$$

and the decomposition of the posterior is:

$$
\begin{aligned}
p(\Omega_\omega^* \mid \mathbf{Y}) &= p(z^*, \tilde{\Gamma}^*, \sigma^{2*}, \gamma_\alpha^{2*}, \gamma_\beta^{2*}, \xi^{2*}, q_\alpha^*, q_\beta^* \mid \mathbf{Y}) \\
&= p(z^* \mid \mathbf{Y}) \times p(\tilde{\Gamma}^* \mid z^*, \mathbf{Y}) \times p(\sigma^{2*} \mid z^*, \tilde{\Gamma}^*, \mathbf{Y}) \times p(\gamma_\alpha^{2*} \mid z^{\alpha*}, \tilde{\Gamma}^*, \sigma^{2*}, \mathbf{Y}) \times p(\gamma_\beta^{2*} \mid z^{\beta*}, \tilde{\Gamma}^*, \sigma^{2*}, \mathbf{Y}) \\
&\times p(\xi^{2*} \mid z^*, \tilde{\Gamma}^*, \sigma^{2*}, \gamma_\alpha^{2*}, \gamma_\beta^{2*}, \mathbf{Y}) \times p(q_\alpha^* \mid z^{\alpha*}, \tilde{\Gamma}^*, \sigma^{2*}, \gamma_\alpha^{2*}, \gamma_\beta^{2*}, \xi^{2*}, \mathbf{Y}) \times p(q_\beta^* \mid z^{\beta*}, \tilde{\Gamma}^*, \sigma^{2*}, \gamma_\alpha^{2*}, \gamma_\beta^{2*}, \xi^{2*}, \mathbf{Y})
\end{aligned}
$$

According to Equation A.21, we can compute each component of the posterior distribution based on the specific values of each MCMC sample, with only some adjustments required.

$$
\begin{aligned}
\widehat{p}(z^* \mid \mathbf{Y}) &= \frac{1}{G} \sum_{g=1}^{G} p(z^* \mid \sigma^{2(g)}, \gamma_\alpha^{2(g)}, \gamma_\beta^{2(g)}, \xi^{2(g)}, q_\alpha^{(g)}, q_\beta^{(g)}, \mathbf{Y}) \\
\widehat{p}(\tilde{\Gamma}^* \mid z^*, \mathbf{Y}) &= \frac{1}{G} \sum_{g=1}^{G} p(\tilde{\Gamma}^* \mid z^*, \sigma^{2(g)}, \gamma_\alpha^{2(g)}, \gamma_\beta^{2(g)}, \xi^{2(g)}, \mathbf{Y}) \\
\widehat{p}(\sigma_i^{2*} \mid z^*, \tilde{\Gamma}^*, \mathbf{Y}) &= p(\sigma_i^{2*} \mid z^*, \tilde{\Gamma}^*, \mathbf{Y}) \\
\widehat{p}(\gamma_\alpha^{2*} \mid z^{\alpha*}, \tilde{\Gamma}^*, \sigma^{2*}, \mathbf{Y}) &= p(\gamma_\alpha^{2*} \mid z^{\alpha*}, \tilde{\boldsymbol{\alpha}}_1^*, \mathbf{Y}) \\
\widehat{p}(\gamma_\beta^{2*} \mid z^{\beta*}, \tilde{\Gamma}^*, \sigma^{2*}, \mathbf{Y}) &= p(\gamma_\beta^{2*} \mid z^{\beta*}, \tilde{\boldsymbol{\beta}}_1^*, \mathbf{Y}) \\
\widehat{p}(\xi^{2*} \mid z^*, \tilde{\Gamma}^*, \sigma^{2*}, \gamma_\alpha^{2*}, \gamma_\beta^{2*}, \mathbf{Y}) &= p(\xi^{2*} \mid \alpha_0^*, \boldsymbol{\beta}_0^*, \mathbf{Y}) \\
\widehat{p}(q_\alpha^* \mid z^{\alpha*}, \tilde{\Gamma}^*, \sigma^{2*}, \gamma_\alpha^{2*}, \xi^{2*}, \mathbf{Y}) &= p(q_\alpha^* \mid z^{\alpha*}, \mathbf{Y}) \\
\widehat{p}(q_\beta^* \mid z^{\beta*}, \tilde{\Gamma}^*, \sigma^{2*}, \gamma_\beta^{2*}, \xi^{2*}, \mathbf{Y}) &= p(q_\beta^* \mid z^{\beta*}, \mathbf{Y})
\end{aligned}
$$

The posterior distribution of z^* is

$$
\begin{aligned}
p(z \mid -) = &\prod_{i=1}^{N} (2\pi)^{\frac{-T_i}{2}} \times \prod_{i=1}^{N} \frac{S_0/2^{v_0/2}}{\Gamma(v_0/2)} \times q_\alpha^{s(z^\alpha)} (1-q_\alpha)^{L-s(z^\alpha)} \left(\frac{1}{\gamma_\alpha^2}\right)^{\frac{s(z^\alpha)}{2}} \times q_\beta^{s(z^\beta)} (1-q_\beta)^{KL-s(z^\beta)} \left(\frac{1}{\gamma_\beta^2}\right)^{\frac{s(z)}{2}} \\
&\times |V_1|^{\frac{1}{2}} \left[\prod_{i=1}^{N} \left(\frac{1}{\sigma_i^2}\right)^{\frac{T_i+v_0}{2}+1}\right] \times \exp\left(-\sum_{i=1}^{N} \frac{S_i+S_0}{2\sigma_i^2} - \frac{1}{2}\sum_{i=1}^{N} \left(\left(\sigma_i^{-2}\tilde{\mathcal{W}}_i^\top R_i\right)^\top \widehat{\tilde{\Gamma}}_i\right) + \frac{1}{2}\tilde{\Gamma}_1^\top V_1^{-1} \tilde{\Gamma}_1\right).
\end{aligned}
$$

When no sparsity, the prior of Γ^* is:

$$
p(\Gamma^*) = \prod_{l=1}^{1+K+L+K\times L} \left[\left(\frac{1}{2\pi\xi^{2*}}\right)^{\frac{1}{2}} \exp\left(-\frac{\Gamma_l^*}{2\xi^{2*}}\right)\right]
$$

The posterior is:

- $\Gamma^* \sim \mathcal{N}(\Gamma_1, V_1)$ where $V_1 = \left(\sum_{i=1}^{N} \sigma_i^{-2} \mathcal{W}_i^\top \mathcal{W}_i + A\right)^{-1}$, $A = \text{diag}\left(1/\xi^{2(g)}\right)$ and

$\Gamma_1 = V_1 \left(\sum_{i=1}^{N} \sigma_i^{-2} \mathcal{W}_i^\top \mathbf{R}_i \right)$, and the dimension of $\mathcal{W}_i$ is $1 + L + K + K \times L$. Thus,

$$p(\Gamma^* \mid \sigma^{2(g)}, \xi^{2(g)}) = (2\pi)^{-\frac{1+L+K+K\times L}{2}} |V_1|^{-1/2} \exp\left(-\frac{1}{2} (\Gamma^* - \Gamma_1)^\top V_1^{-1} (\Gamma^* - \Gamma_1) \right)$$

- σ_i^{2*}

$$p(\sigma_i^{2*} \mid \Gamma^*) \sim \mathcal{IG}\left(\frac{v_{0,i} + T_i}{2}, \frac{S_{0,i} + S_i + \left(\Gamma^* - \widehat{\Gamma}_i \right)^\top \mathcal{W}_i^\top \mathcal{W}_i \left(\Gamma^* - \widehat{\Gamma}_i \right)}{2} \right)$$

where $S_i = (\mathbf{R}_i - W_i^\top \widehat{\Gamma}_i)^\top (\mathbf{R}_i - W_i^\top \widehat{\Gamma}_i)$ and $\widehat{\Gamma}_i = \left(\mathcal{W}_i^\top \mathcal{W}_i \right)^{-1} \mathcal{W}_i^\top \mathbf{R}_i$ is the OLS estimator for the i-th asset.

- ξ^{2*}

$$p(\xi^{2*} \mid \Gamma^*) \sim \mathcal{IG}\left(\frac{C^*}{2}, \frac{D^*}{2} \right),$$

where $C^* = C + (1 + L + K + K \times L)$ and $D^* = D + \alpha_0^{*2} + \boldsymbol{\beta}_0^{*\top} \boldsymbol{\beta}_0^* + \boldsymbol{\alpha}_1^{*\top} \boldsymbol{\alpha}_1^* + \boldsymbol{\beta}_1^{*\top} \boldsymbol{\beta}_1^*$.