

# Rethinking Sparsity: Parametric Portfolios and Firm Characteristics\*

Daniele Bianchi<sup>†</sup>

Pedro H. M. Venturi<sup>†</sup>

First draft: July 2024.

This draft: December 17, 2024

## Abstract

We investigate the economic trade-off between variable selection and shrinkage in designing optimal parametric portfolios based on a large set of firm characteristics. Using a flexible Bayesian inference approach, we show that a sparsity-inducing prior reduces model uncertainty but leads to high-turnover, under-diversified portfolios with concentrated exposures to a few characteristics. In contrast, a prior favoring ridge-type shrinkage fosters better diversification across more characteristics, yielding portfolios more aligned with mean-variance efficiency, particularly when accounting for realistic transaction costs. Our findings caution against over-reliance on sparsity-inducing methods when predicting the cross-section of stock returns using firm characteristics.

**Keywords:** Parametric Portfolios, Firm Characteristics, Stock Returns, Asset Pricing, Bayesian Inference, Variable Selection, Shrinkage.

**JEL codes:** G10, G11, G12, C22.

---

\*We are thankful to Martijn Boons, Andrea Carriero, Julio Crego, Victor DeMiguel, Nicholas Hirschey, Dimitris Korobilis, Hedibert Freitas Lopes, Cathy Yi-Hsuan Chen, Markus Pelger, and the seminar participants at Queen Mary University of London, Nova School of Business and Economics, and the Adam Smith Business School, University of Glasgow, as well as participants at the CFE-CMStatistics 2024 conference, for their helpful comments and suggestions.

<sup>†</sup>School of Economics and Finance, Queen Mary University of London. E-mail: [d.bianchi@qmul.ac.uk](mailto:d.bianchi@qmul.ac.uk)  
Web: [whitesphd.com](http://whitesphd.com)

<sup>†</sup>School of Economics and Finance, Queen Mary University of London. E-mail: [p.h.moravisventuri@qmul.ac.uk](mailto:p.h.moravisventuri@qmul.ac.uk)

# 1 Introduction

Predicting the cross-section of stock returns and its implications for portfolio decisions is a core challenge in empirical asset pricing. While firm characteristics play a crucial role in explaining return variation, the question of which to prioritize remains unresolved, particularly in high-dimensional settings. Empirical approaches often focus on either selecting a subset of characteristics with the strongest explanatory power or acknowledging that all characteristics may contribute to some extent, even if their individual impacts are small.

The choice between a sparse model (emphasizing variable selection) and a dense model (emphasizing ridge-type shrinkage) extends beyond issues of interpretability, carrying significant implications for asset pricing. For example, [DeMiguel et al. \(2020\)](#) show that incorporating transaction costs can materially influence the number of characteristics investors should consider when maximizing mean-variance portfolio utility. Differently, [Bryzgalova et al. \(2023\)](#) highlight that pervasive model uncertainty can lead to an incomplete understanding of the stochastic discount factor (SDF) in equity markets if not properly accounted for.

This paper investigates the economic trade-off between variable selection and shrinkage – between sparse and dense models – when designing optimal portfolio policies in high dimensions. Specifically, we leverage the flexibility of a heavy-tailed Bayesian prior to capture the sensitivity of portfolio weights to a large set of firm characteristics within the framework of a parametric portfolio rule, as in [Brandt et al. \(2009\)](#). This approach connects to broader economic theory, as the first-order condition for an investor’s optimal portfolio under unconstrained mean-variance utility is equivalent to the SDF (e.g., [Kozak et al., 2020](#)).

We assume a priori that each characteristic influences investors’ utility with a certain probability  $q$ . If a firm characteristic is relevant, its effect  $\theta_j$  on the portfolio policy is modeled using a Student- $t$  distribution with  $\nu$  degrees of freedom and variance scaled by  $\gamma^2$ . Smaller

values of  $\gamma^2$  impose greater shrinkage on  $\theta_j$ , while the thickness of the prior tails, controlled by  $\nu$ , governs the balance between sparsity and ridge-type shrinkage in determining the sensitivity of portfolio weights to firm characteristics. We explore different tail assumptions, ranging from very heavy tails (e.g.,  $\nu = 4$ ) to those approaching a Normal distribution (e.g.,  $\nu = 100$ ). We also propose an economic rationale to calibrating  $\nu$  based on aggregate transaction costs.

Our empirical analysis examines 131 firm characteristics from [Chen and Zimmermann \(2021\)](#) across an unbalanced panel of 21,418 stocks spanning January 1980 to December 2023. A heavier-tailed prior reduces model uncertainty, concentrating portfolio exposures at the intensive margin – relatively larger allocations to fewer characteristics. In contrast, a thin-tailed prior induces greater shrinkage, distributing exposures at the extensive margin with smaller allocations across a broader set of characteristics. These differences have important implications for mean-variance utility, particularly when accounting for trading frictions.

To demonstrate this, we compare the performance of portfolios derived from various prior specifications against an equal-weight benchmark (e.g., [DeMiguel et al., 2009](#)) and a standard parametric portfolio policy based on size, value, and momentum, as in [Brandt et al. \(2009\)](#). Our findings show that priors prioritizing ridge-type shrinkage over sparsity, when predicting individual stock returns based on firm characteristics, deliver superior out-of-sample economic performance, especially in the presence of transaction costs.

These results are robust to a series of additional empirical tests. Drawing on the intuition of [Avramov et al. \(2023a\)](#), we evaluate portfolio policies derived from different priors under various economic restrictions. First, we construct optimal portfolios using a restricted universe of common stocks (CRSP share codes 10 and 11) listed exclusively on the NYSE, while also imposing ex-post portfolio constraints such as limited leverage and no short sales. Second, we assess whether the profitability of portfolios under different priors is more pronounced during high limits-to-arbitrage market states, including periods of high volatility and tight

financial conditions. The results indicate that, to the extent firm characteristics predict cross-sectional stock returns, the outperformance of our Bayesian parametric portfolios persists when investing in relatively cheap-to-trade stocks and during periods of lower volatility and higher market liquidity. This suggests that portfolio profitability is not driven by difficult-to-arbitrage stocks or high limits-to-arbitrage market conditions.

We further evaluate parametric portfolio policies based on alternative regularization priors, such as the Bayesian lasso (Park and Casella, 2008) and the horseshoe (Carvalho et al., 2009), as well as Bayesian variable selection methods like the mixture of normals proposed by George and McCulloch (1993) and the Normal spike-and-slab of Giannone et al. (2021). This analysis provides insights into the trade-off between shrinkage and variable selection beyond our proposed prior structure.

Overall, the findings reinforce our main conclusion: addressing model uncertainty through stricter variable selection results in extreme, under-diversified portfolio allocations with sub-optimal out-of-sample performance. In contrast, a less restrictive approach that induces greater ridge-type shrinkage produces more diversified, cost-effective portfolios better aligned with mean-variance efficiency. These results underscore the risks of over-relying on variable selection methods. Disregarding evidence highlighting the role of model uncertainty and the importance of shrinkage as a regularization tool risks creating the “illusion” that more interpretable, sparsity-inducing approaches can deliver superior economic outcomes.

It is important to acknowledge both the strengths and limitations of our approach. The prior formulation we adopt encompasses popular dimension reduction methods. For instance, setting  $q = 1$  and a large  $\nu$  is equivalent to a diffuse ridge prior, which can be interpreted as a regression on the principal components of the characteristics, applying less shrinkage to the more significant components (e.g., Marquardt, 1970; Smith and Campbell, 1980; Bańbura

et al., 2015; De Mol et al., 2024).<sup>1</sup> This highlights the generality of our approach, which accommodates scenarios where the cross-section of stock returns is driven by a few common components. Another important advantage is that our Bayesian inferential procedure fully characterizes the uncertainty surrounding the share of non-zero weight sensitivities, allowing this uncertainty to be estimated jointly with the identification of relevant characteristics.

However, a key limitation of our framework is its potential underperformance relative to non-linear methods if the linear parametric model fails to capture the full relationship between stock returns and firm characteristics. Our primary focus, however, is on understanding the trade-off between ridge-type shrinkage and sparsity from a high-dimensional portfolio perspective. In this context, adopting a complex non-linear modeling framework would likely hinder the economic interpretability of the results without delivering substantial benefits for the scope of this paper. That said, our framework can be extended to incorporate non-linear transformations of the characteristics, introducing additional flexibility, as suggested in Kelly et al. (2024).

On a separate note, it is important to emphasize that the definition of sparsity is not invariant to transformations of the characteristics. For instance, a model may be sparse in the rotated space of the characteristics if only a few principal components are relevant for prediction. By contrast, it may appear dense in the “natural” space of the original characteristics, since a small number of common components can combine all of them (e.g., Chernozhukov et al., 2017). In this paper, we focus on the economic trade-off between variable selection and ridge-type shrinkage in the original space of untransformed characteristics. The primary motivation for this choice is to facilitate comparison with the literature on lasso-type variable selection, which typically assumes sparsity in the original characteristics (e.g., Chinco et al., 2019; Freyberger et al., 2020). More importantly, analyzing sparsity patterns

---

<sup>1</sup>Conversely, setting  $q = 1$  and a small  $\nu$  yields a conventional diffuse Student- $t$  shrinkage prior, which lacks variable selection properties (e.g., Carvalho et al., 2009; Armagan and Zaretzki, 2010).

in the original space of characteristic-managed portfolios is more coherent with the optimal parametric portfolio choice framework proposed in [Brandt et al. \(2009\)](#).

Our work contributes to a large literature exploring the cross-sectional variation of stock returns based on firm characteristics, including [Hou et al. \(2015\)](#); [Harvey et al. \(2016\)](#); [Green et al. \(2017b\)](#); [Kelly et al. \(2019\)](#); [Freyberger et al. \(2020\)](#); [Haddad et al. \(2020\)](#); [Kozak et al. \(2020\)](#); [Chen and Zimmermann \(2021\)](#); [Bryzgalova et al. \(2023\)](#), among others. Our analysis is closely related to [DeMiguel et al. \(2020\)](#) and [Bryzgalova et al. \(2023\)](#).

Unlike [DeMiguel et al. \(2020\)](#), we leverage the flexibility of a heavy-tailed Bayesian prior, allowing for an explicit investigation of the trade-off between variable selection and shrinkage in shaping optimal portfolio compositions. In contrast to [Bryzgalova et al. \(2023\)](#), we focus on testing a high-dimensional parametric portfolio policy in an out-of-sample setting while explicitly accounting for transaction costs.

A second strand of literature to which we contribute involves the use of Bayesian methods in empirical asset pricing. Bayesian tools have been extensively applied in various domains, including asset allocation (e.g., [Pettenuzzo et al., 2014](#)), model selection (e.g., [Pástor and Stambaugh, 2000](#); [Chib et al., 2020](#); [Avramov et al., 2023b](#)), performance evaluation (e.g., [Busse and Irvine, 2006](#); [Harvey and Liu, 2019](#)), return predictability (e.g., [Avramov, 2002](#)), and asset pricing tests (e.g., [Jensen et al., 2022](#)), among others.

Our empirical framework leverages the flexibility of a heavy-tailed spike-and-slab specification, as in [Fava and Lopes \(2021\)](#), which extends the normal spike-and-slab approach of [Giannone et al. \(2021\)](#). Unlike these studies, we focus on examining the out-of-sample economic trade-off between sparsity and shrinkage in constructing parametric portfolios based on a multitude of firm characteristics.

## 2 Parametric portfolio policy

Consider  $N_t$  stocks available at a given time  $t$ . Each stock  $i$  has an excess return  $r_{i,t+1}$  over the period  $[t, t + 1]$  and a  $k$ -dimensional vector of stock characteristics  $\widehat{\mathbf{x}}_{i,t} = (\widehat{x}_{i,t}^1, \dots, \widehat{x}_{i,t}^k)$  observed at time  $t$ . The investor’s objective is to choose the optimal portfolio weights  $\mathbf{w}_t = (w_{1,t}, \dots, w_{N_t,t})^\top$  to maximize the utility of the portfolio return  $r_{p,t+1} = \mathbf{w}_t^\top r_{t+1}$ .

Following the approach in [Brandt et al. \(2009\)](#), we define the optimal portfolio choice as a linear function of firm characteristics:

$$\mathbf{w}_t = \mathbf{w}_t^b + \frac{1}{N_t} \widehat{\mathbf{X}}_t \boldsymbol{\theta}, \quad (1)$$

where  $\mathbf{w}_t^b$  represents the benchmark portfolio allocation,  $\widehat{\mathbf{X}}_t$  is the  $N_t \times k$  matrix of standardized characteristics, and  $\boldsymbol{\theta}^\top = (\theta_1, \dots, \theta_k)^\top$  is the vector of “sensitivities” that tilts the portfolio toward characteristics that enhance investor utility. The matrix  $\widehat{\mathbf{X}}_t$  is standardized cross-sectionally to have zero mean and unit variance across all stocks, ensuring that deviations from the benchmark portfolio sum to zero (e.g., [DeMiguel et al., 2020](#)). This implies that the portfolio weights sum to one as long as the benchmark weights do. The normalization by  $1/N_t$  ensures the portfolio rule can accommodate an arbitrary number of stocks.<sup>2</sup>

The coefficients  $\theta_1, \dots, \theta_k$  are constant across assets, meaning that the optimal portfolio depends solely on the characteristics rather than on the stocks themselves. The return on the parametric portfolio at time  $t + 1$  can therefore be expressed as:

$$r_{p,t+1} = \mathbf{w}_t^\top r_{t+1} = r_{t+1}^b + \boldsymbol{\theta}^\top f_{t+1}, \quad (2)$$

---

<sup>2</sup>Doubling the number of stocks without changing the cross-sectional distribution of the characteristics would result in twice as aggressive allocations, even though the underlying investment opportunities remain unchanged.

where  $f_{t+1} = \frac{1}{N_t} \widehat{\mathbf{X}}_t^\top r_{t+1}$  represents the returns on the characteristic-managed portfolios, and  $r_{t+1}^b = \mathbf{w}_t^{b,\top} r_{t+1}$  is the return on the benchmark portfolio.

Equation (2) effectively rotates the investment universe from individual stocks to the space of characteristics. Thus, the complex problem of investing in a large number of stocks reduces to estimating the contribution of each characteristic-managed portfolio to investor utility beyond the benchmark allocation (e.g., [Kelly and Xiu, 2023](#)). We use an equally weighted portfolio as the benchmark ( $w_{i,t}^b = 1/N_t \forall i, t$ ), motivated by empirical evidence suggesting its long-term outperformance over other benchmarks, such as the value-weighted portfolio (see, e.g., [DeMiguel et al., 2009](#)).

We assume that the investor optimizes mean-variance utility, which allows for determining the vector  $\theta_1, \dots, \theta_k$  through a linear projection of characteristic-managed portfolios onto the benchmark allocation return.<sup>3</sup> Specifically, the investor maximizes the expected utility:

$$\max_{\boldsymbol{\theta}} \frac{1}{T} \sum_{t=0}^{T-1} u(r_{p,t+1}) = \frac{1}{T} \sum_{t=0}^{T-1} \left[ r_{t+1}^b + \boldsymbol{\theta}^\top f_{t+1} - \frac{\zeta}{2} (r_{t+1}^b + \boldsymbol{\theta}^\top f_{t+1})^2 \right], \quad (3)$$

where  $\zeta$  denotes risk aversion. The first-order condition for maximizing utility leads to:

$$\widehat{\boldsymbol{\theta}} = \frac{1}{\zeta} \left( \frac{1}{T} \sum_{t=0}^{T-1} f_{t+1}^\top f_{t+1} \right)^{-1} \frac{1}{T} \sum_{t=0}^{T-1} f_{t+1} (1 - \zeta r_{t+1}^b) = (F^\top F)^{-1} F^\top Y, \quad (4)$$

where  $F$  is the  $T \times k$  matrix of characteristic-managed portfolio returns scaled by risk aversion, and  $Y = \mathbf{1} - \zeta R^b$ , with  $\mathbf{1}$  and  $R^b$  representing  $T \times 1$  vectors of ones and benchmark returns, respectively. As a result,  $\widehat{\boldsymbol{\theta}}$  is equivalent to a least-squares projection of  $Y$  onto  $F$ , allowing the optimal portfolio to implicitly account for expected returns, variances, and covariances on characteristic-managed portfolios as they affect investor utility (e.g., [Britten-Jones, 1999](#)).

---

<sup>3</sup>[Kandel and Stambaugh \(1996\)](#) show that if only the first two conditional moments matter for portfolio choice, the optimal investment rule under a power utility is proportional to a mean-variance utility rule.

## 2.1 A spike-and-slab prior with heavy tails

The existing literature suggests that the number of characteristics containing potentially useful information on stock returns is large (see [Chen and Zimmermann, 2021](#), and references therein). Several regularization and dimension reduction techniques have been proposed, including ridge regression (e.g., [Kozak et al., 2020](#)), principal components analysis (e.g., [Haddad et al., 2020](#)), and the lasso (e.g., [Freyberger et al., 2020](#)), to address the challenges of overfitting in such high-dimensional settings.

We adopt the mixture prior proposed by [Fava and Lopes \(2021\)](#) (henceforth GLP-t), which extends the framework of [Giannone et al. \(2021\)](#) by relaxing the assumption of normality for the non-negligible entries in  $\theta_j, j = 1, \dots, k$  (the slab) while retaining a Dirac mass at zero for negligible entries (the spike). Specifically, the prior on  $\theta_j, j = 1, \dots, k$  is defined as:

$$\theta_j \mid \sigma^2, \gamma^2, q \sim \begin{cases} \mathcal{T}_\nu(0, \sigma^2 \gamma^2) & \text{with probability } q, \\ 0 & \text{with probability } 1 - q. \end{cases} \quad (5)$$

Here,  $\nu$  represents the degrees of freedom of the Student- $t$  distribution,  $q$  is the probability of including a given characteristic, and  $\gamma^2$  controls the degree of shrinkage applied to the weight tilts  $\theta_j$ . This setup ensures that each  $\theta_j$  is either zero with probability  $1 - q$  or drawn from a Student- $t$  distribution with variance  $\mathbb{V}(\theta_j) = \frac{\nu}{\nu-2} \sigma^2 \gamma^2$  with probability  $q$ .

The remaining priors for  $q$ ,  $\gamma^2$ , and  $\sigma^2$  follow [Giannone et al. \(2021\)](#). Specifically, we assume  $p(\sigma^2) \propto 1/\sigma^2$  and specify the marginal prior for  $q \sim \text{Beta}(a, b)$ , representing the proportion of characteristics an investor considers. The prior distribution for the shrinkage parameter  $\gamma^2$  is implied by the function:

$$\gamma^2 = \frac{1}{k\bar{v}_x q} \cdot \frac{R^2}{1 - R^2} \quad (6)$$

where  $k$  is the number of characteristics,  $\bar{v}_x = \mathbb{E}[\hat{\sigma}_j^2]$ , and  $\hat{\sigma}_j^2$  represents the sample variance of the  $j$ -th managed portfolio return. The prior for  $R^2 \sim \text{Beta}(A, B)$  represents the expected sample variance of the benchmark returns explained by the managed portfolios and weight sensitivities relative to the error.

We note that setting a direct prior on  $R^2$  allows us to apply this framework to models of varying sizes while remaining agnostic to the exact nature of the dimensionality issue. Figure 1 illustrates the implications of different parameter choices. Notably, varying  $q$  has only a modest effect on prior shrinkage ( $\gamma^2$ ), whereas assuming a high  $R^2$  can significantly reduce shrinkage, especially when sparsity is high (i.e.,  $q$  is small). This highlights that if only a few firm characteristics explain much of the stock variation, then less regularization is necessary (i.e.,  $\gamma^2$  is large). We use an uninformative prior setup with  $a = b = 1$  and  $A = B = 1$ , corresponding to a uniform distribution with  $E[R^2] = E[q] = 0.5$ .

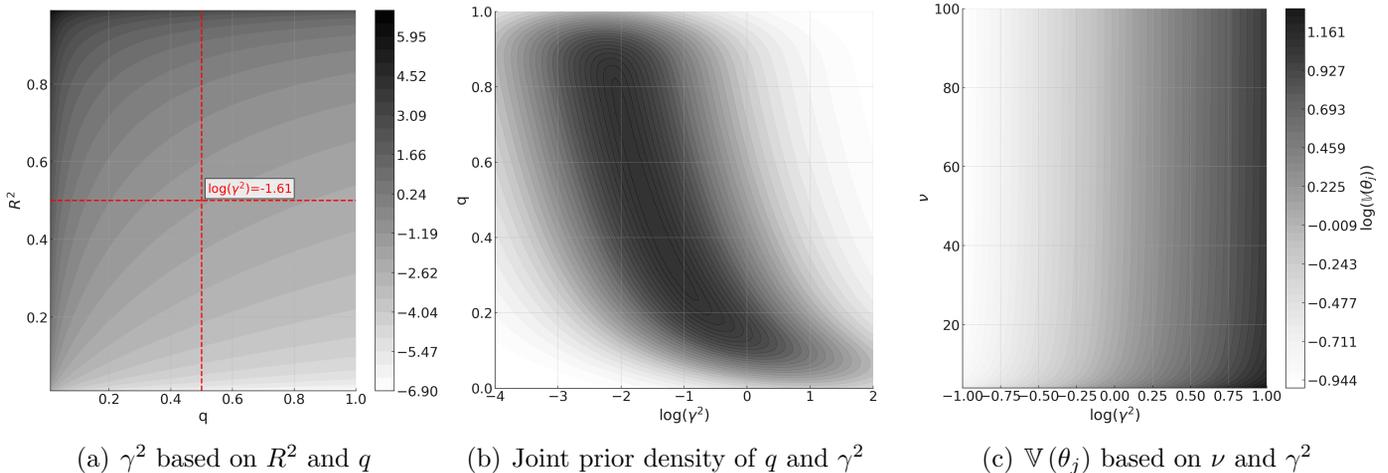


Figure 1: **Prior comparative statics.** The left panel shows the contours of  $\log(\gamma^2)$  as a function of  $R^2$  and  $q$  (see Eq.(6)). The middle panel shows the joint prior density of  $q$  and  $\gamma^2$ . The right panel shows the contours of the log of prior variance  $\log(\mathbb{V}(\theta_j))$  based on  $\nu$  and  $\gamma^2$ .

Equation (6) implies a negative correlation between  $q$  and  $\gamma^2$ , reflecting the common belief that sparsity and shrinkage act as substitutes when addressing the curse of dimensionality

(e.g., [Abadie and Kasy, 2019](#)). The middle panel of [Figure 1](#) illustrates this relationship. To facilitate interpretation, we represent the joint density of  $q$  and  $\log(\gamma^2)$  instead of  $q$  and  $\gamma^2$ . A flat prior on  $R^2$  and  $q$  results in a pronounced negative correlation between  $q$  and  $\gamma^2$ : the lower (higher) the probability of including a predictor and the overall model size, the higher (lower) the prior variance  $\mathbb{V}(\theta_j)$ .

The right panel, however, shows that the prior variance is also influenced by  $\nu$ . Specifically, the figure demonstrates that a heavier-tailed Student- $t$  prior results in a larger  $\log(\mathbb{V}(\theta_j))$  for a given level of shrinkage  $\log(\gamma^2)$ .<sup>4</sup> This implies an even stronger negative correlation between sparsity and shrinkage for smaller values of  $\nu$ .

Overall, [Figure 1](#) suggests that ignoring shrinkage may lead to overly sparse weight sensitivities, potentially biasing inferences about the optimal portfolio policy. The interplay between  $q$  and  $\gamma^2$  becomes more pronounced as the prior tails, governed by  $\nu$ , become thicker. We will revisit this point in greater detail when we discuss the parameter posterior estimates.

## 2.2 An economic rationale to calibrate $\nu$

[Figure 1](#) illustrates how different values of  $\nu$  imply different priors on  $\theta_j$ . Specifically, larger values of  $\nu$  correspond to smaller prior variances, all else being equal. This has important implications for optimal the mean-variance portfolio allocation. To understand this, recall from [Eq.\(1\)](#) that the optimal portfolio weight for the  $i$ -th stock is given by

$$w_{i,t} = w_{i,t}^b + (\hat{x}_{i,t}^1 \theta_1, \dots, \hat{x}_{i,t}^k \theta_k).$$

Thus, greater shrinkage of  $\theta_j$  towards zero reduces the sensitivity of the  $i$ -th portfolio weight to characteristic  $j$ , which in turn affects the allocation strategy.

---

<sup>4</sup>We report the contours of  $\log(\mathbb{V}(\theta_j))$  instead of  $\mathbb{V}(\theta_j)$  to make the interpretation more straightforward.

In the main empirical analysis we explore different values of  $\nu$  ranging from 4 to 100, covering both heavy-tailed (Student- $t$ ) and thin-tailed (Normal) distributions. Additionally, we propose an economic rationale for calibrating  $\nu$  based on a proxy for aggregate transaction costs. This approach is motivated by the idea that a smaller  $\nu$  potentially imply larger liquidity needs and rebalancing costs, contingent on the elasticity estimate  $\hat{\theta}_j$ .

To capture this intuition, we define a measure of aggregate transaction costs:

$$TC_t = \frac{1}{N_t} \left| \sum_{i=1}^{N_t} \sum_{j=1}^k \hat{x}_{i,t}^j \frac{\eta_{i,t}}{2} \right|, \quad t = 1, \dots, T \quad (7)$$

where  $\eta_{i,t}$  represents the bid-ask spread for asset  $i$  at time  $t$ , serving as a proxy for individual trading costs (e.g., Bessembinder and Venkataraman, 2010). The latter is computed using the Corwin and Schultz (2012) approximation, scaled by the stock price. The absolute value ensures that aggregate transaction costs are positive. In addition, Eq.(7) accounts for the possibility that rebalancing different characteristics can reduce overall transaction costs. For instance, characteristics with positive and negative values can offset one another, thereby reducing the magnitude of the term  $\sum_{j=1}^k \hat{x}_{i,t}^j \frac{\eta_{i,t}}{2}$  (e.g., DeMiguel et al., 2020).

Figure C.2 in Appendix C reports the cross-sectional distribution of the half bid-ask spread for the entire out-of-sample period (left panel) and averages it across groups of stocks sorted by market capitalization (right panel). The figures confirm that using the effective bid-ask spread as a proxy for transaction costs aligns with the assumptions in Brandt et al. (2009), who model transaction costs as decreasing with firm size and over time.

We note that the value of  $TC_t \ll 1, \forall t$ . Thus, a straightforward plug-in calibration is unfeasible, as none of the moments of the Student- $t$  distribution are defined for  $\nu = TC_t < 1$ . To address this issue, we rescaled  $TC_t$  to fall within the interval  $[4, 100]$  (TC1) or multiplied

$TC_t$  by 1,000 (TC2).<sup>5</sup>

## 2.3 Posterior inference

The Student- $t$  component in the spike-and-slab prior from Eq.(5) can be described as a scale mixture of normals of the form:

$$\theta_j \mid \sigma^2, \gamma^2, q, \lambda_j^2 \sim \mathcal{N}(0, \sigma^2 \gamma^2 \lambda_j^2) \quad (8)$$

where  $\lambda_j^2 \sim \mathcal{IG}(\nu/2, \nu/2)$  follows an inverse-Gamma distribution with both scale and shape parameters equal to  $\nu/2$ .

This formulation is particularly useful because it allows us to express the posterior distribution of  $\theta_j$  in closed form while preserving the heavy-tail structure of the prior. Specifically, when integrated over  $\lambda_j^2$ , the marginal distribution of  $\theta_j$  is a Student- $t$  distribution with  $\nu$  degrees of freedom (see [Andrews and Mallows, 1974](#), and Appendix A for a formal proof). As a result, the joint posterior distribution of  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  is a multivariate normal distribution of the form:

$$\theta_1, \dots, \theta_k \mid \text{rest} \sim \mathcal{N} \left( \underbrace{\Sigma_\theta^{-1} F^\top Y}_{\mu_\theta}, \sigma^2 \underbrace{(F^\top F + D^{-1})^{-1}}_{\Sigma_\theta^{-1}} \right) \quad (9)$$

where  $D = \gamma^2 \text{diag}(z_1 \lambda_1^2, \dots, z_k \lambda_k^2)$ , and  $z_j \sim \text{Bernoulli}(q)$  acts as an indicator that selects the  $j = 1, \dots, k$  characteristics with probability  $q$ . Appendix B.1 provides the complete derivation. Note that the posterior in Eq.(9) can be further simplified by integrating out  $\lambda_j^2$ .

---

<sup>5</sup>We ensure that  $\nu$  lies between 4 and 100 by applying the following linear transformation to  $TC_t$ :

$$TC_t^{\text{scaled}} = 4 + (100 - 4) \frac{TC_t - \min(TC_t)}{\max(TC_t) - \min(TC_t)}.$$

The posterior distribution of  $\lambda_j^2, j = 1, \dots, k$  is given by:

$$\lambda_j^2 \mid \text{rest} \sim \mathcal{IG} \left( \frac{\nu + 1}{2}, \frac{\theta_j^2}{2\sigma^2\gamma^2} + \frac{\nu}{2} \right) \quad (10)$$

where  $\frac{\nu+1}{2}$  and  $\frac{\theta_j^2}{2\sigma^2\gamma^2} + \frac{\nu}{2}$  are the shape and scale parameters, respectively (see Appendix B.2 for a detailed derivation). For the remaining posteriors of  $\sigma^2$ ,  $R^2$ , and  $q$ , please refer to Appendices B.3 and B.5. Following Giannone et al. (2021), we sample posterior draws by discretizing the support of  $R^2, q \in [0, 1]$ , using an interlacing approach with two grids defined over the unit interval, and then evaluating the joint posterior distribution.

We note that it is also possible to estimate  $\nu$  directly from the data. Appendix B.6 outlines a strategy for estimating  $\nu$  using a Gamma prior distribution,  $\nu \sim \mathcal{G}(\alpha, \beta)$ . Since the posterior distribution is not available in closed form, a Metropolis-Hastings (MH) algorithm would be required for estimation.<sup>6</sup> While estimating  $\nu$  allows for greater flexibility, it also introduces additional complexity that is beyond the scope of this paper. In this respect, calibrating  $\nu$  like we do may introduce some rigidity, but it better aligns with our objective of transparently investigating the trade-off between sparsity and shrinkage to leverage the information contained in firm characteristics.

### 3 Data and full-sample estimates

We collect firm characteristics from the [www.openassetpricing.com](http://www.openassetpricing.com) website (see Chen and Zimmermann, 2021, for more details).<sup>7</sup> The initial dataset includes 212 U.S. firm characteristics, each signed such that a larger value implies a higher expected return.<sup>8</sup> To ensure

<sup>6</sup>For an overview of the random walk MH algorithm, see, for example, Gelman et al. (2013).

<sup>7</sup>We use version 1.4.1, released in October 2024.

<sup>8</sup>Price, Size, and STreversal are downloaded from the Center for Research in Security Prices (CRSP) and matched to the initial set of firm characteristics.

consistency, we retained the original sign of all characteristics based on the documentation provided by the authors.

We restrict our sample to continuous characteristics available from January 1980 to December 2023, resulting in a total of 154 characteristics. These characteristics are merged with monthly stock return data from the Center for Research in Security Prices (CRSP). Our sample includes all common stocks (share codes 10 and 11) listed on the NYSE, AMEX, or NASDAQ exchanges. Following [Green et al. \(2017a\)](#) and [Chen and Zimmermann \(2021\)](#), we incorporate delisting returns as per [Shumway and Warther \(1999\)](#).<sup>9</sup> We remove observations with missing returns and exclude extreme returns exceeding 250% or falling below -100%. Additionally, we filter out stocks with market capitalizations below the 20th percentile, thereby excluding very small and illiquid stocks.

We further refine the sample by excluding characteristics with missing data for more than 60% of stocks across the sample or for more than 90% of stocks during any single period if this occurs multiple times. Remaining missing values are imputed monthly using the cross-sectional median for each characteristic (e.g., [Gu et al., 2020](#)). Each characteristic is winsorized cross-sectionally at the 1st and 99th percentiles, following [Green et al. \(2017a\)](#), and subsequently normalized to have a cross-sectional mean of zero and a standard deviation of one, as recommended by [Brandt et al. \(2009\)](#) and [DeMiguel et al. \(2020\)](#).

The final dataset comprises 131 characteristics, spanning an unbalanced panel of 21,418 stocks, with a minimum of 2,837, a maximum of 5,921, and an average of 4,061 stocks per month from January 1980 to December 2023. [Figure C.1](#) in [Appendix C](#) presents the sample mean, volatility, and kurtosis of the managed portfolios used in the main empirical study. The descriptive statistics reveal a strong alignment between return volatility and kurtosis, indicating that extreme returns significantly contribute to the dispersion of returns. This

---

<sup>9</sup>Returns adjusted for delisting account for 1,537 out of 2,144,343 observations in our final dataset.

provides prima facie evidence that a prior with heavy tails may be justified by the data. Furthermore, the lack of correlation between kurtosis and mean returns suggests that high kurtosis reflects heightened risk rather than superior performance.

### 3.1 Understanding the trade-off between sparsity and shrinkage

We analyze the relationship between sparsity and shrinkage using posterior estimates based on the full sample of returns. Specifically, we examine how the prior thickness ( $\nu$ ) affects  $q$  and  $\gamma^2$  and thus the portfolio policy  $w_{i,t} = w_{i,t}^b + \sum_{j=1}^k \theta_j \hat{x}_{i,t}^j$  via portfolio tilts  $\theta_j, j = 1, \dots, k$ .

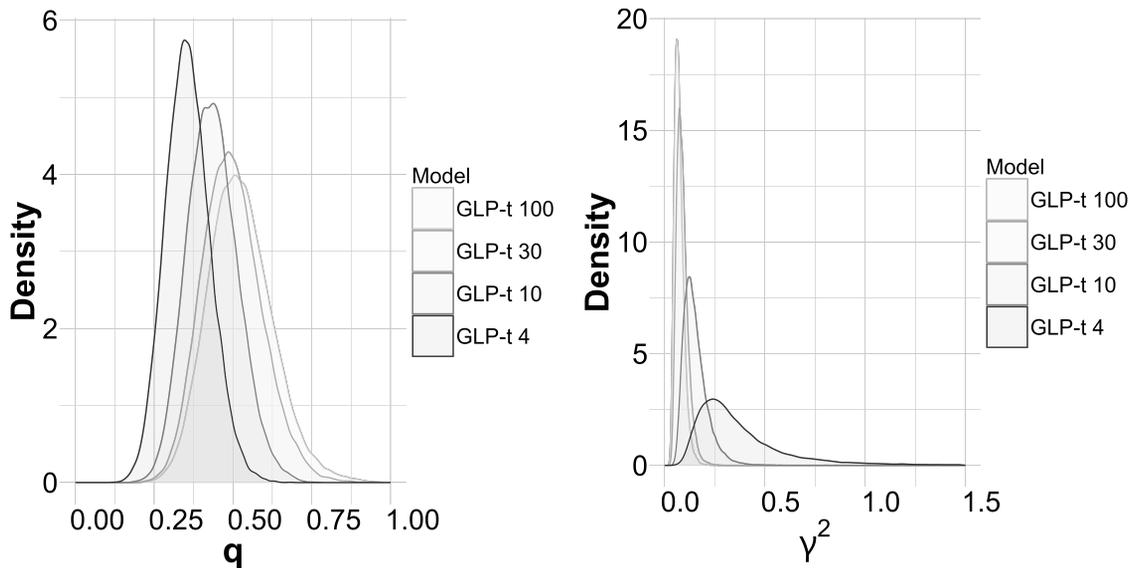


Figure 2: **Posterior estimates of  $q$  and  $\gamma^2$ .** The figure shows the in-sample posterior estimates of  $q$  (left panel) and  $\gamma^2$  (right panel) for the prior with  $\nu = [4, 10, 30, 100]$ . The sample period is from January 1980 to December 2023.

Figure 2 presents the posterior estimates for different degrees of freedom ( $\nu = [4, 10, 30, 100]$ ) of the Student- $t$  prior. Thick-tailed priors ( $\nu$  low) favor simpler models by selecting fewer characteristics (smaller  $q$ ) and applying less stringent shrinkage (larger  $\gamma^2$ ) to portfolio tilts ( $\theta_j$ ). This approach reduces uncertainty about which characteristics maximise investor’s utility while allowing their influence on portfolio weights to remain relatively unconstrained.

Conversely, thin-tailed priors ( $\nu$  high) include a broader set of characteristics (larger  $q$ ) but enforce stronger shrinkage (smaller  $\gamma^2$ ), thus imposing greater regularization of portfolio tilts.

As shown in Figure D.3 in Appendix D.1, a larger  $\nu$  (e.g.,  $\nu = 100$ ) results in pervasive model uncertainty, meaning no distinct sparsity pattern emerges among firm characteristics. In contrast, lower  $\nu$  values reduce uncertainty by enabling the confident exclusion of characteristics with weaker signals. Notably, a prior robust to outlying signals (e.g.,  $\nu = 4$ ) does not alter the selection of strong characteristics but significantly decreases the inclusion probability of those weakly associated with future stock returns.<sup>10</sup>

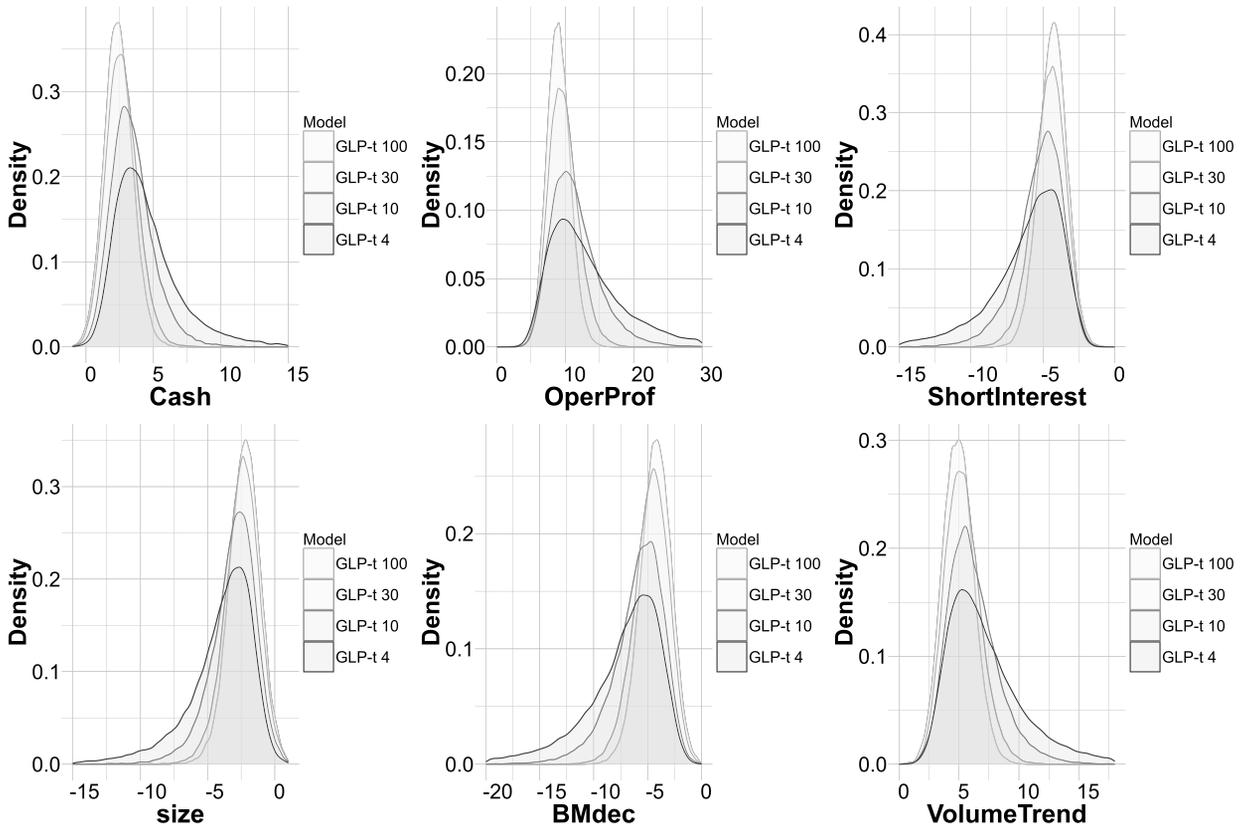


Figure 3: **Posterior estimates of  $\theta_j$ .** The figure shows the posterior estimates for some firm characteristics selected by all prior specifications. The sample period is from January 1980 to December 2023.

<sup>10</sup>This aligns with the intuition that heavier-tailed priors are better suited to identifying relevant characteristics in the presence of extreme signals (e.g., [Carvalho et al., 2009](#)). Table D.1 lists the variables selected for different values of  $\nu$  based on a conventional 50% posterior inclusion probability cutoff.

Figure 3 illustrates the effect of prior tails on the posterior distribution of portfolio tilts. The figure presents the posterior estimates of  $\theta_j$  for a few selected characteristics, including cash holdings, short interest, operating profitability, and size, across all prior specifications. Heavier-tailed priors produce wider posterior distributions; as  $\nu$  decreases, the dispersion of  $\theta_j$  increases, indicating greater uncertainty in the estimated sensitivities.

In contrast, larger values of  $\nu$  result in stronger shrinkage on  $\theta_j$ , effectively dampening the influence of firm characteristics on the portfolio policy. Consequently, the portfolio weights remain closer to the benchmark  $w_{i,t}^b$ , leading to a more diversified portfolio with limited tilts. This point will be explored in more details in Section 4.2, where we examine the out-of-sample portfolio composition and performance under different prior assumptions.

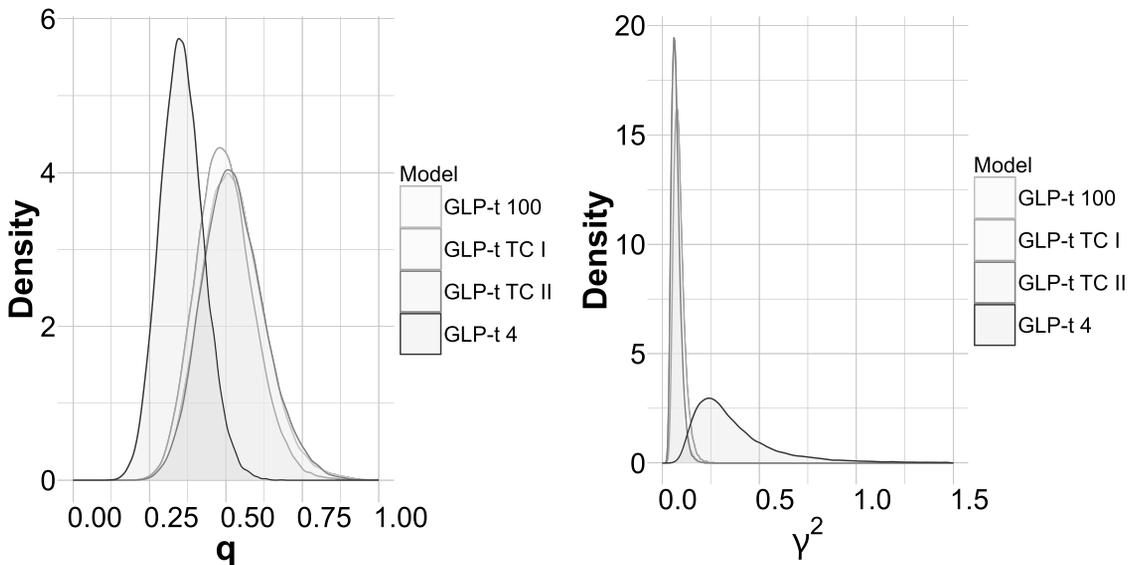


Figure 4: **Posterior estimates when  $\nu$  is calibrated based on transaction costs.** The figure shows the in-sample posterior estimates of  $q$  (left panel) and  $\gamma^2$  (right panel) for the prior with  $\nu = [4, TC1, TC2, 100]$ . The sample period is from January 1980 to December 2023.

Figure 4 presents the posterior estimates of  $q$  and  $\gamma^2$  when  $\nu$  is calibrated based on transaction costs. For the full-sample implementation,  $\nu$  is computed as  $\nu = \frac{1}{T} \sum_{t=1}^T TC_t$ , where  $TC_t$  is scaled as discussed in Section 2.2. The results indicate that incorporating transaction cost

information into the prior calibration tends to favor a thin-tailed prior specification, leading to less sparsity and greater shrinkage in the set of portfolio tilts. This finding aligns with the intuition of [DeMiguel et al. \(2020\)](#), suggesting that transaction costs can lead to less sparsity in the set of characteristics used to maximize a mean-variance portfolio policy.

## 4 Out-of-sample analysis

In-sample estimates of  $\theta_j$ ,  $q$ , and  $\gamma^2$  provide valuable insights into the properties of the prior and its implications for optimal portfolio choice in a static setting. We now discuss the recursive, 240-month training-window estimates of the sparsity  $q$  and the shrinkage  $\gamma^2$  parameters as a function of  $\nu$ . These recursive estimates form the foundation for the real-time implementation of the parametric portfolio policy we will discuss next.

### 4.1 Recursive estimates and real-time portfolios

Figure 5 presents the recursive estimates of  $\sqrt{\gamma^2}$  (left panel) and  $q$  (right panel) for various degrees of freedom  $\nu$ .<sup>11</sup> The out-of-sample period is from January 2000 to December 2023. The posterior mean estimates exhibit considerable variation over time. For instance, the estimate of  $q$  for  $\nu = 100$  peaks at 0.6 following the dot-com bubble and gradually declines to 0.4 leading up to the COVID-19 outbreak. This results in portfolio choices based on a larger set of characteristics while enforcing stronger shrinkage on portfolio tilts.

In contrast, the prior with  $\nu = 4$  maintains a more persistent focus on selecting a smaller, more predictive subset of firm characteristics, with the posterior mean of  $q$  ranging from 0.22 to below 0.1 over the out-of-sample period. This results in portfolios that tilt strongly toward a narrower set of firm characteristics with higher predictive power. Furthermore,

---

<sup>11</sup>We report the  $\sqrt{\gamma^2}$  to increase readability due to the scale of  $\gamma^2$ .

smaller values of  $\nu$  correspond to higher  $\gamma^2$ , assigning greater impact to the most predictive characteristics in the optimal portfolio.

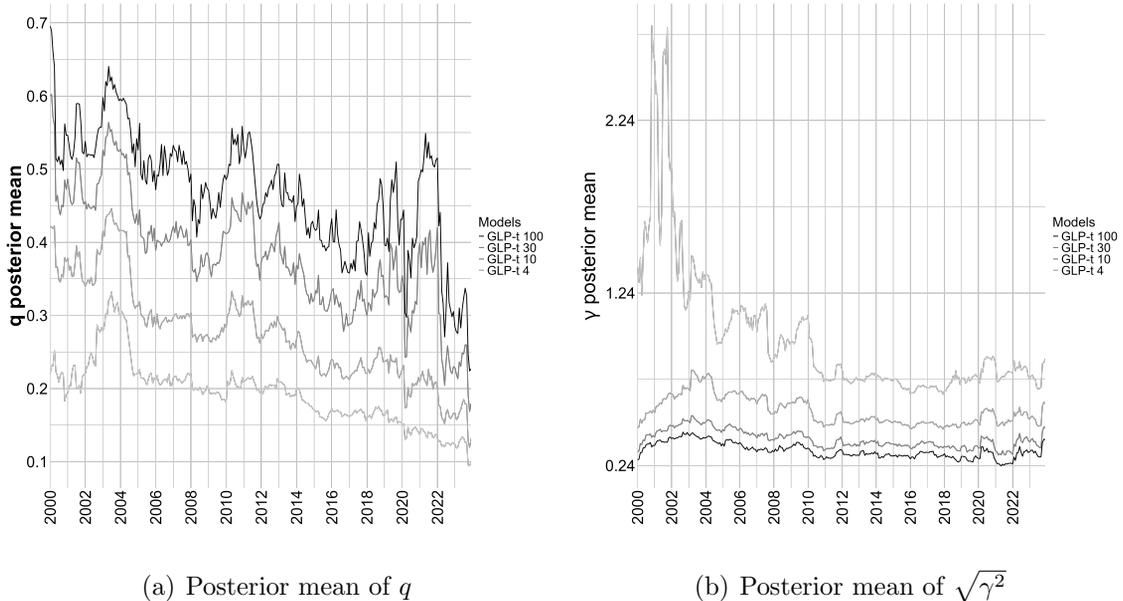


Figure 5: **Recursive estimates of  $q$  and  $\gamma^2$ .** The figure shows the posterior mean of  $q$  (left panel) and  $\sqrt{\gamma^2}$  (right panel) for different levels of  $\nu = [4, 10, 30, 100]$ . The estimates are based on a rolling window of 240 months. The out-of-sample period is from January 2000 to December 2023.

Figure D.4 in Appendix D.2 illustrates the temporal patterns of  $\sqrt{\gamma^2}$  and  $q$  posterior estimates when  $\nu$  is calibrated at each period based on the most recent trading cost estimate as in Eq.(7). The trajectories of the posterior estimates resemble those of thin-tailed priors, albeit with slightly more erratic fluctuations, reflecting the volatile nature of  $TC_t$  over the out-of-sample period.

Overall, consistent with the full-sample estimates, there is a negative relationship between sparsity and shrinkage. This inverse relationship is less pronounced when transaction costs are used to calibrate  $\nu$ , resulting in persistently smaller (larger) values of  $\gamma^2$  ( $q$ ) over time. These findings underscore the importance of tailoring prior tails based on the investor's objectives.

## 4.2 Portfolio composition and characteristics exposures

The recursive posterior estimates reveal that the extent to which prior tails account for extreme returns significantly influences both the number of firm characteristics included in the portfolio policy and the degree of shrinkage applied to each weight sensitivity,  $\theta_j, j = 1, \dots, k$ . To investigate the practical investment implications, we calculate the portfolio's Herfindahl-Hirschman Index (HHI) as  $\text{HHI}_t = \sum_{i=1}^{N_t} w_{it}^2$ , where  $w_{it}$  represents the portfolio weight of asset  $i$ , and  $N_t$  denotes the total number of assets at time  $t$ . Examining the HHI over time allows us to analyze how different prior specifications (e.g., heavy-tailed versus thin-tailed priors) influence portfolio composition and diversification.

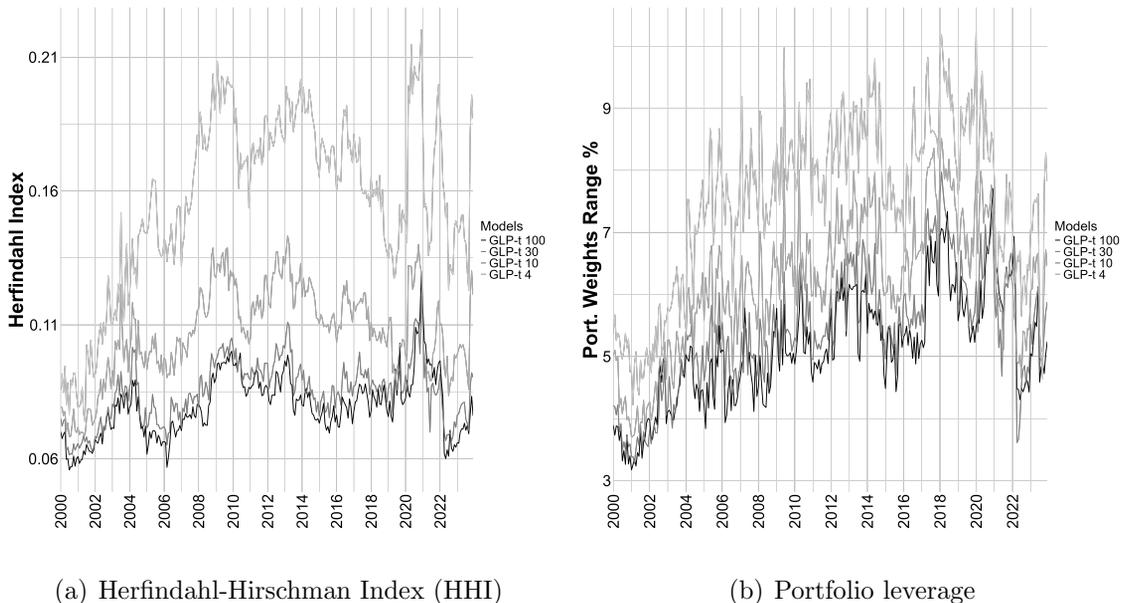


Figure 6: **Portfolio diversification and leverage.** The figure shows the Herfindahl-Hirschman Index (HHI) (left panel) and the weights range (right panel) obtained for the recursive parametric portfolio allocation for different prior specifications. The optimal allocation is based on a rolling window of 240 months. The sample period is from January 2000 to December 2023.

Figure 6 displays the HHI index over the out-of-sample period. To enhance readability, the HHI is rescaled from 0 (a perfectly equal-weighted portfolio) to 1 (a fully concentrated

portfolio where a single asset holds all the weight). Panel (a) demonstrates that heavy-tailed priors lead to higher concentration, with fewer stocks capturing larger capital allocations (higher HHI). For instance, the portfolio HHI for a prior with  $\nu = 4$  consistently exceeds that for  $\nu = 100$  by more than a factor of two throughout the out-of-sample period. Notably, diversification levels for  $\nu = 4$  and  $\nu = 10$  are relatively similar.

Periods of market stress, such as the Great Financial Crisis and the COVID-19 pandemic, exacerbate portfolio concentration across all priors, as indicated by an increase in the HHI index. This finding highlights the sensitivity of portfolio diversification to external shocks, regardless of the prior tails assumption.

We also examine the leverage implied by each prior by calculating the spread between the maximum and minimum portfolio weights,  $|\max(w_{i,t}) - \min(w_{i,t})|$ , at each time  $t$ . A larger spread between long and short positions indicates higher liquidity requirements to implement the portfolio allocation (e.g., [Patton and Weller, 2020](#)). Panel (b) of [Figure 6](#) shows that heavier-tailed priors result in more extreme portfolio weights, with the spread between  $\max(w_{i,t})$  and  $\min(w_{i,t})$  reaching as high as 10% during the COVID-19 pandemic. This spread is more than 50% higher than that observed with thin-tailed priors.

[Figure D.5](#) in [Appendix D.2](#) demonstrates that, consistent with the recursive estimates of  $q$  and  $\gamma^2$  (see [Figure D.4](#)), calibrating  $\nu$  based on transaction costs encourages portfolio diversification and leverage patterns that, while more erratic, remain broadly comparable to those observed with a thin-tailed prior specification. This effect is particularly pronounced for the TC2 calibration. Overall, calibrating  $\nu$  using aggregate transaction costs results in lower portfolio concentration (lower HHI) and reduced liquidity requirements (less leverage).

**4.2.1 Portfolios exposure to characteristics.** To examine the fundamental investment properties of the optimal portfolios derived from different prior tail assumptions, we recon-

struct the portfolio exposures to different characteristics. These exposures represent the weighted average of characteristics at the portfolio level, where the weights are determined by maximizing the investor’s utility under the prior (see Section 2). Specifically, the portfolio exposure to a given characteristic  $j$  is calculated as:

$$\widehat{x}_{p,t}^j = \frac{1}{N_t} \sum_{i=1}^{N_t} w_{i,t} \widehat{x}_{i,t}^j, \quad (11)$$

where  $N_t$  denotes the total number of stocks in the portfolio at time  $t$ ,  $w_{i,t}$  is the weight assigned to stock  $i$ , and  $\widehat{x}_{i,t}^j$  is the value of characteristic  $j$  for stock  $i$  at time  $t$ .

Since the cross-sectional mean of each standardized characteristic is always zero, Eq.(11) directly maps the weight sensitivities  $\theta_j$  to the actual portfolio exposure for the corresponding characteristic. This provides an intuitive interpretation of how the investor’s utility-maximizing weights translate into tangible portfolio characteristics under different prior specifications.

Figure 7 presents a heatmap of the absolute value of portfolio characteristics  $|\widehat{x}_{p,t}^j|$  over the out-of-sample period from January 2000 to December 2023. To enhance interpretability, we display the values obtained for the  $\nu = 4$ ,  $\nu = 100$ , and TC2 specifications.<sup>12</sup> The characteristics with the highest exposure remain consistent across prior specifications.

For example, portfolios consistently exhibit exposure to illiquidity (Amihud, 2002), operating profitability (Fama and French, 2006), short interest (Dechow et al., 2001), analysts’ forecast dispersion (Diether et al., 2002), co-skewness (Ang et al., 2006), and size (Banz, 1981), among others. The scale of the exposures is inversely related to the value of  $\nu$ ; smaller  $\nu$  implies more sparsity, which in turn results in a larger exposure to fewer characteristics (intensive margin). In contrast, a thin-tailed prior ( $\nu = 100$ ) produces a smaller but broader

---

<sup>12</sup>Results for the  $\nu = 10, 30$ , TC1 specifications are available upon request from the authors.

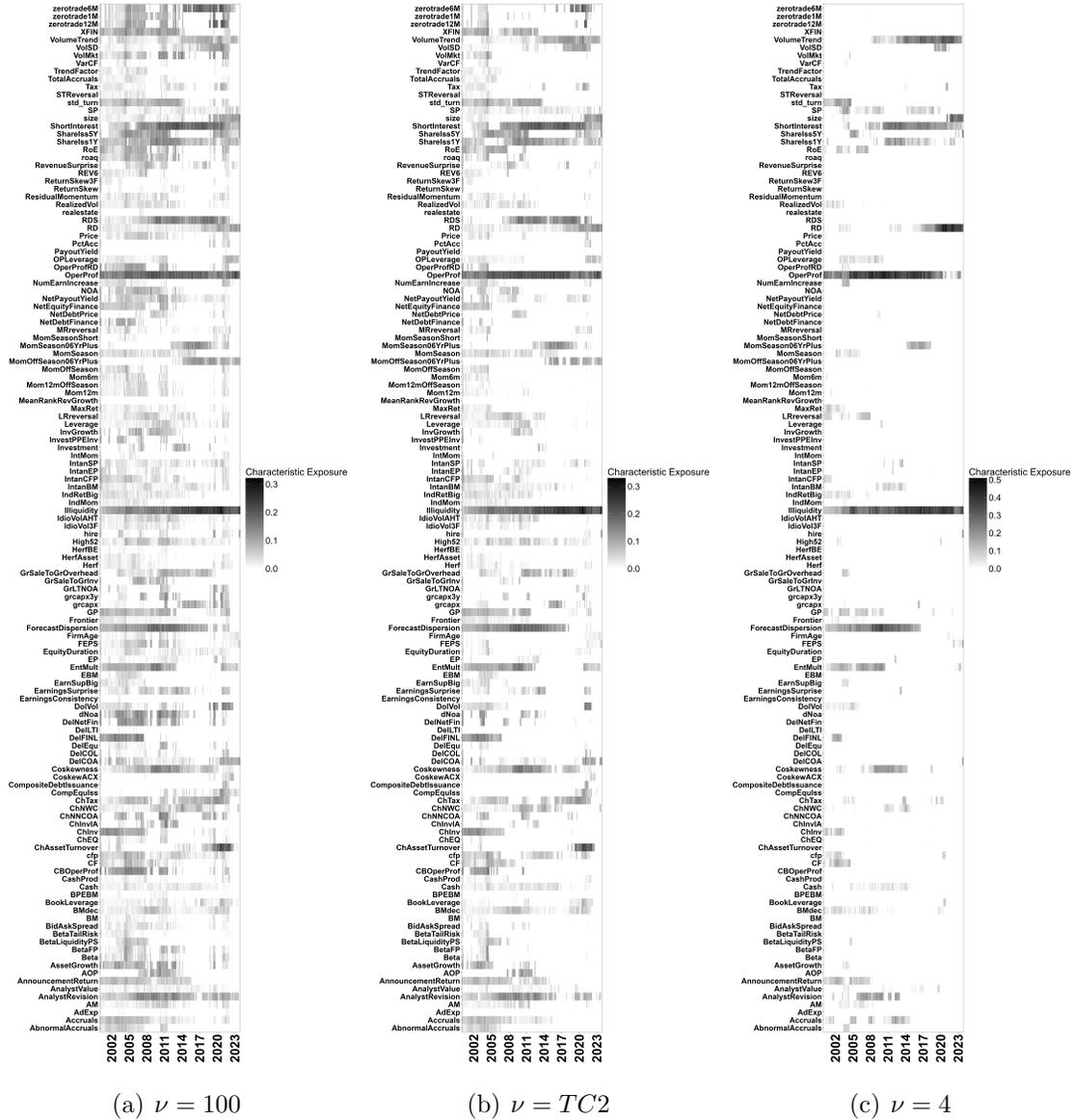


Figure 7: **Portfolio exposure to characteristics.** The figure shows the portfolio exposure to individual characteristics based on different assumptions on the prior tails. The portfolio exposure to a given characteristic is calculated as in Eq.(11). The optimal allocation is based on a rolling window of 240 months. The out-of-sample period is from January 2000 to December 2023.

exposure to firm characteristics (extensive margin). This inverse relationship also persists when calibrating the prior tail thickness using aggregate transaction costs (see Figure 7).

## 5 Out-of-sample portfolio allocation

The analysis of portfolio characteristics highlights the trade-off between sparsity and composition driven by the prior tails. Heavier-tailed priors ( $\nu = 4$ ) concentrate portfolio exposure on a narrower set of firm characteristics, potentially exploiting stronger predictive signals at the expense of increased portfolio concentration. Conversely, thinner-tailed priors ( $\nu = 100$ ) encourage broader diversification, trading firm characteristics at the extensive margin while reducing deviations from the benchmark.

We now evaluate the practical implications of these findings for out-of-sample portfolio performance and investor’s realised utility. Each month, we construct the optimal portfolio  $w_{i,t} = w_{i,t}^b + \sum_{j=1}^k \widehat{\theta}_j \widehat{x}_{i,t}^j$ , where weights are based on the recursive estimates of  $\widehat{\theta}_j$  at time  $t$ . For the allocation, we select the  $j$ -th characteristic whose posterior inclusion probability exceeds 50%.<sup>13</sup> Then, the out-of-sample portfolio return at  $t + 1$  is  $r_{p,t+1} = \sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}$ .

It is important to emphasize that our findings are robust to alternative cutoff of the posterior inclusion probability. For instance, in Section 6.1, we explore a more data-driven approach, whereby a characteristic is excluded from the portfolio policy if its posterior inclusion probability falls below  $1 - \text{mean}(q)$ . This threshold dynamically adjusts the number of characteristics entering the portfolio rule based on the overall sparsity level. All key conclusions remain unchanged, highlighting the robustness of our findings to variations in the inclusion criteria.

To evaluate the economic value of these portfolio strategies, we analyze both the Sharpe ratio (SR) and metrics such as the implied performance fee and the certainty equivalent return ( $\Delta\text{CER}$ ) relative to the equal-weight (EW) portfolio (see DeMiguel et al., 2009). Following

---

<sup>13</sup>Barbieri and Berger (2004) provide theoretical support for the median probability model, which selects variables with inclusion probabilities greater than 0.5. This approach is optimal for prediction under mild regularity conditions.

Della Corte et al. (2008), we compute the implied performance fee  $f$ , which represents the maximum fee an investor would pay to switch from the EW portfolio to another competing strategy.<sup>14</sup> The realized certainty equivalent return is computed as  $CER_p = \bar{r}_p - \frac{\zeta}{2}\hat{\sigma}_p^2$ , where  $\bar{r}_p$  and  $\hat{\sigma}_p^2$  are the mean and variance of portfolio returns over the out-of-sample period. Similarly, the certainty equivalent return of the EW portfolio is denoted as  $CER_{EW}$ . The spread  $\Delta CER = CER_p - CER_{EW}$  quantifies the economic utility a risk-averse, mean-variance investor would gain by investing in a given strategy vis-a-vis the equal-weight benchmark.

Table 1 summarizes the out-of-sample portfolio performance and descriptive statistics for portfolio weights. Panel A examines performance without transaction costs, showing that heavier-tailed priors ( $\nu = 4$ ) achieve the highest mean return (0.125) but also the highest volatility (0.168), resulting in an annualized Sharpe ratio (SR) of 2.54. In contrast, thinner-tailed priors ( $\nu = 100$ ) produce lower mean returns (0.100) but achieve the highest SR (3.122) due to lower volatility (0.110). The latter is lower than the volatility produced by the original Brandt et al. (2009) (BSV) approach (0.122), which relies solely on size, value, and momentum as stock characteristics. The null hypothesis that the Sharpe ratios are comparable to the EW portfolio is strongly rejected based on the block-bootstrap method of Ledoit and Wolf (2008). Additionally, both the performance fee and  $\Delta CER$  are lowest for  $\nu = 4$ , highlighting the economic cost of under-diversified strategies associated with thicker-tailed priors.

Panel B provides descriptive statistics of the portfolio weights, further illustrating the trade-off between sparsity and diversification across prior specifications. Heavier-tailed priors ( $\nu = 4$ ) result in more extreme allocations, with maximum weights reaching 3.965% and

---

<sup>14</sup>For mean-variance utility, the implied fee  $f$  is calculated as:

$$\sum_{t=1}^T (r_{p,t} - \text{fee}) - \frac{\zeta}{2} (r_{p,t} - \text{fee})^2 = \sum_{t=1}^T r_{Bench,t} - \frac{\zeta}{2} r_{Bench,t}^2, \quad (12)$$

where  $\zeta = 5$  is the risk aversion parameter in our application, and  $r_{Bench,t}$  denotes the benchmark portfolio returns, set to EW in this case.

**Panel A:** Portfolio performance w/o transaction costs

	EW	BSV	GLP-t( $\nu$ )					
			4	10	30	100	TC1	TC2
Mean	0.014	0.061	0.125	0.110	0.102	0.100	0.107	0.100
Std	0.061	0.122	0.168	0.141	0.119	0.110	0.126	0.111
ES(5%)	-0.118	-0.159	-0.159	-0.147	-0.100	-0.097	-0.117	-0.091
SR (annual)	0.725	1.697	2.549	2.678	2.956	3.122	2.920	3.099
$pval(\Delta SR)_{boot}$		0.003	0.000	0.000	0.000	0.000	0.000	0.000
$\Delta CER$		0.011	0.012	0.029	0.042	0.047	0.040	0.047
Performance fee		0.014	0.025	0.046	0.058	0.062	0.057	0.062

**Panel B:** Portfolio weight statistics

Max w (%)	0.030	0.673	3.965	3.163	2.665	2.497	3.016	2.520
Min w (%)	0.030	-0.843	-3.557	-3.019	-2.732	-2.629	-2.959	-2.653
Mean  w  (%)	0.030	0.175	0.441	0.371	0.338	0.334	0.362	0.333
Mean ( $w < 0$ ) (%)	0.000	-0.178	-0.420	-0.364	-0.336	-0.336	-0.356	-0.334
Turnover	0.109	1.106	6.901	5.864	5.703	6.023	6.059	5.967
Average HHI	0.000	0.017	0.157	0.108	0.085	0.080	0.102	0.080

**Panel C:** Portfolio performance net of transaction costs

Mean	0.013	0.053	0.070	0.066	0.061	0.057	0.063	0.058
Std	0.061	0.119	0.152	0.129	0.106	0.098	0.114	0.099
ES(5%)	-0.120	-0.169	-0.205	-0.188	-0.137	-0.139	-0.156	-0.131
SR (annual)	0.680	1.494	1.569	1.743	1.933	1.972	1.882	1.976
$pval(\Delta SR)_{boot}$		0.006	0.007	0.004	0.002	0.000	0.002	0.001
$\Delta CER$		0.007	0.007	0.011	0.021	0.023	0.019	0.023
Performance fee		0.009	0.009	0.014	0.026	0.027	0.023	0.027

Table 1: **Out-of-sample portfolios.** This table reports the out-of-sample portfolio performance without (Panel A) and with (Panel B) transaction costs. Transaction costs are proxied by the stock-specific half bid-ask spread and are imputed each month based on the portfolio rebalancing (turnover) as in [DeMiguel et al. \(2009\)](#). Panel B reports a series of descriptive statistics based on average values over the out-of-sample period. The out-of-sample period is from January 2000 to December 2023.

minimum weights dropping to  $-3.557\%$ . Larger leverage is coupled by reduced diversification as shown by the HHI, which is highest for  $\nu = 4$  (0.157). In contrast, thinner-tailed priors ( $\nu = 100$ ) exhibit broader diversification, characterized by smaller mean absolute weights (0.334%), less concentration (0.080), and reduced turnover (6.023). The latter is calculated as  $\sum |w_{i,t} - w_{i,t-1}^+|$ , where  $w_{i,t}$  represents the portfolio weight of asset  $i$  at time  $t$ , and  $w_{i,t-1}^+ \equiv w_{i,t-1}(1 + r_{i,t})$  denotes the adjusted weight from the previous period.

Panel C incorporates transaction costs into the performance analysis. Following [DeMiguel et al. \(2009\)](#), the impact of trading costs on portfolio performance is modeled as  $r_{p,t}^{Net} = (1 + r_{p,t}) \left( 1 - \sum_{i=1}^{N_t} \frac{\eta_{i,t}}{2} |w_{i,t} - w_{i,t-1}^+| \right) - 1$ , where  $\eta_{i,t}$  denotes the bid-ask spread for asset  $i$  at time  $t$  as proxied by [Corwin and Schultz \(2012\)](#). The half bid-ask spread represents a proxy for individual trading costs linked to liquidity (e.g., [Bessembinder and Venkataraman, 2010](#)). Accounting for transaction costs reduces mean returns across all specifications, with heavier-tailed priors ( $\nu = 4$ ) experiencing the largest decline (from 0.125 to 0.070).

Nevertheless, thinner-tailed priors ( $\nu = 100$ ) maintain the highest Sharpe ratio (1.972) across strategies, demonstrating resilience to transaction costs through broader diversification and reduced turnover. Additionally, the implied performance fee remains higher for thinner-tailed priors, underscoring their economic value even when transaction costs are considered. Transaction-cost-calibrated priors strike a balance, achieving performance and diversification metrics comparable to thinner-tailed priors while explicitly addressing the practical implications of trading costs.

## 5.1 Performance under economic restrictions

[Karolyi and Van Nieuwerburgh \(2020\)](#) argue that it is crucial to examine the economic underpinnings of complex statistical methods that seek to extract additional profits by leveraging information embedded in firm characteristics. To address this, we follow the logic in [Avramov et al. \(2023a\)](#) and extend the main out-of-sample portfolio implementation to comprehensively assess the impact of economic restrictions on performance across different prior specifications.

In the cross-section, we restrict the universe of stocks to those that are relatively inexpensive to trade by focusing exclusively on NYSE-listed stocks (e.g., [Simon et al., 2023](#)). In the time series, we investigate whether portfolio profitability diminishes when ex-post constraints are applied to portfolio weights. These constraints, imposed at each time  $t$ , either

limit the leverage permitted for trading a given stock or prohibit short positions (e.g., [Jones and Lamont, 2002](#)).

**5.1.1 Stocks listed on NYSE.** We consider a sample of common stocks (share codes 10 and 11) traded exclusively on the NYSE. This results in a substantially smaller unbalanced panel of 5,186 stocks, with a minimum of 1,149, a maximum of 1,848, and an average of 1,405 stocks per month. These stocks typically have larger market capitalizations and higher liquidity, which arguably could represent a more realistic investment universe for liquidity-constrained investors.

To account for the sample differences, we recalculated the characteristic-managed portfolios using the new cross-section and re-estimated all model parameters. Consistent with the main empirical framework, we constructed the optimal portfolio based on the recursive estimates of  $\theta_j, q, \gamma^2$  up to time  $t$ . For the allocation at time  $t$ , we selected the  $j$ -th characteristic with a posterior inclusion probability exceeding 50%. [Section 6.1](#) also explore the portfolio performance based on a  $1 - \text{mean}(q)$  cutoff.

Panel A of [Table 2](#) presents the portfolio weight statistics. Portfolios constructed in this sample exhibit greater concentration compared to the full sample. For example, under the  $\nu = 4$  prior, the HHI is significantly higher (0.342) compared to the full sample (0.157). This indicates that portfolios invested in NYSE-only stocks tend to deviate more strongly from the equal-weight benchmark  $w_{i,t}^b$ .

Weight magnitudes also differ significantly between the samples. Portfolios invested in NYSE-only stocks allocate larger capital on average. For instance, the mean  $|w|$  under  $\nu = 4$  is 1.022% in the NYSE sample, compared to 0.441% in the full sample. These differences are likely driven by the smaller, more liquid universe of NYSE stocks, where larger and more frequent portfolio adjustments are more feasible compared to the full cross-section of stocks

**Panel A:** Portfolio weight statistics

	EW	BSV	GLP-t( $\nu$ )					
			4	10	30	100	TC1	TC2
Max w (%)	0.075	1.372	8.620	7.045	6.135	5.806	6.730	5.943
Min w (%)	0.075	-1.424	-6.909	-5.748	-6.132	-6.098	-5.985	-6.124
Mean  w  (%)	0.075	0.290	1.022	0.901	0.960	0.946	0.943	0.955
Mean ( $w < 0$ ) (%)	0.000	-0.310	-0.987	-0.871	-0.948	-0.938	-0.928	-0.946
Turnover	0.094	1.011	10.182	8.982	8.785	8.442	9.348	8.592
Average HHI	0.001	0.023	0.342	0.238	0.240	0.228	0.247	0.233

**Panel B:** Portfolio performance net of transaction costs

Mean	0.011	0.013	0.038	0.033	0.035	0.036	0.034	0.035
Std	0.060	0.081	0.196	0.133	0.114	0.108	0.113	0.108
ES(5%)	-0.129	-0.179	-0.448	-0.279	-0.261	-0.243	-0.247	-0.245
SR (annual)	0.574	0.482	0.639	0.828	1.019	1.094	1.000	1.090
$pval(\Delta SR)_{boot}$		0.604	0.873	0.478	0.063	0.041	0.082	0.043
$\Delta CER$		-	-	0.006	0.013	0.016	0.014	0.016
Performance Fee		-	-	0.007	0.016	0.018	0.016	0.019

Table 2: **Out-of-sample portfolios for NYSE-listed common stocks.** This table reports the out-of-sample portfolio weights (Panel A) and performance net of transaction costs (Panel B) for a sub-sample of stocks traded only on the NYSE. Transaction costs are proxied by the stock-specific half bid-ask spread and are imputed each month based on the portfolio rebalancing (turnover) as in [DeMiguel et al. \(2009\)](#). The out-of-sample period is from January 2000 to December 2023.

listed on NYSE/AMEX/NASDAQ exchanges.

Despite differences across samples, consistent with the main results, a heavier-tailed prior ( $\nu = 4$ ) consistently generates more concentrated portfolios, characterized by larger weight magnitudes and higher turnover, compared to thinner-tailed priors ( $\nu = 100, TC1, TC2$ ). These findings confirm that the trade-off between sparsity and shrinkage remains robust even within the NYSE subsample.

Panel B of Table 2 reports portfolio performance net of transaction costs (see Panel C of Table 1 for full-sample results). Risk-adjusted returns are consistently lower across all priors in the NYSE sample. The smaller, more liquid stock universe leads to moderately lower mean returns, while portfolio volatility remains relatively comparable to that of the

full sample. Certainty equivalent returns (CER) follow a similar pattern, with higher values observed in the full sample.

Nevertheless, thin-tailed prior specifications deliver substantially higher economic utility compared to smaller values of  $\nu$ , the equal-weight benchmark, and the original BSV approach. For instance, the Sharpe ratio (SR) for  $\nu = 100$  is 1.094 on an annualized basis, significantly outperforming the SRs of 0.574, 0.482, and 0.639 achieved by EW, BSV, and  $\nu = 4$ , respectively. Transaction-cost-calibrated priors, such as TC2, closely follow, with an SR of 1.090. A heavier-tailed prior also yields significantly lower SRs,  $\Delta\text{CER}$ , and performance fees once transaction costs are accounted for, underscoring the persistent limitations of sparsity even in a more efficient and liquid stock universe.

**5.1.2 Imposing portfolio constraints.** Panel A of Table 3 presents the out-of-sample portfolio performance net of transaction costs for leverage-constrained portfolios ( $w_{i,t} \in (-3\%, 3\%) \forall i, t$ ). Mean returns and Sharpe ratios remain largely consistent with those of the unconstrained portfolio allocation (see Panel C in Table 1). Performance fees and certainty equivalent returns ( $\Delta\text{CER}$ ) also exhibit minimal differences between leverage-constrained and unconstrained portfolios. These results indicate that portfolio constraints primarily temper extreme portfolio positions without significantly impacting the overall risk-return profile.

In contrast, Panel B shows that the imposition of no-short-sales constraints ( $w_{i,t} \geq 0 \forall i, t$ ) has a pronounced effect on portfolio performance. Mean returns decline significantly across all prior specifications, reflecting the restrictive nature of long-only portfolios, which limits their ability to leverage negative predictive signals (e.g., Brennan and Lo, 2010). Sharpe ratios are also markedly lower under no-short-sales constraints, although they remain significantly higher than those of the benchmark EW portfolio. For example, with  $\nu = 100$ , the SR decreases from 1.968 in the unconstrained allocation to 1.040. This compares to 0.680 of the

**Panel A:** Leverage constraints  $w_{i,t} \in (-3\%, 3\%)$ 

	EW	BSV	GLP-t( $\nu$ )					
			4	10	30	100	TC1	TC2
Mean	0.013	0.053	0.068	0.066	0.061	0.057	0.063	0.058
Std	0.061	0.119	0.149	0.129	0.106	0.098	0.113	0.099
ES(5%)	-0.120	-0.169	-0.203	-0.187	-0.137	-0.139	-0.155	-0.131
SR (annual)	0.680	1.494	1.556	1.738	1.930	1.968	1.875	1.972
$pval(\Delta SR)_{boot}$		0.006	0.008	0.004	0.002	0.000	0.002	0.001
$\Delta CER$		0.007	0.007	0.011	0.021	0.023	0.018	0.023
Performance fee		0.009	0.009	0.014	0.026	0.027	0.023	0.027

**Panel B:** No-short sales constraints  $w_{i,t} \geq 0$ 

Mean	0.013	0.021	0.021	0.021	0.020	0.020	0.020	0.020
Std	0.061	0.061	0.063	0.060	0.060	0.061	0.060	0.061
ES(5%)	-0.120	-0.110	-0.115	-0.110	-0.110	-0.111	-0.111	-0.111
SR (annual)	0.680	1.108	1.058	1.103	1.076	1.040	1.088	1.046
$pval(\Delta SR)_{boot}$		0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\Delta CER$		0.007	0.006	0.007	0.007	0.006	0.007	0.006
Performance fee		0.008	0.007	0.008	0.007	0.006	0.007	0.006

Table 3: **Out-of-sample portfolios with allocation constraints.** This table reports the out-of-sample portfolio performance with transaction costs with leverage (Panel A) and no-short sales (Panel B) constraints. Transaction costs are proxied by the stock-specific half bid-ask spread and are imputed each month based on the portfolio rebalancing (turnover) as in [DeMiguel et al. \(2009\)](#). The out-of-sample period is from January 2000 to December 2023.

EW benchmark (p-value = 0.000). Similarly, the performance fee and  $\Delta CER$  are substantially lower, underscoring the diminished economic performance in the absence of short positions.

**5.1.3 Limits to arbitrage.** Economic theory suggests that lower trading frictions and greater arbitrage activity enhance price efficiency. In contrast, when limits-to-arbitrage are binding, characteristic-based trading strategies can become more profitable due to temporary mispricings. For example, theoretical work predicts that higher volatility reduces market makers' ability to provide liquidity because of tighter funding constraints and reduced risk appetite (e.g., [Brunnermeier and Pedersen, 2009](#); [Adrian and Shin, 2010](#)). This liquidity shortfall, particularly during financial turmoil, can exacerbate mispricings and amplify anomaly

payoffs. Supporting this mechanism, [Chordia et al. \(2014\)](#) show that increased stock market liquidity weakens equity return anomalies, likely due to improved arbitrage activity and the correction of mispricings.

To explore this point, we test whether the profitability of investment strategies under different prior specifications is more pronounced during high limits-to-arbitrage market states, such as periods of elevated volatility and tight financial conditions. We classify financial conditions using the National Financial Conditions Index (NFCI), published by the Federal Reserve Bank of Chicago.<sup>15</sup> The NFCI aggregates 105 indicators spanning money markets, debt and equity markets, banking systems, and shadow banking sectors. Standardized values above zero indicate tighter-than-average financial conditions (reduced liquidity), while values below zero reflect looser-than-average conditions, typically associated with greater market liquidity. Similarly, we define high versus low volatility states based on the median monthly value of the implied volatility index (VIX) of S&P 500 options.

Table 4 reports portfolio performance net of transaction costs during periods of tight versus loose financial conditions. Across all prior specifications, Sharpe Ratios are consistently higher in loose financial conditions (Panel B) compared to tight conditions (Panel A). Within each panel, increasing the degree of shrinkage (higher  $\nu$ ) leads to better performance, with SRs peaking at  $\nu = 100$ : 1.624 under tight conditions and 2.485 under loose conditions.

Notably, the performance differential between  $\nu = 100$  and  $\nu = 4$  is more pronounced when financial conditions are tight. Under tight conditions, the SR improves from 1.194 ( $\nu = 4$ ) to 1.624 ( $\nu = 100$ ), a substantial increase of 36%. In contrast, the improvement under loose financial conditions is smaller but still notable, with the SR rising from 2.144 ( $\nu = 4$ ) to 2.485 ( $\nu = 100$ ), a gain of approximately 16%.

---

<sup>15</sup>Current NFCI data can be accessed at <https://www.chicagofed.org/research/data/nfci/current-data>.

**Panel A:** Tight financial conditions

	EW	BSV	GLP-t( $\nu$ )					
			4	10	30	100	TC1	TC2
Mean	0.009	0.056	0.062	0.063	0.058	0.054	0.058	0.055
Std	0.075	0.137	0.175	0.151	0.122	0.112	0.124	0.112
ES(5%)	-0.142	-0.203	-0.244	-0.234	-0.173	-0.173	-0.183	-0.156
SR (annual)	0.355	1.372	1.194	1.407	1.619	1.624	1.600	1.652
$pval(\Delta SR)_{boot}$		0.004	0.039	0.014	0.005	0.004	0.006	0.004
$\Delta CER$		0.006	-0.018	0.001	0.018	0.021	0.017	0.022
Performance Fee		0.008	-	0.002	0.023	0.026	0.022	0.026

**Panel B:** Loose financial conditions

Mean	0.018	0.050	0.079	0.070	0.063	0.060	0.067	0.061
Std	0.044	0.099	0.125	0.103	0.089	0.082	0.103	0.084
ES(5%)	-0.070	-0.107	-0.134	-0.097	-0.070	-0.078	-0.095	-0.078
SR (annual)	1.298	1.703	2.144	2.302	2.411	2.485	2.234	2.447
$pval(\Delta SR)_{boot}$		0.162	0.039	0.029	0.018	0.012	0.031	0.015
$\Delta CER$		0.008	0.013	0.022	0.024	0.026	0.020	0.025
Performance Fee		0.009	0.019	0.027	0.028	0.029	0.025	0.028

Table 4: **Out-of-sample portfolios based on financial conditions.** This table reports the out-of-sample performance net of transaction costs during periods of tight financial conditions (top panel) vs. loose financial conditions (bottom panel). Financial conditions are identified based on the National Financial Conditions Index (NFCI), published by the Federal Reserve Bank of Chicago. Transaction costs are proxied by the stock-specific half bid-ask spread and are imputed each month based on the portfolio rebalancing (turnover) as in [DeMiguel et al. \(2009\)](#). The out-of-sample period is from January 2000 to December 2023.

Table [D.2](#) in Appendix [D.3](#) reports out-of-sample performance net of transaction costs during periods of high volatility (Panel A) and low volatility (Panel B). Under high volatility, the SR improves notably as the degree of shrinkage ( $\nu$ ) increases: at  $\nu = 4$ , the Sharpe Ratio is 1.437, rising to 1.887 at  $\nu = 100$ —an improvement of approximately 31%. In low-volatility periods, higher shrinkage ( $\nu = 100$ ) also enhances performance, but the relative gain is smaller, suggesting that shrinkage plays a less critical role when market conditions are calmer and liquidity is greater.

Overall, the spread in Sharpe Ratios between high and low volatility states, as well as between tight and loose financial conditions, narrows as  $\nu$  increases. This indicates that

greater shrinkage stabilizes performance across varying market environments. These findings underscore the importance of ridge-type shrinkage over sparsity-inducing priors, particularly during periods of heightened risk, financial stress, and elevated limits to arbitrage.

## 6 Additional portfolio results

We evaluate the robustness of the main empirical findings to variations in the cutoff used for selecting characteristics based on posterior inclusion probabilities and the type of Bayesian prior employed to estimate portfolio tilts. Additionally, we conduct time-series factor-spanning regressions, where the realized portfolio returns for a given prior specification are regressed on the returns of the equal-weight benchmark and a range of control factors. These include the size and value factors from [Fama and French \(1993\)](#) and the momentum factor from [Jegadeesh and Titman \(1993\)](#).<sup>16</sup>

### 6.1 Data-driven selection threshold

Table 5 presents portfolio performance under an “adaptive” selection criterion, where a characteristic is excluded if its posterior inclusion probability falls below  $1 - \text{mean}(q)$ . For example, if  $\text{mean}(q) = 0.4$  in a given period, all  $\theta_j$  with posterior inclusion probabilities exceeding 0.6 are selected. This approach reflects the underlying logic that stronger evidence in favor of sparsity (smaller  $q$ ) should lead to stricter selection criteria.

Panel A shows that the portfolio weight statistics exhibit similar ranges for maximum and minimum weights, average absolute weights ( $|w|$ ), and turnover. For example, the maximum weight for the  $\nu = 100$  specification is 2.452%, compared to 2.497% in Table 1. Similarly, the turnover for the same specification is 5.895, compared to 6.023 based on the 50% cutoff.

---

<sup>16</sup>Data on the size, value, and momentum factors are sourced from Kenneth French’s website. We thank the authors for making the data publicly available.

**Panel A:** Portfolio weight statistics

	EW	BSV	GLP-t( $\nu$ )					
			4	10	30	100	TC1	TC2
Max w (%)	0.030	0.673	3.845	3.077	2.611	2.452	2.936	2.468
Min w (%)	0.030	-0.843	-3.249	-2.741	-2.586	-2.568	-2.799	-2.565
Mean  w  (%)	0.030	0.175	0.405	0.343	0.320	0.323	0.339	0.320
Mean ( $w < 0$ ) (%)	0.000	-0.178	-0.382	-0.332	-0.318	-0.325	-0.331	-0.321
Turnover	0.109	1.106	5.914	5.149	5.391	5.895	5.678	5.840
Average HHI	0.000	0.017	0.145	0.098	0.079	0.076	0.095	0.076

**Panel B:** Portfolio performance net of transaction costs

Mean	0.013	0.053	0.054	0.055	0.058	0.056	0.059	0.057
Std	0.061	0.119	0.187	0.143	0.115	0.101	0.125	0.101
ES(5%)	-0.120	-0.169	-0.318	-0.233	-0.137	-0.134	-0.156	-0.134
SR (annual)	0.680	1.494	0.981	1.301	1.693	1.879	1.620	1.899
$pval(\Delta SR)_{boot}$		0.006	0.412	0.063	0.004	0.000	0.007	0.000
$\Delta CER$		0.007	0.001	0.007	0.013	0.021	0.009	0.021
Performance Fee		0.009	0.001	0.006	0.017	0.025	0.012	0.025

Table 5: **Out-of-sample portfolios based on  $1 - \text{mean}(q)$  selection cutoff.** This table reports the out-of-sample portfolio weights (Panel A) and performance net of transaction costs (Panel B) based on a  $1 - \text{mode}(q)$  cutoff to select a given characteristics. Transaction costs are proxied by the stock-specific half bid-ask spread and are imputed each month based on the portfolio rebalancing (turnover) as in [DeMiguel et al. \(2009\)](#). The out-of-sample period is from January 2000 to December 2023.

The HHI, which measures portfolio concentration, is also largely consistent across selection criteria. For instance, the  $TC2$  specification has an HHI of 0.076 in Table 5, compared to 0.080 in Table 1.

These results indicate that both selection cutoffs produce relatively diversified portfolios for thin-tailed priors, with lower concentration compared to thick-tailed priors. That is, the overall allocation patterns remain largely unchanged. Panel B also demonstrates that portfolio performance metrics net of transaction costs, including mean returns, standard deviations, and Sharpe ratios, exhibit only minor variations across selection criteria.

Overall, the findings suggest that the choice of selection cutoff has a relatively modest effect on portfolio allocations and performances net of transaction costs. The data-driven

1 – mean( $q$ ) cutoff results in slightly lower turnover and reduced concentration (lower HHI), while the 50% cutoff delivers marginally higher economic performance. These results indicate that both selection criteria yield consistent portfolio outcomes, underscoring the robustness of the portfolio policy.

## 6.2 Alternative priors

In addition to the Student- $t$  prior, we evaluate the out-of-sample performance of two widely used shrinkage priors, such as the Bayesian lasso (Park and Casella, 2008) and the horseshoe (Carvalho et al., 2009), as well as popular Bayesian variable selection priors such as the mixture of normals of George and McCulloch (1993) and the normal dirac spike-and-slab of Giannone et al. (2021).

Park and Casella (2008) extended the work of Tibshirani (1996) by proposing a prior of the form:

$$\theta_j \sim \mathcal{N}(0, \sigma^2 \lambda_j^2), \quad \lambda_j^2 \sim \mathcal{E}\left(\frac{\gamma^2}{2}\right), \quad \gamma^2 \sim \text{IG}(a, b),$$

where  $\mathcal{E}\left(\frac{\gamma^2}{2}\right)$  denotes an exponential distribution with rate parameter  $\frac{\gamma^2}{2}$ .<sup>17</sup> This formulation is analogous to the penalty term in conventional lasso regression; a larger  $\gamma^2$  concentrates the prior more tightly around zero. The horseshoe prior (HS) takes the form:

$$\theta_j \sim \mathcal{N}(0, \sigma^2 \gamma^2 \lambda_j^2), \quad \lambda_j^2 \sim \mathcal{C}^+(0, 1), \quad \gamma^2 \sim \mathcal{C}^+(0, 1),$$

where  $\mathcal{C}^+(0, 1)$  represents the half-Cauchy distribution on the positive reals with a scale parameter of one.

Finally, we also examine the original stochastic search variable selection (SSVS) of George

---

<sup>17</sup>Tibshirani (1996) first observed that the frequentist lasso estimate could be interpreted as a Bayes posterior mode under a Laplace prior.

and McCulloch (1993) and the normal spike-and-slab proposed by Giannone et al. (2021) (GLP). SSVS represents a computationally convenient approach to Bayesian variable selection which is based on a mixture of two normal distributions (e.g., Narisetty et al., 2019). The GLP prior shares the same structure as the specification outlined in Section 2.1, except that the Student-t distribution in the mixture is replaced with a conventional normal distribution with mean zero and variance  $\gamma^2\sigma^2$ , where  $\gamma^2$  governs the degree of shrinkage applied to  $\theta_j, j = 1, \dots, k$ . Appendix E provides additional details on the posterior distributions for both the Bayesian lasso, the horseshoe, and the SSVS prior. For a comprehensive explanation of the normal spike-and-slab prior, we refer the reader to Giannone et al. (2021).

We note that the posterior estimates of  $\theta_j, j = 1, \dots, k$  under the Bayesian lasso and the horseshoe prior are non-sparse, making direct comparisons with variable selection tools such as our heavy-tailed spike-and-slab prior challenging. To address this limitation, we implement the Signal Adaptive Variable Selector (SAVS) algorithm proposed by Ray and Bhattacharya (2018) to ex-post induce sparsity in the posterior estimates  $\hat{\theta}_j, j = 1, \dots, k$ , conditional on a given prior. Appendix E.1 provides a detailed discussion of this method and its key advantages.<sup>18</sup>

Table 6 presents results for the full and the NYSE-only samples. Regarding portfolio weights (Panel A), the Bayesian lasso (Blasso) and SSVS produce significantly more concentrated portfolios (HHI equal to 0.346 and 0.149 respectively) and substantially higher turnover (12.319 and 8.210) compared to a thin-tailed specification such as GLP (HHI = 0.080, turnover = 6.235). This pattern persists even when the investable universe is restricted to stocks listed on NYSE. In contrast, the horseshoe prior generates portfolio compositions comparable to those of the GLP specification, exhibiting similar levels of concentration and

---

<sup>18</sup>The SAVS algorithm is an automatic procedure in which the degree of sparsity directly depends on the effectiveness of the shrinkage applied to  $\hat{\theta}_j$ . This property makes it a natural tool for comparing different estimation methods.

**Panel A:** Portfolio weight statistics

	EW	BSV	Blasso	HS	SSVS	GLP	EW	BSV	Blasso	HS	SSVS	GLP
	Full sample						NYSE only					
Max w (%)	0.030	0.673	5.890	2.507	4.007	2.455	0.075	1.372	12.718	5.634	7.479	5.623
Min w (%)	0.030	-0.843	-5.761	-2.347	-3.438	-2.624	0.075	-1.424	-12.848	-4.233	-7.274	-5.973
Mean  w  (%)	0.030	0.175	0.721	0.282	0.460	0.338	0.075	0.290	2.043	0.665	1.037	0.928
Mean ( $w < 0$ ) (%)	0.000	-0.178	-0.720	-0.274	-0.457	-0.344	0.000	-0.310	-2.048	-0.630	-1.047	-0.920
Turnover	0.109	1.106	12.319	4.526	8.120	6.235	0.094	1.011	15.522	7.030	9.721	8.212
Average HHI	0.000	0.017	0.346	0.065	0.149	0.080	0.001	0.023	1.036	0.139	0.296	0.218

**Panel B:** Portfolio performance net of transaction costs

	Full sample						NYSE only					
Mean	0.013	0.053	0.028	0.048	0.053	0.056	0.011	0.013	0.027	0.030	0.024	0.035
Std	0.061	0.119	0.139	0.095	0.123	0.093	0.060	0.081	0.191	0.118	0.141	0.105
ES(5%)	-0.120	-0.169	-0.299	-0.134	-0.217	-0.123	-0.129	-0.179	-0.373	-0.252	-0.318	-0.232
SR (annual)	0.680	1.494	0.659	1.716	1.454	2.041	0.574	0.482	0.456	0.834	0.553	1.131
$pval(\Delta SR)_{boot}$		0.006	0.947	0.003	0.017	0.000		0.604	0.716	0.399	0.956	0.072
$\Delta CER$		0.007	-0.025	0.018	0.008	0.025		-0.006	-0.060	-0.009	-0.025	0.003
Performance fee		0.009	-	0.020	0.006	0.029		-	-	0.001	-	0.004

Table 6: **Out-of-sample portfolios based on alternative priors.** This table reports the out-of-sample portfolio weights (Panel A) and performance net of transaction costs (Panel B) for alternative prior specifications, such as the Bayesian lasso (Park and Casella, 2008), the horseshoe (Carvalho et al., 2009), the mixture of normals prior of George and McCulloch (1993), and the dirac spike-and-slab of Giannone et al. (2021). We report the results for the full sample of stocks and the NYSE-only subsample. Transaction costs are proxied by the stock-specific half bid-ask spread and are imputed each month based on the portfolio rebalancing (turnover) as in DeMiguel et al. (2009). The out-of-sample period is from January 2000 to December 2023.

turnover.

Panel B reports the out-of-sample portfolio performance net of transaction costs. The GLP prior achieves a SR for the full sample that is two times higher than that of the Blasso and around 50% higher than the SSVS. The gap expands for stocks listed on the NYSE, with the SR of Blasso and SSVS less than half of GLP. Realized utility, as measured by  $\Delta CER$  and performance fees, also strongly favors the thin-tailed spike-and-slab specification.

These findings reinforce the intuition from the main empirical results: a more conservative approach that emphasizes shrinkage over sparsity delivers superior out-of-sample economic utility and risk-adjusted returns when transaction costs are taken into account.

### 6.3 Spanning regressions

We follow [Barillas and Shanken \(2017\)](#) and perform factor-spanning regressions to test the incremental mean-variance efficiency of each prior portfolio compared to the equal-weight benchmark. Specifically, we estimate the following time-series regression:

$$r_{p,t} = \alpha + \beta r_{EW,t} + \phi' \text{controls} + \varepsilon_t,$$

where  $r_{p,t}$  represents the return net of transaction costs for the portfolio under a given prior, and  $r_{EW,t}$  denotes the return net of transaction costs for the benchmark equal-weight portfolio. The set of controls include the size and value factors from [Fama and French \(1993\)](#) and the momentum factor from [Jegadeesh and Titman \(1993\)](#).

	BSV		GLP-t( $\nu$ )				
		4	10	30	100	TC1	TC2
	Full sample						
$\alpha$	0.031	0.075	0.069	0.064	0.058	0.067	0.059
p-value $\alpha$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Appraisal Ratio	0.427	0.508	0.555	0.615	0.604	0.610	0.612
	NYSE sample						
$\alpha$	-0.005	0.034	0.030	0.030	0.031	0.030	0.031
p-value $\alpha$	0.066	0.004	0.000	0.000	0.000	0.000	0.000
Appraisal Ratio	-0.109	0.177	0.228	0.282	0.299	0.275	0.296

**Table 7: Spanning regressions.** This table reports the results of spanning regressions of the form  $r_{p,t} = \alpha + \beta r_{EW,t} + \gamma' \text{controls} + \varepsilon_t$ , where the portfolio return  $r_{p,t}$  and the EW benchmark return  $r_{EW,t}$  are constructed based on the full cross section of stocks (top panel) and the NYSE-only sample (bottom panel). Returns are net of transaction costs. Transaction costs are proxied by the stock-specific half bid-ask spread and are imputed each month based on the portfolio rebalancing (turnover) as in [DeMiguel et al. \(2009\)](#). The out-of-sample period is from January 2000 to December 2023.

Table 7 reports the regression results for both the full sample of stocks and the subsample consisting of NYSE-only stocks. For each specification, we present the regression for the

original BSV approach and all Student- $t$  prior configurations ( $\nu = [4, 10, 30, 100, \text{TC1}, \text{TC2}]$ ). Additionally, we calculate and report the excess Sharpe ratio, or “appraisal ratio”, defined as  $\alpha/\sigma_\varepsilon$ , where  $\sigma_\varepsilon$  is the root mean squared error. This metric evaluates the extent to which the portfolio allocation under a given prior enhances the slope of the mean-variance efficient frontier relative to the EW benchmark (e.g., [Fama and French, 2018](#)).

The findings reveal that the EW portfolio return fails to account for the net returns generated by BSV or any of the prior specifications. A thin-tailed prior ( $\nu = 100$ ) or a prior calibrated to transaction costs ( $\nu = \text{TC2}$ ) yields substantially higher appraisal ratios compared to both heavy-tailed specifications and the original BSV. This advantage is particularly pronounced in the NYSE-only sample, where BSV generates a negative appraisal ratio, while  $\nu = 100$  achieves an appraisal ratio nearly twice that of  $\nu = 4$ .

These results reinforce the main empirical findings in [Section 5](#): (1) portfolio returns net of transaction costs generated by different priors consistently outperform those of the benchmark equal-weight portfolio and BSV, and (2) thin-tailed priors that prioritise shrinkage over sparsity outperform heavy-tailed, sparsity-inducing priors in expanding the mean-variance efficient frontier.

## 7 Conclusions

This paper explore the role of firm characteristics for optimal portfolio allocation and ultimately for cross-sectional stock returns predictability. By employing a flexible Bayesian variable selection prior, we investigate the trade-off between sparsity and shrinkage on mean-variance portfolio performance.

Our findings emphasize the pervasive nature of model uncertainty and question the practicality of sparsity from a portfolio allocation perspective. Heavy-tailed priors, while reducing

uncertainty about the most relevant firm characteristics for return prediction, produce more concentrated portfolios with extreme weights and elevated turnover. In contrast, thin-tailed priors or those calibrated to transaction costs favor the inclusion of firm characteristics at the extensive margin rather than the intensive margin. This approach results in more diversified portfolios with lower turnover and superior economic performance. These findings underscore the need to balance sparsity and shrinkage in portfolio allocation models, particularly in the presence of market frictions.

Overall, our study demonstrates that the choice between sparse and dense models to map firm characteristics into the cross-section of stock returns can have importance economic implications. Broadly incorporating firm characteristics enhances portfolio diversification and delivers more robust out-of-sample performance, especially when transaction costs are accounted for.

## References

- Abadie, A. and Kasy, M. (2019). Choosing among regularized estimators in empirical economics: The risk of machine learning. *Review of Economics and Statistics*, 101(5):743–762.
- Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *Journal of financial intermediation*, 19(3):418–437.
- Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1):31–56.
- Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1):99–102.
- Ang, A., Chen, J., and Xing, Y. (2006). Downside risk. *The review of financial studies*, 19(4):1191–1239.
- Armagan, A. and Zaretzki, R. L. (2010). Model selection via adaptive shrinkage with t priors. *Computational Statistics*, 25:441–461.
- Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458.

- Avramov, D., Cheng, S., and Metzker, L. (2023a). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science*, 69(5):2587–2619.
- Avramov, D., Cheng, S., Metzker, L., and Voigt, S. (2023b). Integrating factor models. *The Journal of Finance*, 78(3):1593–1646.
- Bañbura, M., Giannone, D., and Lenza, M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of forecasting*, 31(3):739–756.
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1):3–18.
- Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, 32(3):870–897.
- Barillas, F. and Shanken, J. (2017). Which alpha? *The Review of Financial Studies*, 30(4):1316–1338.
- Bessembinder, H. and Venkataraman, K. (2010). Bid-ask spreads: Measuring trade execution costs in financial markets. *Encyclopedia of quantitative finance*, pages 184–190.
- Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9):3411–3447.
- Brennan, T. J. and Lo, A. W. (2010). Impossible frontiers. *Management Science*, 56(6):905–923.
- Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671.
- Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *The review of financial studies*, 22(6):2201–2238.
- Bryzgalova, S., Huang, J., and Julliard, C. (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *The Journal of Finance*, 78(1):487–557.
- Busse, J. A. and Irvine, P. J. (2006). Bayesian alphas and mutual fund persistence. *The Journal of Finance*, 61(5):2251–2288.
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2009). Handling sparsity via the horseshoe. In *Artificial intelligence and statistics*, pages 73–80. PMLR.
- Chen, A. Y. and Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review*, *Forthcoming*.
- Chernozhukov, V., Hansen, C., and Liao, Y. (2017). A lava attack on the recovery of sums of dense and sparse signals. *The Annals of Statistics*, pages 39–76.
- Chib, S., Zeng, X., and Zhao, L. (2020). On comparing asset pricing models. *The Journal of Finance*, 75(1):551–577.
- Chinco, A., Clark-Joseph, A. D., and Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492.

- Chordia, T., Subrahmanyam, A., and Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, 58(1):41–58.
- Corwin, S. A. and Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance*, 67(2):719–760.
- De Mol, C., Giannone, D., and Reichlin, L. (2024). The asymptotic equivalence of ridge and principal component regression with many predictors. *Econometrics and Statistics*.
- Dechow, P. M., Hutton, A. P., Meulbroeck, L., and Sloan, R. G. (2001). Short-sellers, fundamental analysis, and stock returns. *Journal of financial Economics*, 61(1):77–106.
- Della Corte, P., Sarno, L., and Thornton, D. L. (2008). The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value. *Journal of Financial Economics*, 89(1):158–174.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the  $1/n$  portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.
- DeMiguel, V., Martin-Utrera, A., Nogales, F. J., and Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, 33(5):2180–2222.
- Diether, K. B., Malloy, C. J., and Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The journal of finance*, 57(5):2113–2141.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.
- Fama, E. F. and French, K. R. (2006). Profitability, investment and average returns. *Journal of Financial Economics*, 82(3):491–518.
- Fama, E. F. and French, K. R. (2018). Choosing factors. *Journal of financial economics*, 128(2):234–252.
- Fava, B. and Lopes, H. F. (2021). The illusion of the illusion of sparsity: An exercise in prior sensitivity. *Brazilian Journal of Probability and Statistics*, 35(4):699–720.
- Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC press, 3rd edition.
- George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica Sinica*, pages 339–373.
- Giannone, D., Lenza, M., and Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437.

- Green, J., Hand, J. R., and Zhang, X. F. (2017a). The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, 30(12):4389–4436.
- Green, J., Hand, J. R. M., and Zhang, X. F. (2017b). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies*, 30(12):4389–4436.
- Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.
- Haddad, V., Kozak, S., and Santosh, S. (2020). Factor timing. *The Review of Financial Studies*, 33(5):1980–2018.
- Harvey, C. R. and Liu, Y. (2019). Cross-sectional alpha dispersion and performance evaluation. *Journal of Financial Economics*, 134(2):273–296.
- Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.
- Hauzenberger, N., Huber, F., and Onorante, L. (2021). Combining shrinkage and sparsity in conjugate vector autoregressive models. *Journal of Applied Econometrics*, 36(3):304–327.
- Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.
- Huber, F., Koop, G., and Onorante, L. (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, 39(3):669–683.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91.
- Jensen, T. I., Kelly, B., and Pedersen, L. H. (2022). Is there a replication crisis in finance? *The Journal of Finance*.
- Jones, C. M. and Lamont, O. A. (2002). Short-sale constraints and stock returns. *Journal of Financial Economics*, 66(2-3):207–239.
- Kandel, S. and Stambaugh, R. F. (1996). On the predictability of stock returns: an asset-allocation perspective. *The Journal of Finance*, 51(2):385–424.
- Karolyi, G. A. and Van Nieuwerburgh, S. (2020). New methods for the cross-section of returns. *The Review of Financial Studies*, 33(5):1879–1890.
- Kelly, B., Malamud, S., and Zhou, K. (2024). The virtue of complexity in return prediction. *The Journal of Finance*, 79(1):459–503.
- Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.
- Kelly, B. T. and Xiu, D. (2023). Financial machine learning. Technical report, National Bureau of Economic Research.

- Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.
- Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.
- Loughran, T. and Wellman, J. W. (2011). New evidence on the relation between the enterprise multiple and average stock returns. *Journal of Financial and Quantitative Analysis*, 46(6):1629–1650.
- Makalic, E. and Schmidt, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182.
- Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612.
- Narisetty, N. N., Shen, J., and He, X. (2019). Skinny gibbs: A consistent and scalable gibbs sampler for model selection. *Journal of the American Statistical Association*.
- Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the american statistical association*, 103(482):681–686.
- Pástor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381.
- Patton, A. J. and Weller, B. M. (2020). What you see is not what you get: The costs of trading market anomalies. *Journal of Financial Economics*, 137(2):515–549.
- Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, 114(3):517–553.
- Ray, P. and Bhattacharya, A. (2018). Signal adaptive variable selector for the horseshoe prior. *arXiv preprint arXiv:1810.09004*.
- Shumway, T. and Warther, V. A. (1999). The delisting bias in crsp’s nasdaq data and its implications for the size effect. *The Journal of Finance*, 54(6):2361–2379.
- Simon, F., Weibels, S., and Zimmermann, T. (2023). Deep parametric portfolio policies. Technical report, CFR Working Paper.
- Smith, G. and Campbell, F. (1980). A critique of some ridge regression methods. *Journal of the American Statistical Association*, 75(369):74–81.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Supplementary appendix for:

## Rethinking Sparsity: Parametric Portfolios and Firm Characteristics

In this appendix, we provide a more detailed description of the posterior derivations outlined in the main text. The steps follow the algorithm proposed by [Giannone et al. \(2021\)](#) with the modifications required by the Student-t component of the spike-and-slab (e.g., [Fava and Lopes, 2021](#)). We also provide descriptive statistics for the returns on characteristic-managed portfolios used in the main empirical analysis. Finally, we also provide additional in-sample and out-of-sample empirical results.

### A Student- $t$ as a scale mixture of normals

We show that the Student- $t$  with  $\theta_j \sim \mathcal{T}_\nu(0, \sigma^2\gamma^2)$  can be derived from a scale mixture of normals  $\theta_j \sim N(0, \sigma^2\gamma^2\lambda_j^2)$  with  $\lambda_j^2 \sim \mathcal{IG}(\nu/2, \nu/2)$ . The joint distribution can be written as  $p(\theta_j, \lambda_j^2) = p(\theta_j | \lambda_j^2)p(\lambda_j^2)$  with

$$p(\theta_j | \lambda_j^2) = \frac{1}{\sqrt{2\pi\sigma^2\gamma^2\lambda_j^2}} \exp\left(-\frac{\theta_j^2}{2\sigma^2\gamma^2\lambda_j^2}\right) \quad p(\lambda_j^2) = \frac{(\frac{\nu}{2})^{\nu/2}}{\Gamma(\nu/2)} (\lambda_j^2)^{-(\nu/2+1)} \exp\left(-\frac{\nu}{2\lambda_j^2}\right)$$

The marginal distribution of  $\theta_j$  is obtained by integrating out  $\lambda_j^2$  as  $p(\theta_j) = \int_0^\infty p(\theta_j, \lambda_j^2) d\lambda_j^2 = \int_0^\infty p(\theta_j | \lambda_j^2) p(\lambda_j^2) d\lambda_j^2$ . Substituting the corresponding densities

$$\begin{aligned}
p(\theta_j) &= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2\gamma^2\lambda_j^2}} \exp\left(-\frac{\theta_j^2}{2\sigma^2\gamma^2\lambda_j^2}\right) \cdot \frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\Gamma(\nu/2)} (\lambda_j^2)^{-(\nu/2+1)} \exp\left(-\frac{\nu}{2\lambda_j^2}\right) d\lambda_j^2 \\
&= \frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\sqrt{2\pi\sigma^2\gamma^2}\Gamma(\nu/2)} \cdot \underbrace{\int_0^\infty (\lambda_j^2)^{-(\nu/2+1/2)} \exp\left(-\left(\frac{\theta_j^2}{2\sigma^2\gamma^2} + \frac{\nu}{2}\right) \frac{1}{\lambda_j^2}\right) d\lambda_j^2}_{\text{this simplifies to } \frac{\Gamma(\frac{\nu+1}{2})}{\left(\frac{\theta_j^2}{2\sigma^2\gamma^2} + \frac{\nu}{2}\right)^{(\nu+1)/2}}} \\
&= \frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\sqrt{2\pi\sigma^2\gamma^2}\Gamma(\nu/2)} \cdot \frac{\Gamma(\frac{\nu+1}{2})}{\left(\frac{\theta_j^2}{2\sigma^2\gamma^2} + \frac{\nu}{2}\right)^{(\nu+1)/2}} \tag{A.1}
\end{aligned}$$

Simplifying the constant in Eq.(A.1) we obtain

$$p(\theta_j) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi\nu\sigma^2\gamma^2}} \left(1 + \frac{\theta_j^2}{\nu\sigma^2\gamma^2}\right)^{-(\nu+1)/2} \tag{A.2}$$

which is the probability density function of the Student-t distribution with mean zero, variance  $\sigma^2\gamma^2$ , and  $\nu$  degrees of freedom.

## B Gibbs sampler

Let us define  $F$  as a  $T \times k$  matrix of returns on the  $k$  characteristic-managed portfolios and  $\theta$  the corresponding  $k$ -dimensional vector of regression coefficients. It is useful to rewrite the model in terms of a set of latent variables  $z = [z_1, \dots, z_k]$  which takes value 1 if the corresponding managed portfolio is included in the parametric portfolio and 0 otherwise. Given the prior structure outlined in the main text, the joint posterior of the model parameters

takes the form

$$\begin{aligned}
p(\theta, \sigma^2, R^2, z, q, \lambda \mid Y, F) &\propto p(Y \mid F, \theta, \sigma^2) \cdot p(\theta \mid \sigma^2, \lambda, z) \cdot p(z \mid q) \cdot p(\lambda) \cdot p(\sigma^2) \cdot p(R^2, q) \\
&\propto \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{T}{2}} \exp \left( -\frac{1}{2\sigma^2} (Y - F\theta)^\top (Y - F\theta) \right) \\
&\quad \cdot \prod_{i=1}^k \left( \frac{1}{\sqrt{2\pi\sigma^2\gamma^2\lambda_i}} \exp \left( -\frac{\theta_i^2}{2\sigma^2\gamma^2\lambda_i} \right) \right)^{z_i} (\delta(\theta_i))^{1-z_i} \\
&\quad \cdot \prod_{i=1}^k \lambda_i^{-\frac{\nu}{2}-1} \exp \left( -\frac{\nu}{2\lambda_i} \right) \cdot q^{s(z)} (1-q)^{k-s(z)} \\
&\quad \cdot q^{a-1} (1-q)^{b-1} \cdot (R^2)^{A-1} (1-R^2)^{B-1} \cdot \frac{1}{\sigma^2}
\end{aligned}$$

where  $\gamma^2 = \frac{1}{k\bar{\nu}_x q} \cdot \frac{R^2}{1-R^2}$ ,  $\delta(\cdot)$  is a Dirac-delta function, and  $s(z) = \sum_{i=1}^k z_i$ . As customary in the Bayesian literature, we can sample from the joint posterior distribution by using a Gibbs sampling algorithm. We can sample from the joint posterior using a Gibbs sampling algorithm with blocks (i)  $\theta$ , (ii)  $\sigma^2$ , (iii)  $\lambda_i^2$ , (iv)  $z$ , and (v)  $R^2, q$ . In the following, we specify the conditional posterior for each sampler block.

## B.1 Posterior distribution of $\theta$

To simplify the derivation denote with  $\theta_z, \lambda_z$  the non-zero components of  $\theta, \lambda$ , i.e.,  $\theta_z = (\theta_j : z_j = 1)$  and  $\lambda_z^2 = (\lambda_j^2 : z_j = 1)$  and with  $F_z$  the selected managed portfolios. The hierarchical prior for  $\theta_z$  can be written as,

$$p(\theta_z \mid \sigma^2, \gamma^2, \lambda_z^2) \propto \exp \left( -\frac{1}{2\sigma^2} \theta_z^\top D_z^{-1} \theta_z \right)$$

where  $D_z = \gamma^2 \text{diag}(\lambda_z^2)$  is a diagonal matrix. The conditional posterior of  $\theta_z$  takes the form,

$$\begin{aligned}
p(\theta_z \mid Y, \sigma^2, \gamma^2, \lambda_z^2) &\propto p(Y \mid \theta_z, \sigma^2) p(\theta_z \mid \sigma^2, \gamma^2, \lambda_z^2) \\
&\propto (2\pi\sigma^2)^{-\frac{T+s(z)}{2}} |D_z|^{-\frac{1}{2}} \exp \left( -\frac{1}{2\sigma^2} \left[ (Y - F_z\theta_z)^\top (Y - F_z\theta_z) + \theta_z^\top D_z^{-1} \theta_z \right] \right) \\
&\propto \exp \left( -\frac{1}{2\sigma^2} \left[ Y^\top Y - 2Y^\top F_z\theta_z + \theta_z^\top (F_z^\top F_z + D_z^{-1}) \theta_z \right] \right)
\end{aligned}$$

This is recognised as the kernel of a multivariate normal distribution

$$\theta_z \mid \text{rest} \sim N(\mu_{\theta_z}, \sigma^2 \Sigma_{\theta_z}^{-1}) \quad (\text{B.1})$$

where  $\mu_{\theta_z} = \Sigma_{\theta_z}^{-1} F_z^\top Y$  and  $\Sigma_{\theta_z} = (F_z^\top F_z + D_z^{-1})$ . We note that the posterior  $p(\theta_z^2 \mid \text{rest})$  can be further simplified by marginalising over  $\lambda_z$  as  $D_z = \gamma^2 \mathbb{E}[\text{diag}(\lambda_z)]$ , such that  $D_z = \gamma^2 \frac{\nu}{\nu-2} I_{s(z)}$ . As a result, the marginal distribution of  $\theta_z$  is a multivariate normal of the form

$$\theta_z \mid \text{rest} \sim N(\mu_{\theta_z}, \sigma^2 \Sigma_{\theta_z}^{-1}) \quad (\text{B.2})$$

where

$$\mu_{\theta_z} = \left( F_z^\top F_z + \frac{1}{\gamma^2} \frac{\nu-2}{\nu} I_{s(z)} \right)^{-1} F_z^\top Y \quad (\text{B.3})$$

and

$$\Sigma_{\theta_z} = F_z^\top F_z + \frac{1}{\gamma^2} \frac{\nu-2}{\nu} I_{s(z)} \quad (\text{B.4})$$

## B.2 Posterior distribution of $\lambda_j^2$

The prior for  $\lambda_j^2$  is:

$$p(\lambda_j^2) \propto (\lambda_j^2)^{-\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2\lambda_j^2}\right)$$

whereas the conditional prior of  $\theta_j$  given  $\lambda_j^2$  is:

$$p(\theta_j \mid \sigma^2, \gamma^2, \lambda_j^2) \propto \frac{1}{\sqrt{\sigma^2 \gamma^2 \lambda_j^2}} \exp\left(-\frac{\theta_j^2}{2\sigma^2 \gamma^2 \lambda_j^2}\right)$$

The posterior distribution of  $\lambda_j^2$  is proportional to the product of the likelihood and the prior  $p(\lambda_j^2 \mid \theta_j, \sigma^2, \gamma^2) \propto p(\theta_j \mid \sigma^2, \gamma^2, \lambda_j^2) p(\lambda_j^2)$  such that,

$$p(\lambda_j^2 \mid \theta_j, \sigma^2, \gamma^2) \propto (\lambda_j^2)^{-\frac{\nu+1}{2}-1} \exp\left(-\frac{1}{2\lambda_j^2} \left(\frac{\theta_j^2}{\sigma^2 \gamma^2} + \nu\right)\right)$$

This is recognised as the kernel of an inverse-gamma distribution, such that

$$\lambda_j^2 \mid \text{rest} \sim \mathcal{IG}\left(\frac{\nu+1}{2}, \frac{\theta_j^2}{2\sigma^2 \gamma^2} + \frac{\nu}{2}\right) \quad (\text{B.5})$$

where  $\alpha = \frac{\nu+1}{2}$  and  $\beta = \frac{\theta_j^2}{2\sigma^2\gamma^2} + \frac{\nu}{2}$  are the shape and scale parameters, respectively.

### B.3 Posterior distribution of $\sigma^2$

To derive the full conditional distribution of  $\sigma^2$  we consider the joint posterior distribution  $p(\sigma^2, \theta_z | \text{rest}) \propto p(Y | F_z, \theta_z, \sigma^2) \cdot p(\sigma^2, \theta_z | \lambda_z, z) \cdot p(\sigma^2)$  and then marginalising over  $\theta_z$ . Given the Jeffreys' prior for  $\sigma^2$ , i.e.,  $p(\sigma^2) \propto \frac{1}{\sigma^2}$ , we obtain

$$p(\sigma^2, \theta_z | \lambda_z, z) = p(\theta_z | \sigma^2, \lambda_z, z) \cdot p(\sigma^2) \propto (\sigma^2)^{-\left(\frac{s(z)}{2}+1\right)} \exp\left(-\frac{1}{2\sigma^2}\theta_z^\top D_z^{-1}\theta_z\right)$$

Substituting the expressions for the likelihood and the joint prior, we get,

$$\begin{aligned} p(\sigma^2 | \text{rest}) &= \int p(Y | \theta_z, \sigma^2) p(\sigma^2, \theta_z | \lambda_z, z) d\theta_z \\ &\propto (\sigma^2)^{-\frac{T+s(z)}{2}+1} \int \exp\left(-\frac{1}{2\sigma^2} [(Y - F_z\theta_z)^\top (Y - F_z\theta_z) + \theta_z^\top D_z^{-1}\theta_z]\right) d\theta_z \\ &\propto (\sigma^2)^{-\frac{T+s(z)}{2}+1} \exp\left(-\frac{1}{2\sigma^2} S\right) \end{aligned}$$

where  $S = Y^\top Y - Y^\top F_z (F_z^\top F_z + D_z^{-1})^{-1} F_z^\top Y$  the generalised residual sum of squares. We note that the posterior  $p(\sigma^2 | \text{rest})$  can be further simplified by marginalising over  $\lambda_z$  as  $D_z = \gamma^2 \mathbb{E}[\text{diag}(\lambda_z)]$ , such that  $D_z = \gamma^2 \frac{\nu}{\nu-2} I_{s(z)}$ . As a result, the marginal posterior for  $\sigma^2$  is recognised as an Inverse-Gamma distribution with shape  $\frac{T+s(z)}{2}$  and scale  $S_\nu$  parameters, i.e.,

$$\sigma^2 | \text{rest} \sim \mathcal{IG}\left(\frac{T + s(z)}{2}, \frac{S_\nu}{2}\right) \quad (\text{B.6})$$

where  $S_\nu = Y^\top Y - Y^\top F_z (F_z^\top F_z + \frac{\nu-2}{\nu}\gamma^{-2}I_{s(z)})^{-1} F_z^\top Y$ .

### B.4 Posterior distribution of $z_i$

To derive the full conditional distribution of  $z_i$  we begin by defining the joint posterior  $p(\theta_z, \sigma^2, z | \text{rest}) \propto p(Y | F_z, \theta_z, \sigma^2) \cdot p(\sigma^2, \theta_z, z)$ . This is defined by multiplying the joint

prior and the likelihood as

$$p(\theta_z, \sigma^2, z \mid \text{rest}) \propto q^{s(z)}(1-q)^{k-s(z)} (2\pi\sigma^2)^{-\frac{T+s(z)}{2}+1} \prod_{i=1}^{s(z)} (\gamma^2 \lambda_i^2)^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2} (Y - F_z \theta_z)^\top (Y - F_z \theta_z) + \sum_{i=1}^{s(z)} \frac{\theta_{z,i}^2}{\gamma^2 \lambda_i^2}\right) \quad (\text{B.7})$$

This can be simplified by marginalising over  $\lambda_i^2$  as

$$p(\theta_z, \sigma^2, z \mid \text{rest}) \propto q^{s(z)}(1-q)^{k-s(z)} (2\pi\sigma^2)^{-\frac{T+s(z)}{2}+1} \left(\frac{\nu}{\nu-2}\gamma^2\right)^{-\frac{s(z)}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2} (Y - F_z \theta_z)^\top (Y - F_z \theta_z) + \frac{\theta_z^\top \theta_z}{\gamma^2} \cdot \frac{\nu-2}{\nu}\right) \quad (\text{B.8})$$

To obtain the marginal posterior for  $z$  we first integrate out  $\theta_z$  as

$$\begin{aligned} p(\sigma^2, z \mid \text{rest}) &= \int p(\theta_z, \sigma^2, z \mid \text{rest}) d\theta_z \\ &= q^{s(z)}(1-q)^{k-s(z)} (2\pi\sigma^2)^{-\frac{T+s(z)}{2}+1} \left(\frac{\nu}{\nu-2}\gamma^2\right)^{-\frac{s(z)}{2}} \\ &\quad \int \cdot \exp\left(-\frac{1}{2\sigma^2} (Y - F_z \theta_z)^\top (Y - F_z \theta_z) + \frac{\theta_z^\top \theta_z}{\gamma^2} \cdot \frac{\nu-2}{\nu}\right) d\theta_z \\ &= q^{s(z)}(1-q)^{k-s(z)} (2\pi\sigma^2)^{-\frac{T+s(z)}{2}+1} \left(\frac{\nu}{\nu-2}\gamma^2\right)^{-\frac{s(z)}{2}} |\Sigma_\nu|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} S_\nu\right) \end{aligned}$$

where  $S_\nu = Y^\top Y - Y^\top F_z (F_z^\top F_z + \frac{\nu-2}{\nu}\gamma^{-2}I_{s(z)})^{-1} F_z^\top Y$  and  $\Sigma_\nu = (F_z^\top F_z + \frac{\nu-2}{\nu}\frac{1}{\gamma^2}I_{s(z)})$ . Then, we integrate out  $\sigma^2$  as

$$\begin{aligned} p(z \mid \text{rest}) &= \int p(\sigma^2, z \mid \text{rest}) d\sigma^2 \\ &= q^{s(z)}(1-q)^{k-s(z)} |\Sigma_\nu|^{-\frac{1}{2}} S_\nu^{-\frac{T}{2}} \left(\frac{\nu}{\nu-2}\gamma^2\right)^{-\frac{s(z)}{2}} \quad (\text{B.9}) \end{aligned}$$

Therefore, we can use the same strategy as in [Giannone et al. \(2021\)](#) and draw iteratively  $z_i \mid \text{rest}, z_{-i}$  for  $i = 1, \dots, k$  using a nested Gibbs sampler.

## B.5 Posterior of $R^2$ and $q$

We follow [Giannone et al. \(2021\)](#) and derive the joint posterior distribution for  $R^2$  and  $q$  as

$$p(R^2, q | \text{rest}) = p(\theta | \sigma^2, \lambda, z) \cdot p(z | q) \cdot p(R^2, q) \quad (\text{B.10})$$

This can be derived similarly to [Giannone et al. \(2021\)](#) by marginalising over  $\lambda_i^2$  such that

$$\begin{aligned} p(R^2, q | \text{rest}) &\propto \left( \frac{\nu}{\nu-2} \gamma^2 \right)^{-\frac{s(z)}{2}} \exp \left( -\frac{\nu-2}{2\sigma^2 \gamma^2 \nu} \theta^\top \text{diag}(z) \theta \right) \\ &\quad \cdot q^{s(z)} (1-q)^{k-s(z)} q^{a-1} (1-q)^{b-1} (R^2)^{A-1} (1-R^2)^{B-1} \end{aligned} \quad (\text{B.11})$$

We can substitute the definition  $\gamma^2 = \frac{1}{k\bar{v}_x q} \cdot \frac{R^2}{1-R^2}$  such that

$$\begin{aligned} p(R^2, q | \text{rest}) &\propto \left( \frac{\nu}{\nu-2} \frac{1}{k\bar{v}_x} \right)^{-\frac{s(z)}{2}} q^{\frac{s(z)}{2}} (R^2)^{-\frac{s(z)}{2}} (1-R^2)^{\frac{s(z)}{2}} q^{s(z)} (1-q)^{k-s(z)} q^{a-1} (1-q)^{b-1} \\ &\quad \cdot \exp \left( -\frac{1}{2\sigma^2} \frac{\nu-2}{2} \frac{k\bar{v}_x q (1-R^2)}{R^2} \theta^\top \text{diag}(z) \theta \right) (R^2)^{A-1} (1-R^2)^{B-1} \\ &\propto \left( \frac{\nu}{\nu-2} \frac{1}{k\bar{v}_x} \right)^{-\frac{s(z)}{2}} \exp \left( -\frac{1}{2\sigma^2} \frac{\nu-2}{\nu} \frac{k\bar{v}_x q (1-R^2)}{R^2} \theta^\top \text{diag}(z) \theta \right) \\ &\quad \cdot q^{s(z)+a-1+\frac{s(z)}{2}} (1-q)^{k-s(z)+b-1} (R^2)^{-\frac{s(z)}{2}+A-1} (1-R^2)^{\frac{s(z)}{2}+B-1} \end{aligned} \quad (\text{B.12})$$

We follow [Giannone et al. \(2021\)](#) and sample from a discrete approximation of this distribution. More specifically, we discretize the support of  $R^2$  and  $q$  by interlacing two grids defined over the unit interval, each with 0.01 increments and finer 0.001 increments near the boundaries.

## B.6 A possible prior for $\nu$

Assuming a prior  $\nu \sim \text{Gamma}(\alpha, \beta)$  and given the likelihood  $\lambda_i^2 | \nu \sim \text{IG}(\frac{\nu}{2}, \frac{\nu}{2})$ , the conditional posterior of  $\nu$  can be derived as:

$$p(\nu | \lambda_i^2, \dots) \propto \nu^{\alpha-1} e^{-\beta\nu} \prod_{i=1}^k (\lambda_i^2)^{-\left(\frac{\nu}{2}+1\right)} \exp \left( -\frac{\nu}{2\lambda_i^2} \right)$$

This can be sampled using a Metropolis-Hastings step:

1. Propose a new value  $\nu_{\text{new}}$  using a proposal distribution (e.g., Gaussian) centered at  $\nu_{\text{current}}$ .
2. Calculate the acceptance ratio:

$$\alpha = \min \left( 1, \frac{p(\nu_{\text{new}} | \lambda_i^2, \dots)}{p(\nu_{\text{current}} | \lambda_i^2, \dots)} \cdot \frac{q(\nu_{\text{current}} | \nu_{\text{new}})}{q(\nu_{\text{new}} | \nu_{\text{current}})} \right)$$

3. Accept  $\nu_{\text{new}}$  with probability  $\alpha$ , otherwise retain  $\nu_{\text{current}}$ .

# C Data and Descriptive Statistics

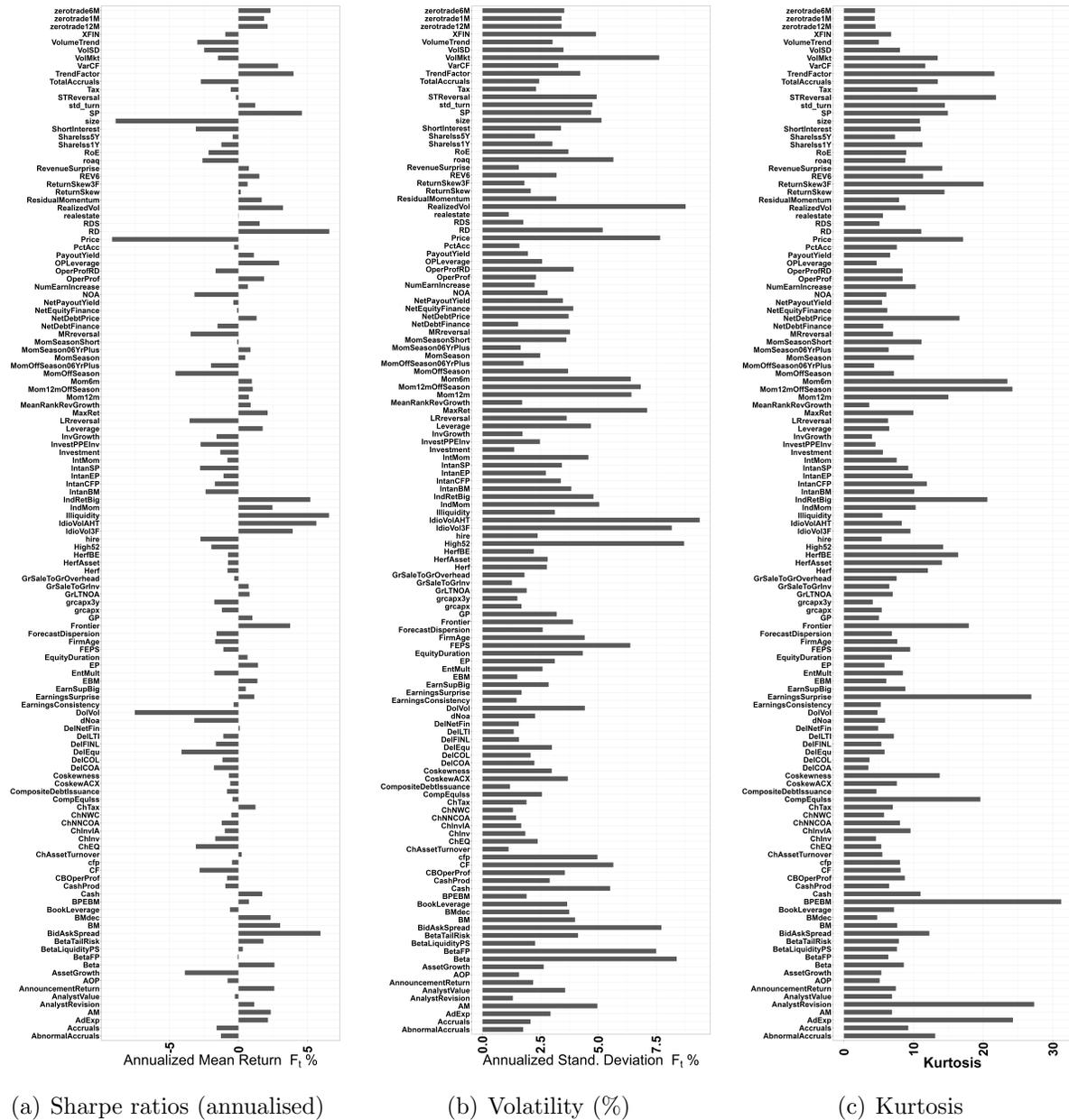


Figure C.1: **Descriptive statistics on characteristic-managed portfolios.** The figure reports the sample average (left), volatility (middle), and kurtosis (right) of the characteristic-managed portfolios considered in the empirical application. The sample period is from January 1980 to December 2023.

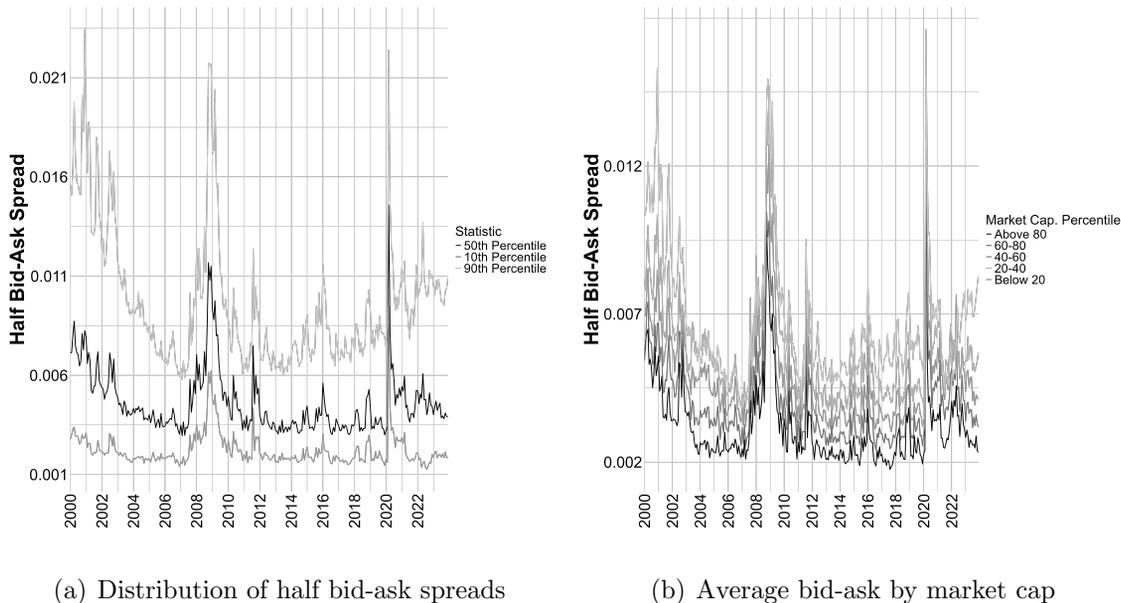


Figure C.2: **Half bid-ask spreads distribution.** The left panel reports the 10th, 50th, and 90th percentile of the cross-sectional distribution of the effective bid-ask spread (in decimals) over the out-of-sample period. The right panel reports the average bid-ask spread (in decimals) for stocks sorted by market capitalisation (quintile sort). The effective bid-ask spread is calculated based on the [Corwin and Schultz \(2012\)](#) divided by the stock price. The out-of-sample period is from January 2000 to December 2023.

## D Additional empirical results

### D.1 Full-sample estimates

Figure 2 in Section 3.1 shows that the *share* of relevant firm characteristics for the cross-section of stock returns changes with the degrees of freedom  $\nu$ ; that is, the degree of average sparsity, as proxied by the posterior density of  $q$ , is smaller for smaller values of  $\nu$ . We now investigate whether the *identity* of these firm characteristics can be recovered for different levels of  $\nu$ . To this end, we calculate the posterior inclusion probability of each firm characteristic in the optimal portfolio choice. Figure D.3 shows the results. Each horizontal stripe corresponds to a firm characteristic, and darker shades denote higher inclusion probabilities.<sup>19</sup> The estimates

<sup>19</sup>Note that the probability of inclusion of a single predictor may differ from  $q$ . The latter can be considered the average probability of inclusion across firm characteristics. As such, it should not coincide with the inclusion probability of a single characteristic.

obtained from different prior specifications are labeled by column.

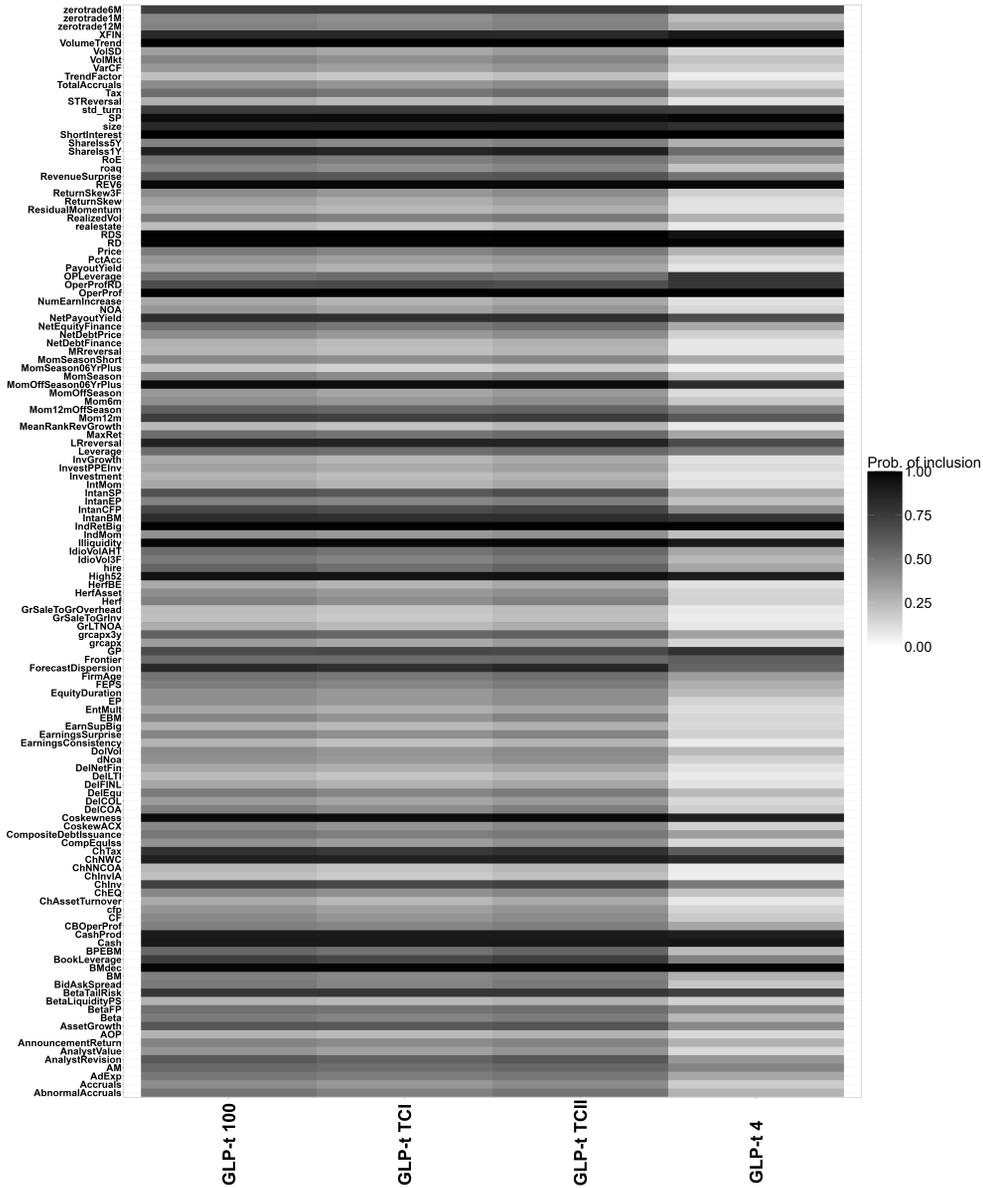


Figure D.3: **In-sample posterior inclusion probabilities.** The figure shows the in-sample posterior probability of inclusion for each characteristic in the cross-section of stock returns for different levels of  $\nu = [4, TC1, TC2, 100]$ . The sample period is from January 1980 to December 2023.

The results show that a heavy-tailed prior does not alter the selection of strong characteristics but reduces the inclusion probability of those weakly associated with future stock returns, thereby reducing the uncertainty about which firm characteristics should be included in the optimal portfolio policy. In this respect, when applying a conventional 50% cutoff for

the selection of a given characteristic (e.g., [Barbieri and Berger, 2004](#)), the identification of strong characteristics does not depend on the prior, but the identification of weaker predictors does. [Table D.1](#) illustrates this point.

When applying a simple threshold of 0.5 (or 50%) to each posterior inclusion probability, some key characteristics—such as liquidity ([Amihud, 2002](#)), operating profitability ([Fama and French, 2006](#)), short interest ([Dechow et al., 2001](#)), analysts’ forecast dispersion ([Diether et al., 2002](#)), 12-month momentum ([Jegadeesh and Titman, 1993](#)), enterprise multiple ([Loughran and Wellman, 2011](#)), co-skewness ([Ang et al., 2006](#)), size ([Banz, 1981](#)), among others—always emerge as significant for the cross-section of stock returns irrespective of the prior tails. However, the number of selected characteristics increases as the value of  $\nu$  increases.

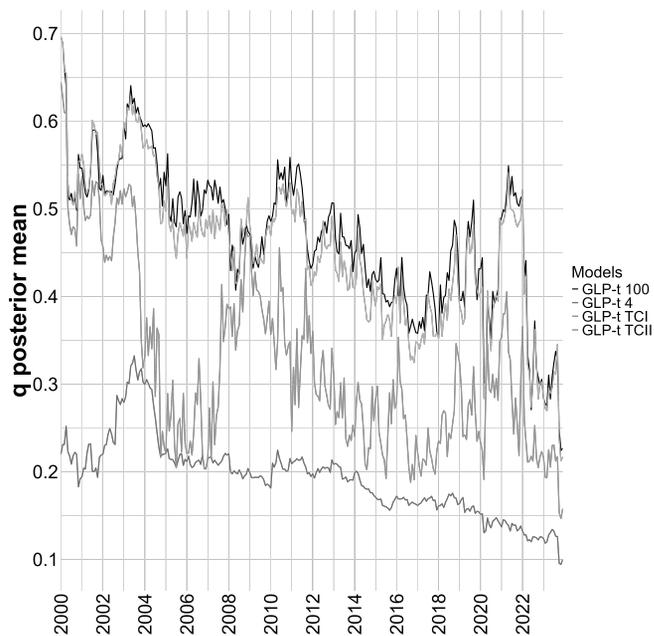
## D.2 Recursive estimates

In the main text we report the recursive posterior estimates of some of the key parameters for  $\nu = [4, 10, 30, 100]$ . In this Section, we report additional recursive estimates for  $\nu = [4, \text{TC1}, \text{TC2}, 100]$ . [Figure D.4](#) reports the recursive posterior mean of  $q$  and  $\gamma^2$ . [Figure D.5](#) reports the recursive estimates of the HHI and the weights range when the prior tails are calibrated as  $\nu = [4, \text{TC1}, \text{TC2}, 100]$ .

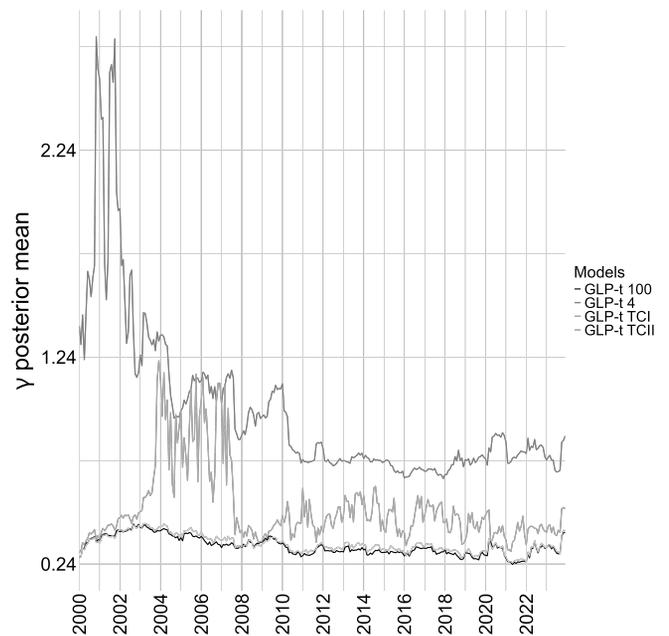
## D.3 Low vs high volatility state

GLP-t( $\nu$ )					
4	10	30	100	TC1	TC2
size	size	size	size	size	size
			AbnormalAccruals		AbnormalAccruals
	AM	AM	AM	AM	AM
		AnalystRevision	AnalystRevision	AnalystRevision	AnalystRevision
	AssetGrowth	AssetGrowth	AssetGrowth	AssetGrowth	AssetGrowth
BetaTailRisk	BetaTailRisk	BetaTailRisk	BetaTailRisk	BetaTailRisk	BetaTailRisk
BMdec	BMdec	BMdec	BMdec	BMdec	BMdec
	BookLeverage	BookLeverage	BookLeverage	BookLeverage	BookLeverage
		BPEBM	BPEBM	BPEBM	BPEBM
Cash	Cash	Cash	Cash	Cash	Cash
CashProd	CashProd	CashProd	CashProd	CashProd	CashProd
	ChInv	ChInv	ChInv	ChInv	ChInv
ChNWC	ChNWC	ChNWC	ChNWC	ChNWC	ChNWC
ChTax	ChTax	ChTax	ChTax	ChTax	ChTax
ForecastDispersion	ForecastDispersion	ForecastDispersion	ForecastDispersion	ForecastDispersion	ForecastDispersion
Frontier	Frontier	Frontier	Frontier	Frontier	Frontier
GP	GP	GP	GP	GP	GP
		grcapx3y	grcapx3y	grcapx3y	grcapx3y
		hire	hire	hire	hire
		IdioVolAHT	IdioVolAHT	IdioVolAHT	IdioVolAHT
Illiquidity	Illiquidity	Illiquidity	Illiquidity	Illiquidity	Illiquidity
IndRetBig	IndRetBig	IndRetBig	IndRetBig	IndRetBig	IndRetBig
IntanBM	IntanBM	IntanBM	IntanBM	IntanBM	IntanBM
	IntanCFP	IntanCFP	IntanCFP	IntanCFP	IntanCFP
	IntanSP	IntanSP	IntanSP	IntanSP	IntanSP
	Leverage	Leverage	Leverage	Leverage	Leverage
LRreversal	LRreversal	LRreversal	LRreversal	LRreversal	LRreversal
		MaxRet	MaxRet	MaxRet	MaxRet
Mom12m	Mom12m	Mom12m	Mom12m	Mom12m	Mom12m
MomOffSeason06YrPlus	Mom12mOffSeason	Mom12mOffSeason	Mom12mOffSeason	Mom12mOffSeason	Mom12mOffSeason
	MomOffSeason06YrPlus	MomOffSeason06YrPlus	MomOffSeason06YrPlus	MomOffSeason06YrPlus	MomOffSeason06YrPlus
		NetEquityFinance	NetEquityFinance	NetEquityFinance	NetEquityFinance
NetPayoutYield	NetPayoutYield	NetPayoutYield	NetPayoutYield	NetPayoutYield	NetPayoutYield
OPLeverage	OPLeverage	OPLeverage	OPLeverage	OPLeverage	OPLeverage
RD	RD	RD	RD	RD	RD
RDS	RDS	RDS	RDS	RDS	RDS
REV6	REV6	REV6	REV6	REV6	REV6
RevenueSurprise	RevenueSurprise	RevenueSurprise	RevenueSurprise	RevenueSurprise	RevenueSurprise
ShareIss1Y	ShareIss1Y	ShareIss1Y	ShareIss1Y	ShareIss1Y	ShareIss1Y
std_turn	std_turn	std_turn	std_turn	std_turn	std_turn
		Tax	Tax	Tax	Tax
XFIN	XFIN	XFIN	XFIN	XFIN	XFIN
zerotrade6M	zerotrade6M	zerotrade6M	zerotrade6M	zerotrade6M	zerotrade6M
		BetaFP	BetaFP	BetaFP	BetaFP
Coskewness	Coskewness	Coskewness	Coskewness	Coskewness	Coskewness
		FirmAge	FirmAge	FirmAge	FirmAge
High52	High52	High52	High52	High52	High52
	OperProf	OperProf	OperProf	OperProf	OperProf
	OperProfRD	OperProfRD	OperProfRD	OperProfRD	OperProfRD
		RoE	RoE	RoE	RoE
	ShortInterest	ShortInterest	ShortInterest	ShortInterest	ShortInterest
	SP	SP	SP	SP	SP
VolumeTrend	VolumeTrend	VolumeTrend	VolumeTrend	VolumeTrend	VolumeTrend

Table D.1: **Characteristics selected for different prior thickness.** The table reports the selected variables based on a 50% cutoff of the posterior inclusion probability for different levels of  $\nu = [4, 10, 30, 100, TC1, TC2]$ . The sample period is from January 1980 to December 2023.



(a) Posterior mean of  $q$



(b) Posterior mean of  $\sqrt{\gamma^2}$

Figure D.4: **Recursive estimates of  $q$  and  $\gamma^2$ .** The figure shows the posterior mean of  $q$  (left panel) and  $\sqrt{\gamma^2}$  (right panel) for different levels of  $\nu = [4, TC1, TC2, 100]$ . The estimates are based on a rolling window of 240 months. The out-of-sample period is from January 2000 to December 2023.

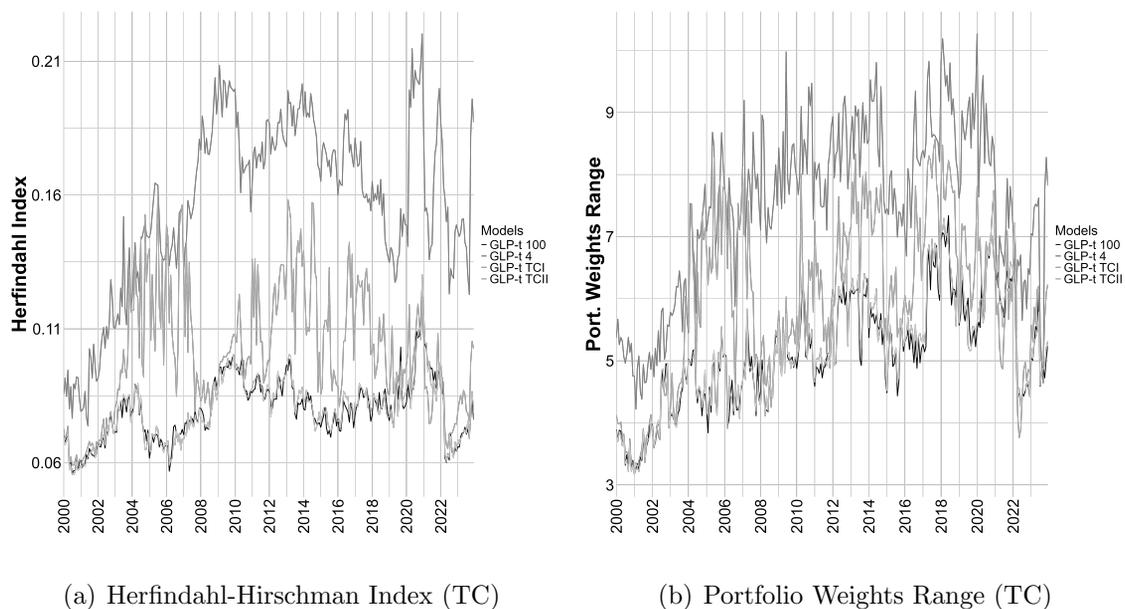


Figure D.5: **Portfolio diversification and leverage.** The figure shows the Herfindahl-Hirschman Index (HHI) (left panel) and the weights range (right panel) obtained for the recursive parametric portfolio allocation for different prior specifications  $\nu = [4, 100, \text{TC1}, \text{TC2}]$ . The value of the HHI is rescaled in the interval  $[0, 1]$ . The optimal allocation is based on a rolling window of 240 months. The sample period is from January 2000 to December 2023.

**Panel A: High volatility**

	EW	BSV	GLP- $t(\nu)$					
			4	10	30	100	TC1	TC2
Mean	0.007	0.063	0.078	0.079	0.075	0.067	0.073	0.069
Std	0.077	0.145	0.185	0.158	0.130	0.121	0.129	0.121
ES(5%)	-0.142	-0.203	-0.244	-0.234	-0.168	-0.182	-0.179	-0.168
SR (annual)	0.247	1.475	1.437	1.699	1.974	1.887	1.915	1.950
$pval(\Delta SR)_{boot}$		0.005	0.007	0.003	0.000	0.000	0.001	0.000
$\Delta CER$		0.009	-0.014	0.009	0.028	0.028	0.026	0.030
Performance Fee		0.012	-	0.014	0.038	0.036	0.036	0.038

**Panel B: Low volatility**

Mean	0.020	0.043	0.062	0.054	0.046	0.047	0.053	0.046
Std	0.040	0.086	0.110	0.091	0.074	0.066	0.094	0.068
ES(5%)	-0.061	-0.106	-0.141	-0.109	-0.085	-0.062	-0.119	-0.065
SR (annual)	1.621	1.663	1.926	2.003	2.094	2.389	1.899	2.262
$pval(\Delta SR)_{boot}$		0.812	0.444	0.360	0.290	0.086	0.509	0.142
$\Delta CER$		0.005	0.008	0.012	0.014	0.018	0.010	0.016
Performance Fee		0.006	0.011	0.014	0.015	0.019	0.012	0.017

Table D.2: **Out-of-sample portfolios based volatility states.** This table reports the out-of-sample performance net of transaction costs during periods of high volatility (Panel A) vs low volatility (Panel B). Transaction costs are proxied by the stock-specific half bid-ask spread and are imputed each month based on the portfolio rebalancing (turnover) as in [DeMiguel et al. \(2009\)](#). The out-of-sample period is from January 2000 to December 2023.

## E Alternative priors

**Bayesian Lasso.** The Bayesian Lasso model was based on [Park and Casella \(2008\)](#). The priors take the following form:

$$\begin{aligned}\theta \mid \sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim N(0, \sigma^2 D), \\ \tau_j^2 \mid \lambda^2 &\sim \text{Exp}\left(\frac{\lambda^2}{2}\right), \\ \lambda^2 &\sim G(a, b)\end{aligned}$$

with  $p(\sigma^2) \propto \frac{1}{\sigma^2}$ ,  $D = \text{diag}(\tau_1^2, \dots, \tau_p^2)$ ,  $a = 1$ , and  $b = 2$ . Considering the hierarchical prior structure proposed by [Park and Casella \(2008\)](#), the conditional posteriors can be derived as:

$$\begin{aligned}\theta \mid \sigma^2, \tau_1^2, \dots, \tau_p^2, y &\sim N(A^{-1}F^\top y, \sigma^2 A^{-1}), \\ \tau_j^2 \mid \theta_j, y &\sim IG\left(\sqrt{\frac{\lambda^2 \sigma^2}{\theta_j^2}}, \lambda^2\right)\end{aligned}$$

and

$$\begin{aligned}\lambda^2 \mid \tau_1^2, \dots, \tau_p^2 &\sim G\left(a + p, b + \frac{\sum_{j=1}^p \tau_j^2}{2}\right), \\ \sigma^2 \mid \theta, y &\sim IG\left(\frac{s_0 + p}{2}, \frac{s + \theta^\top D^{-1} \theta}{2}\right)\end{aligned}$$

where  $A = (F^\top F + D^{-1})^{-1}$ ,  $s = (y - F\theta)^\top (y - F\theta)$ , and  $D = \text{diag}(\tau_1^2, \dots, \tau_p^2)$ .

**Horseshoe.** We implemented the Horseshoe model under a hierarchical specification of the horseshoe prior proposed by [Makalic and Schmidt \(2015\)](#). The prior takes the following form:

$$\begin{aligned}\theta \mid \lambda_1, \dots, \lambda_p, \tau, \sigma^2 &\sim N(0, \sigma^2 \tau^2 \Lambda), \\ \lambda_j^2 \mid \nu_j &\sim IG\left(\frac{1}{2}, \frac{1}{\nu_j}\right), \\ \nu_j &\sim IG\left(\frac{1}{2}, 1\right), \quad j = 1, \dots, p, \\ \tau^2 \mid \xi &\sim IG\left(\frac{1}{2}, \frac{1}{\xi}\right), \\ \xi &\sim IG\left(\frac{1}{2}, 1\right)\end{aligned}$$

where  $\Lambda = \text{diag}(\lambda_1^2, \dots, \lambda_p^2)$  and  $p(\sigma^2) \propto \sigma^{-2}$ . The conditional posteriors can be derived as:

$$\begin{aligned}
\theta &| \dots \sim N(A^{-1}F^\top y, \sigma^2 A^{-1}), \\
\lambda_j^2 &| \dots \sim IG\left(1, \frac{1}{\nu_j} + \frac{\theta_j^2}{2\tau^2\sigma^2}\right), \\
\nu_j &| \dots \sim IG\left(1, 1 + \frac{1}{\lambda_j^2}\right), \quad j = 1, \dots, p, \\
\tau^2 &| \dots \sim IG\left(\frac{1+p}{2}, \frac{1}{\xi} + \frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\theta_j^2}{\lambda_j^2}\right), \\
\xi &| \dots \sim IG\left(1, 1 + \frac{1}{\tau^2}\right), \\
\sigma^2 &| \dots \sim IG\left(\frac{n+p}{2}, \frac{s + \theta^\top D_\lambda^{-1} \theta}{2}\right)
\end{aligned}$$

where  $A = (F^\top F + D_\lambda^{-1})^{-1}$ ,  $s = (y - F\theta)^\top (y - F\theta)$ , and  $D_\lambda = \text{diag}(\tau^2 \lambda_1^2, \dots, \tau^2 \lambda_p^2)$ .

**Stochastic Search Variable Selection** A computationally convenient approach to Bayesian variable selection is the stochastic search variable selection (SSVS) as originally proposed by [George and McCulloch \(1993, 1997\)](#),

$$\theta_j | \gamma_j \sim (1 - \gamma_j) \underbrace{N(0, \tau_0^2)}_{\text{spike}} + \gamma_j \underbrace{N(0, \tau_1^2)}_{\text{slab}}, \quad (\text{E.1})$$

where both  $\tau_0^2$  and  $\tau_1^2$  are fixed, and  $\tau_0^2 \ll \tau_1^2$ . This is a mixture of two continuous distributions, whereby for  $\tau_0^2 \rightarrow 0$  the spike becomes a Dirac at zero. Notice that for  $\tau_0^2 \neq 0$ , the spike is unable to shrink exactly  $\beta_j = 0$ , i.e.,  $\mathcal{H}_0 : \beta_j \approx 0$ .

The elicitation of  $\tau_0^2, \tau_1^2$  is critical for variable selection given the prior in Eq.(E.1). We follow [Narisetty et al. \(2019\)](#) and calibrated the value of the prior variance parameters as  $\tau_0^2 = \text{var}(y)/10 * T$  and  $\tau_1^2 = \text{var}(y) * \max(k^2/100 * T, \log(T))$ , where  $T$  is the length of the sample size and  $k$  the number of firm characteristics. They also recommend to set a prior inclusion probability  $\pi_0 \sim \text{Beta}(0.1, 0.1)$  which corresponds to a uniform prior.

Consider the prior in Eq.(E.1) with  $\gamma_j \sim \text{Beta}(c, d)$  and  $\sigma^2 \sim IG(a, b)$ , the posterior takes

the form

$$\begin{aligned}
\theta | \dots &\sim N(A^{-1}X'y, \sigma^2 A^{-1}), \quad \text{where} \quad A^{-1} = (F'F + D)^{-1}, \\
\sigma^2 | \dots &\sim IG\left(a + \frac{T+k}{2}, b + \frac{s + \theta'D^{-1}\theta}{2}\right), \\
\gamma_j | \dots &\sim \text{Ber}\left(\frac{N(\theta_j|0, \sigma^2\tau_1^2)\pi_0}{N(\theta_j|0, \sigma^2\tau_1^2)\pi_0 + N(\theta_j|0, \sigma^2\tau_0^2)(1-\pi_0)}\right), \quad j = 1, \dots, p \\
\pi_0 | \dots &\sim \text{Beta}\left(c + \sum_{j=1}^k \gamma_j, d + \sum_{j=1}^k (1 - \gamma_j)\right), \quad j = 1, \dots, k
\end{aligned}$$

where  $D$  is a diagonal matrix with elements  $(1 - \gamma_j)\tau_0^2 + \gamma_j\tau_1^2$  (see [George and McCulloch, 1993, 1997](#), for more details).

## E.1 Sparsifying the posterior estimates

The posterior estimates of  $\theta_j, j = 1, \dots, k$  under the Bayesian lasso and the horseshoe are non-sparse and thus can not provide exact differentiation between significant vs non-significant characteristics. The latter is particularly relevant since we ultimately want to assess the performance of alternative priors compared to variable selection tools such as our heavy-tailed spike-and-slab prior.

To address this issue, we build upon [Ray and Bhattacharya \(2018\)](#) and implement a Signal Adaptive Variable Selector (SAVS) algorithm to induce sparsity in  $\hat{\theta}$ , conditional on a given prior. The SAVS is a post-processing algorithm which divides signals and nulls on the basis of the point estimates of the regression coefficients (see, e.g., [Hauzenberger et al., 2021](#)). Specifically, let  $\hat{\theta}_j$  be the posterior estimate of  $\theta_j$  and  $F_j$  be the associated characteristic-managed portfolios. If  $|\hat{\theta}_j| \|F_j\|^2 \leq |\hat{\theta}_j|^{-2}$  we set  $\hat{\theta}_j = 0$ , where  $\|\cdot\|$  denotes the euclidean norm.

The SAVS post-processing to induce sparsity in the posterior estimates is threefold. First, as highlighted by [Ray and Bhattacharya \(2018\)](#), the SAVS represents an automatic procedure in which the sparsity-inducing property directly depends on the effectiveness of the shrinkage performed on  $\hat{\theta}_j$ . This refers to the precision of the posterior mean estimates; that is, the more accurate is  $\hat{\theta}_j$ , the more precise the identification of the non-negligible characteristics. Second, the SAVS is “agnostic” with respect to the shrinkage prior or estimation approach adopted, so it represents a natural tool to compare different estimation methods. Third,

it is decision-theoretically motivated as it grounds on the idea of minimizing the posterior expected loss (see, e.g., [Huber et al., 2021](#)).