Firm characteristics and the cross-section of stock returns: A tale of two tails

Daniele Bianchi[†] Pedro H. M. Venturi[†]

First draft: July 2024. This draft: July 16, 2024

Abstract

We explore the role of firm characteristics to predict the cross-section of stock returns through the lens of a flexible Bayesian variable selection prior embedded in an otherwise conventional parametric portfolio choice. The main results show that model uncertainty is pervasive, and there is little evidence in favour of sparse models. Yet, there is a trade-off between sparsity and shrinkage when maximising the portfolio's expected utility: while a heavy-tailed sparsity-inducing prior reduces uncertainty on which firm characteristics matter, it also produces strikingly less diversified portfolios with more extreme weights. As a result, when transaction costs are factored in, a dense model that allows for selecting many characteristics while shrinking their impact on the optimal portfolio choice is more adequate to capture the out-of-sample variation of stock returns.

Keywords: Firm characteristics, Parametric portfolio choice, Cross-section of stock returns, Asset pricing, Bayesian variable selection, Shrinkage.

JEL codes: G10, G11, G12, C22.

[†]School of Economics and Finance, Queen Mary University of London. E-mail: d.bianchi@qmul.ac.uk Web: whitesphd.com

[†]School of Economics and Finance, Queen Mary University of London. E-mail: p.h.moravisventuri@qmul.ac.uk

1 Introduction

Understanding the variation of stock returns and its implication for portfolio choices is a fundamental goal of empirical asset pricing. Yet, stock returns are noisy and thus notoriously difficult to predict. Economic theories provide helpful guidance in identifying which state variables matter but are often too stylised and fall short in describing all sources of information investors may rely on. This has led researchers and practitioners alike to rely on a proliferation of risk factors and firm characteristics, each one of them allegedly providing significant information to pin down the dynamic of stock returns (e.g., Harvey et al., 2016).

However, which risk factor or firm characteristics are priced in the cross-section is arguably uncertain *a priori* (e.g., Bryzgalova et al., 2023). To address this issue, the empirical practice often focuses on selecting a subset of variables with the highest explanatory power (e.g. Freyberger et al., 2020) or recognise that all available variables might be important, although the impact of some might be small (e.g., Kozak et al., 2020). Yet, choosing between a sparse and a dense model – in the sense of Chernozhukov et al. (2017) – may have important implications for asset pricing and optimal portfolio allocations. For instance, DeMiguel et al. (2020) shows that careful consideration of trading costs in a parametric portfolio choice may lead to selecting many firm characteristics to predict the cross-section of stock returns.

In this paper, we take the viewpoint of an investor who cares not only about returns but also about portfolio risk and is agnostic about whether a sparse or dense stochastic discount factor more likely characterises the variation of stock returns. To this end, we build upon the parametric portfolio approach of Brandt et al. (2009) and make Bayesian inference to assess the *joint* significance of firm characteristics to predict stock returns from a *portfolio perspective*. The latter augments the base asset space by a set of characteristic-managed portfolios so that the portfolio problem reduces to find those firm characteristics that carry meaningful information on future stock returns (e.g., Hjalmarsson and Manchev, 2012).

We conduct Bayesian inference by building upon Fava and Lopes (2021) and extending on Giannone et al. (2021). Specifically, we consider a Dirac spike-and-slab prior, whereby the impact of a given characteristic on the optimal portfolio weights can be nonzero with a certain probability q. When non-zero, such impact is drawn from a Student-t distribution with ν degrees of freedom and variance scaled by the hyperparameter γ^2. The higher γ^2, the higher the prior variance, and therefore, the less shrinkage is performed on each characteristic. The Student-t assumption implies that the smaller the ν, the stronger the sparsity-inducing property of the prior.

In sum, our approach has four key ingredients. First, it allows for sparsity in the firm characteristics that enter the parametric portfolio choice. Second, it shrinks the contribution of each characteristic towards zero as an alternative way to deal with the curse of dimensionality and avoid overfitting. Third, it treats sparsity and shrinkage separately, as they are controlled by different hyperparameters, q and γ^2. Fourth, the smaller the ν, the smaller the uncertainty on to which characteristic matters for the cross-section; that is, a heavier-tailed prior specification implies a more aggressive shrinkage of those characteristics with only a mild correlation with future stock returns while being less restrictive on those with a stronger correlation. As a result, we can address the uncertainty around key features, such as the number of characteristics relevant to capturing the variation in stock returns, their identity, and the strength of their explanatory power.

1.1 Main findings

Empirically, we consider 145 firm characteristics covering 7,675 stocks from January 1985 to November 2022. Our main contribution is threefold. First, we characterise the marginal posterior distribution of the probability of inclusion q as a function of the Student-t degrees of freedom ν. We show that a smaller ν leads to less uncertainty about which characteristics matter for the cross-section. Nevertheless, the posterior estimates of q increase for more restrictive assumptions on the tails of the non-zero coefficients in the spike-and-slab prior. For larger ν, the evidence points towards more than a handful of firm characteristics improving the investors' portfolio utility. Simple recursive posterior estimates based on rolling windows show that most firm characteristics provide a weak signal for the cross-sectional variation of stock returns. Yet, there is considerable time variation in the amount of sparsity over the set of firm characteristics.

Second, we provide evidence that the joint posterior distribution of q and γ^2 exhibits a negative correlation, which becomes steeper for smaller values of ν. The amount of sparsity (shrinkage) on the set of characteristics is larger (smaller) the smaller the value of ν. We show that this negative correlation has profound implications for optimal portfolio choices. While a heavy-tailed sparsity-inducing prior reduces uncertainty on which firm characteristics matter, it also produces more concentrated portfolios with more trading at the intensive margin. This raises questions about the value of sparsity for maximising investors' expected utility.

Our third contribution addresses these questions based on an in-sample and a recursive, real-time parametric portfolio implementation with and without considering transaction costs. The results show that a heavier-tailed prior that induces more sparsity generates substantially lower out-of-sample risk-adjusted returns than priors with more restrictive distribution tail assumptions. These results hold in particular when considering transaction costs, both in-sample and out-of-sample and for different portfolio constraints. As a result, from an economic perspective, we provide evidence that a dense model that allows for selecting many characteristics while shrinking their impact on the optimal portfolio choice is more adequate to capture the out-of-sample variation of stock returns.

Overall, the empirical evidence suggests that model uncertainty is pervasive, and ignoring it – as well as the evidence in favour of denser models when considering transaction costs – may lead to potentially misleading assumptions on the degree of sparsity needed to summarise the information content in firm characteristics and its role to understand the dynamics of stock returns. In Giannone et al. (2021), this is referred to as an "illusion of sparsity" which may not necessarily supported by the data. Our results provide an economic rationale for their intuition based on an otherwise conventional parametric portfolio choice. These findings serve as a warning against using sparse models without critical judgment when linking firm characteristics to the cross-sectional variation of stock returns, especially when transaction costs are considered (e.g., DeMiguel et al., 2020).

Note that the spike-and-slab formulation we adopted encompasses popular dimension reduction specifications such as ridge regressions (e.g. Giannone et al., 2021). This can be interpreted as a regression on the principal components of the explanatory variables, with less shrinkage on the impact of more important principal components (e.g., Marquardt, 1970; Smith and Campbell, 1980; Bańbura et al., 2015; Kelly et al., 2022). Thus, our work provides a further economic intuition for considering dimension reduction to deal with firm characteristics in the context of asset pricing models.

1.2 Closely related literature

Our work contributes to a large literature that seeks to understand the cross-section of stock returns in high dimensions, such as Hou et al. (2015); Harvey et al. (2016); Green et al. (2017); 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. These approaches focus on cross-sectional regressions where the target variable is the mean returns or risk premiums. Differently, our analysis takes a portfolio perspective, which links firm characteristics and stock returns, by targeting not only the mean of the returns but also portfolio risk. This echoes the approach proposed by DeMiguel et al. (2020). Similarly, we focus on the interplay between sparsity and transaction costs. Differently, we take a Bayesian approach, which allows us to explicitly investigate model uncertainty and the role of sparse vs dense models to understand the dynamics of stock returns.

Our empirical analysis is also linked to the stochastic discount factor (SDF) approach of Kozak et al. (2020). Under mean-variance utility, the first-order condition of the investors' optimal portfolio is akin to the associated SDF. Our results provide statistical and economic evidence to support their intuition that a dense model may be preferable to predict the cross-sectional variation in stock returns. We expand on their results by focusing on Bayesian inference tools with a keen interest in the interplay between sparsity, shrinkage, signal strength, and transaction costs for investors' expected utility.

Finally, another strand of literature we contribute relates to using Bayesian methods for empirical asset pricing. Bayesian tools have been extensively used for asset allocation (e.g., Pettenuzzo et al., 2014), model selection (e.g., Pástor and Stambaugh, 2000; Avramov, 2002; Chib et al., 2020), performance evaluation (e.g., Busse and Irvine, 2006; Harvey and Liu, 2019), and asset pricing tests (e.g. Jensen et al., 2022; Bryzgalova et al., 2023), among others. Similar to Bryzgalova et al. (2023), we elicit a "spike-and-slab" prior to addressing model uncertainty in the set of firm characteristics.

We expand on this literature by assuming a Student-t spike-and-slab prior as in Fava and Lopes (2021). The evidence shows that a heavier-tailed distribution substantially reduces the uncertainty around which and how many firm characteristics matter for the cross-section of stock returns (e.g., Fava and Lopes, 2021). However, this has detrimental economic consequences when transaction costs are considered. We show that one can mitigate this issue by calibrating the prior degrees of freedom based on observable transaction costs. This incentivises shrinkage over sparsity and, therefore, improves investor's expected utility.

2 Parametric portfolio choice

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 $\widehat{\mathbf{x}}_{i,t} = \left(\widehat{x}^1_{i,t}, \ldots, \widehat{x}^k_{i,t}\right)$ observed at time t. The investor's problem is to choose the optimal portfolio weights $\mathbf{w}_t = \left(w_{1,t}, \ldots, w_{N_t,t}\right)^\top$ to maximise the expected utility of the portfolio return $r_{p,t+1} = \mathbf{w}_t^\top r_{t+1}$. We build upon Brandt et al. (2009) and define the optimal portfolio choice as a parametric function of the form

$$\mathbf{w}_t = \mathbf{w}_t^b + \frac{1}{N_t}\widehat{\mathbf{X}}_t\boldsymbol{\theta} \tag{1}$$

where $\mathbf{w}_t^b$ is the benchmark portfolio allocation, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top$ is k-dimensional vector of coefficients to be estimated, and $\widehat{\mathbf{X}}_t$ is an $N_t \times k$ matrix of characteristics. The matrix $\widehat{\mathbf{x}}_t$ is standardised cross-sectionally to have a zero mean and unit variance across all stocks at time t. The normalisation by $1/N_t$ implies that the parametric specification can be applied to an arbitrary number of stocks.[1]

The rationale of Eq.(1) is active portfolio management relative to a passive benchmark. The term $\widehat{\mathbf{X}}_t\boldsymbol{\theta}$ represents the deviation of the optimal portfolio from the benchmark $\mathbf{w}_t^b$. We follow (e.g., DeMiguel et al., 2009) and consider an equal-weight portfolio a benchmark where $w^b_{i,t} = 1/N_t, \forall i, t$, although other conventional strategies such as a value-weighted portfolio can be considered. The characteristics in $\widehat{\mathbf{X}}_t$ are standardised so that the cross-sectional distribution of the characteristics is stationary over time. In addition, the standardisation implies that the cross-sectional average of $\widehat{\mathbf{X}}_t\boldsymbol{\theta}$ is zero, which means that the deviations from the benchmark portfolio sum to zero, and as such, the optimal portfolio weights in $\mathbf{w}_t$ sum to one as far as the benchmark portfolio $\mathbf{w}_t^b$ sums to one.

[1] Doubling the number of stocks without otherwise changing the cross-sectional distribution of the characteristics results in twice as aggressive allocations, although the investment opportunities are fundamentally unchanged.

The coefficients $\boldsymbol{\theta} \in \mathbb{R}^k$ do not vary across assets. This implies that the optimal portfolio choice depends only on the characteristics of the stocks and not the stocks themselves. Thus, Eq.(1) "augments the asset space" from the base assets to the space of characteristic-managed portfolios, and the optimal portfolio problem of investing in thousands of stocks reduces to estimate the information content of firm characteristics for the cross-section of stock returns. The value of $\boldsymbol{\theta}$ that maximises the investor's expected utility depends on the utility function. In the case of mean-variance utility with δ risk aversion, set to $\delta = 5$ in our application, $\boldsymbol{\theta}$ can be found as

$$\begin{aligned}
\max_{\boldsymbol{\theta}} \frac{1}{T}\sum_{t=0}^{T-1} u\left(r_{p,t+1}\right) &= \frac{1}{T}\sum_{t=0}^{T-1}\left[\mathbf{w}_t^{\top} r_{t+1} - \frac{\delta}{2}\left(\mathbf{w}_t^{\top} r_{t+1}\right)^2\right] \\
&= \frac{1}{T}\sum_{t=0}^{T-1}\left[\left(\mathbf{w}_t^b + \frac{1}{N_t}\widehat{\mathbf{X}}_t\boldsymbol{\theta}\right)^{\top} r_{t+1} - \frac{\delta}{2}\left(\left(\mathbf{w}_t^b + \frac{1}{N_t}\widehat{\mathbf{X}}_t\boldsymbol{\theta}\right)^{\top} r_{t+1}\right)^2\right] \\
&= \frac{1}{T}\sum_{t=0}^{T-1}\left[\left(r_{t+1}^b + \boldsymbol{\theta}^{\top} f_{t+1}\right) - \frac{\delta}{2}\left(r_{t+1}^b + \boldsymbol{\theta}^{\top} f_{t+1}\right)^2\right] \qquad (2)
\end{aligned}$$

where $f_{t+1} \equiv \frac{1}{N_t}\widehat{\mathbf{X}}_t^{\top} r_{t+1}$ is the k-dimensional vector of returns on the characteristic-managed portfolios and r_{t+1}^b is the return on the benchmark portfolio. As a result, the first order condition to find the value of $\boldsymbol{\theta}$ that maximises the investors' utility is given by,

$$\frac{1}{T}\sum_{t=0}^{T-1}\left(f_{t+1} - \delta\left(r_{t+1}^b + \boldsymbol{\theta}^{\top} f_{t+1}\right)^{\top} f_{t+1}\right) = 0,$$

such that

$$\begin{aligned}
\widehat{\boldsymbol{\theta}} &= \frac{1}{\delta}\left(\frac{1}{T}\sum_{t=0}^{T-1} f_{t+1}^{\top} f_{t+1}\right)^{-1} \frac{1}{T}\sum_{t=0}^{T-1} f_{t+1}\left(1 - \delta r_{t+1}^b\right) \\
&= \frac{1}{\delta}\left(F^{\top}F\right)^{-1} F^{\top}\underbrace{\left(\mathbf{1} - \delta R^b\right)}_{Y} = \left(F_\delta^{\top} F_\delta\right)^{-1} F_\delta^{\top} Y \qquad (3)
\end{aligned}$$

where $F_\delta = F\delta$ with F the $T \times k$ matrix of characteristic-managed portfolio returns, and $\mathbf{1}$ and R^b are $T \times 1$ vectors of ones and benchmark portfolio returns, respectively. Based on Eq.(3), the estimate $\widehat{\boldsymbol{\theta}}$ can be found based on least squares projection of Y onto F_δ. Solving for the optimal mean-variance portfolio based on this linear regression implicitly considers the dependence of expected returns, variances, and covariances on firm characteristics to the extent that cross-sectional differences in these moments affect the expected utility of the portfolio returns (e.g. Britten-Jones, 1999).

2.1 Bayesian variable selection and shrinkage

The set of firm characteristics $\widehat{\mathbf{x}}_{i,t}$ can be very large (see Chen and Zimmermann, 2021, and the references therein). In this context, the performance of standard estimation techniques, such as ordinary least squares, maximum likelihood, or Bayesian inference with uninformative priors, tends to deteriorate as the number of characteristics increases. This is a well-known curse of dimensionality whereby the estimates $\widehat{\boldsymbol{\theta}}$ can be prone to overfitting. In addition, how many and which firm characteristics matter to capture the variation in stock returns may be uncertain a priori (e.g. Bryzgalova et al., 2023).

When interest lies in regression models with many parameters, several regularisation methods, such as the ridge (e.g., Hoerl and Kennard, 1970) or the lasso regression of Tibshirani (1996), have been proposed to address the risk of overfitting. Kelly et al. (2022) recently showed that a shrinkage "ridgeless" estimator, which regularises the regression estimates, outperforms most competing methods when it comes to predicting the equity premium out of the sample.

(Tibshirani, 1996, Section 5) first noticed that the lasso could be derived as the posterior mode – or the maximum a posterior (MAP) – estimate under a Laplace prior distribution with location zero and scale parameter equal to the inverse of the shrinkage intensity. Ishwaran and Rao (2005) show that a whole family of regularisation schemes can be specified as a hierarchical prior structure of the form $\boldsymbol{\theta}|\sigma^2, d_1, \ldots, d_k \sim \mathcal{N}(0, \sigma^2 D)$ where $D = \text{diag}(d_1, \ldots, d_k)$. The choice of the distribution to model $d_j \sim \pi$ defines what kind of shrinkage or sparsity strategy is being adopted. Popular choices are the two-component spike-and-slab prior of George and McCulloch (1993); Ročková and George (2014), the Laplace prior of Park and Casella (2008), the normal-gamma of Griffin and Brown (2010), the horseshoe of Carvalho et al. (2009), the Dirichlet-Laplace of Bhattacharya et al. (2015), or the Dirac spike-and-slab prior of Giannone et al. (2021).

In this paper, we adopt the Dirac spike-and-slab proposed by Fava and Lopes (2021) (GLP-t henceforth), which expands on Giannone et al. (2021) (GLP henceforth). GLP-t relaxes the assumption of normality for the non-zero component of the spike-and-slab, which is replaced with a more general scale mixture of normal such that

$$\boldsymbol{\theta} \mid \sigma^2, \gamma^2, \lambda_1^2, \ldots, \lambda_k^2, q \sim \mathcal{N}\left(0, \sigma^2 D\right), \qquad \text{with} \qquad D = \text{diag}\left(z_1\gamma^2\lambda_1^2, \ldots, z_k\gamma^2\lambda_k^2\right) \tag{4}$$

where $z_j \sim \text{Bernoulli}(q)$ the indicator that selects the jth variable and $\lambda_j^2 \sim \mathcal{IG}(\nu/2, \nu/2)$ is an inverse-Gamma distribution with scale and shape parameters equal to $\nu/2$. This implies that the marginal distribution of θ_j with respect to λ_j^2 is a Student-t with ν degrees of freedom, i.e., $\int p\left(\theta_j \mid \lambda_j^2\right) p\left(\lambda_j^2\right) d\lambda_j^2 = \mathcal{ST}_\nu(0, \sigma^2\gamma^2)$, (see Andrews and Mallows, 1974, and Appendix **??** for a formal proof). As a result, each of the regression parameters θ_j, $j = 1, \ldots, k$ can be either zero with probability $1 - q$, or a drawn from a distribution with zero mean and variance $\mathbb{V}(\theta_j) = \frac{\nu}{\nu-2}\sigma^2\gamma^2$, with probability q. This is a generalisation of GLP – which assumes $\theta_j \mid \sigma^2, \gamma^2, q \sim \mathcal{N}(0, \sigma^2\gamma^2)$ with probability q and $\theta_j = 0$ with probability $1 - q$.

The remaining priors for q and γ^2 are borrowed from Giannone et al. (2021). Specifically, we elicit the mapping $\gamma^2 = \frac{1}{k\bar{v}_x q} \cdot \frac{R^2}{1-R^2}$ and specify the marginal priors for $q \sim \mathcal{B}(a, b)$ and

$R^2 \sim \mathcal{B}(a,b)$ as Beta distributions with support $[0,1]$. Here $\overline{v}_x$ is redefined as $\mathbb{E}\left[\widehat{\sigma}_j^2\right]\frac{\nu}{\nu-2}$ where $\widehat{\sigma}_j^2$ represents the sample variance of the return on the jth characteristic-managed portfolio. Given this prior structure, we sample the posterior distribution by sequentially drawing from the conditional distributions of the parameters as defined in Proposition 1.

Proposition 1 (Posterior distributions)**.** *Let $Y = F_\delta \boldsymbol{\theta} + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_k)$ and assume an uninformative Jeffrey's prior for $\sigma^2 \propto 1/\sigma^2$. Define $D = diag(z_1\gamma^2\lambda_1^2, \ldots, z_k\gamma^2\lambda_k^2)$ and assume the prior structure as in Eq.(4). The posterior distributions of $\boldsymbol{\theta}, \sigma^2$ and $\lambda_j, j = 1, \ldots, k$ take the form:*

$$\boldsymbol{\theta} \mid data \sim \mathcal{N}\left(\Sigma F_\delta^\top Y, \sigma^2\Sigma\right), \quad with \quad \Sigma = \left(F_\delta^\top F_\delta + D^{-1}\right)^{-1} \tag{5a}$$

$$\lambda_j^2 \mid data \sim \mathcal{IG}\left(\frac{\nu+1}{2}, \frac{\nu}{2} + \frac{\theta_j^2}{2\sigma^2\gamma^2}\right) \qquad j = 1, \ldots, k \tag{5b}$$

$$\sigma^2 \mid data \sim \mathcal{IG}\left(\frac{T}{2}, \frac{Y^\top Y - \widehat{\boldsymbol{\theta}}^\top \Sigma^{-1}\widehat{\boldsymbol{\theta}}}{2}\right) \tag{5c}$$

where $\widehat{\boldsymbol{\theta}} = \Sigma F_\delta^\top Y$ is the posterior mean of $\boldsymbol{\theta} \mid data$.

We follow Giannone et al. (2021) and sample posterior draws by discretizing the support of $R^2, q \in [0,1]$ by interlacing two grids defined over the unit interval and then evaluate the joint posterior distribution. Appendix **??** provides the proof of Proposition 1 and additional details on the posterior distributions of R^2 and q. We now discuss the implications of the Student-t assumption for the non-zero θ_j in turn.

Prior properties and hyperparameters Equation 4 implies that when θ_j is non-zero, it is drawn from a Student-t distribution with ν degrees of freedom and variance scaled by the hyperparameter γ^2. The hyperparameter γ^2 controls the degree of shrinkage. The larger the γ^2 for a given σ^2, the smaller the shrinkage, as θ_j will be a priori more likely to be distant from zero. The hyperparameter ν also affects the posterior concentration of θ_j around zero, conditional on γ^2. Figure 1 shows this case in point. A smaller ν corresponds to a smaller and more dispersed λ_j^2 (see top-left panel), which, in turn, puts a higher probability on extreme values of θ_j (see bottom-left panel). This reduces the parameter shrinkage for a given level of γ^2 and σ^2, as the thicker tails of the Student-t inflates the prior variance (see right panel). This simple comparative static underscores that a higher (lower) ν tightens (widens) the distribution of λ_j^2, leading to a more concentrated (dispersed) distribution of θ_j around zero with lower (higher) variance for a given γ^2. As a result, ν plays a key role in shaping the prior shrinkage of θ_j.

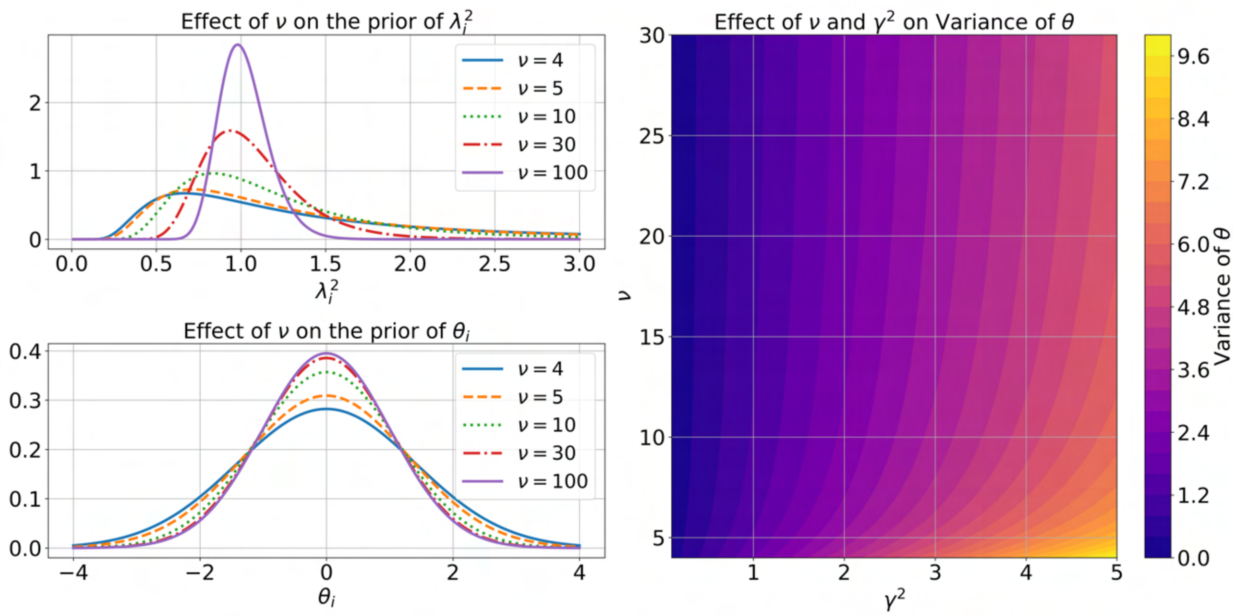


Figure 1: **The role of ν on prior shrinkage**. The left panel shows the effect of ν on the prior for λ_i^2 and θ_i for a fixed level of $\sigma^2 = \gamma^2 = 1$. The right panel shows the impact of ν and γ^2 on the prior variance of the non-zero coefficients θ_i.

The prior $R^2 \sim \mathcal{B}(a, b)$ as Beta distributions with support $[0, 1]$ is appealing for our parametric portfolio choice because it has the interpretation of a prior on the regression R^2, meaning the share of the cross-sectional variation of stock returns explained by characteristic-managed portfolios. As a result, it can be used independently on the portfolio size. In addition, eliciting a prior on R^2 is more agnostic on using sparsity vs shrinkage to deal with a large set of stock characteristics to the extent that they have some explanatory power for the cross-section of stock returns.

Figure 2 shows this case in point. The right panel shows that for a given level of prior R^2, assuming more or less sparsity a priori, i.e., varying q, does not significantly affect the prior shrinkage γ^2. Conversely, assuming that the R^2 is very large can reduce the prior shrinkage, especially for low sparsity q. This embeds the intuition that if firm characteristics explain a large portion of the cross-sectional variation of stock returns there is less need of regularising the coefficient estimates. In the empirical application, we opt for an uninformative approach and set $a = b = 1$, corresponding to a uniform distribution, such that $E\left[R^2\right] = E\left[q\right] = 0.5$.

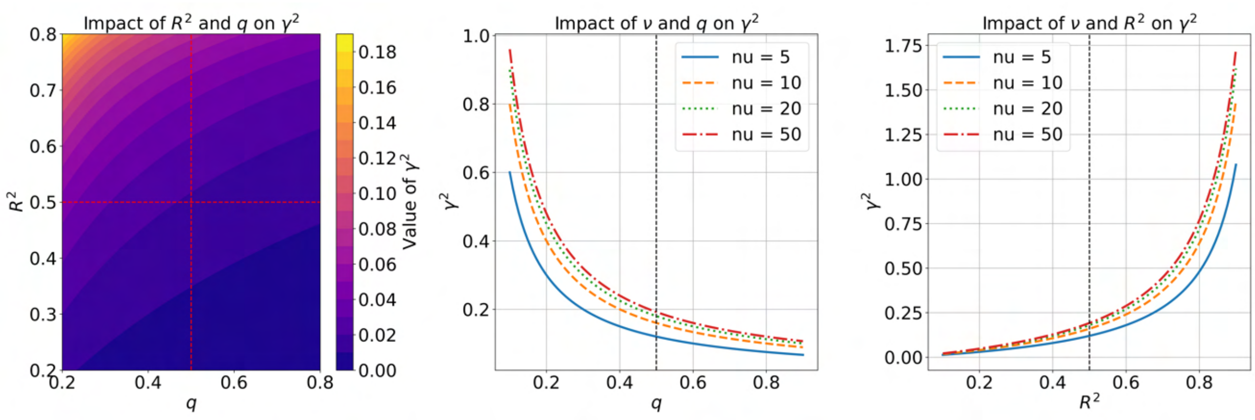


Figure 2: **The role of the marginal priors on** γ^2. The left panel shows the effect of R^2 and q on γ^2 for a fixed value of $k = 10$ and $\overline{v}_x = 1$. The middle panel shows the impact of q on γ^2 for different levels of ν. The right panel shows the impact of R^2 on γ^2 for different levels of ν. The dashed horizontal and vertical lines represent the prior mean of R^2 and q, as they are both drawn from a $\mathcal{B}(1,1)$.

The middle panel of Figure 2 shows a third property of the prior R^2; that is, there is a negative relationship between the hyperparameters γ^2 and q for different levels of ν. This embeds the tradeoff between addressing the curse of dimensionality based on a sparse model with few characteristics and a flat prior for the non-zero θ_j or a dense model with many characteristics but with a prior for θ_j more concentrated around zero (e.g. Chernozhukov et al., 2017). Smaller values of ν mitigate such tradeoff as the negative correlation between q and γ^2 flattens (see also Figure 1). Similarly, the right panel in Figure 2 shows that smaller values of ν correspond to less shrinkage for a given level of R^2 and prior sparsity q.

An economic rationale to calibrate ν. So far, we assumed that the hyperparameter ν for the spike-and-slab is fixed. Figures 1-2 show that different values of ν imply different prior views on θ_j. The larger the ν, the more concentrated around zero is the prior for θ_j. As we will show in the empirical results, ν has first-order importance for optimal portfolio allocation since it ultimately affects the tradeoff between sparsity and shrinkage. To show this take the ith row in the portfolio weight $w_{i,t} = w_{i,t}^b + \widehat{\mathbf{x}}_{i,t}\boldsymbol{\theta}$. The more we shrink the posterior distribution of θ_j toward zero, the smaller the impact of the jth characteristic on the capital allocation on the ith stock.

In addition to fixing ν to cover a wide range of assumptions on the prior tails, we build upon this intuition and propose a simple economic rationale to calibrate ν based on transaction costs. This is based on the premise that an increase in the absolute value of a characteristic j for a given stock i at time t, i.e., $\widehat{x}_{i,t}^j$ leads to higher liquidity needs and rebalancing costs due to $w_{i,t} = w_{i,t}^b + \left(\widehat{x}_{i,t}^1\theta_1, \ldots, \widehat{x}_{i,t}^k\theta_k\right)$. Thus, the posterior distribution of θ_j may directly affect portfolio profitability as larger values of ν may result in a more dispersed posterior for θ_j. To embed this intuition into the calibration of ν, we calibrate its value based on a measure of aggregate transaction costs,

$$TC_t = \frac{1}{N_t}\left|\sum_{i=1}^{N_t}\sum_{j=1}^{k}\widehat{x}_{i,t}^j\eta_{i,t}\right|, \qquad t = 1,\ldots,T \tag{6}$$

where for the full-sample estimates we calibrate $\nu = \sum_{t=1}^{T} TC_t$ and for the real-time, out-of-sample implementation we consider ν as the last observed TC_t for each recursive estimation.

Here $\eta_{i,t}$ is the half bid-ask spread for asset i at time t and captures trading costs based on the liquidity of an individual asset (e.g. Bessembinder and Venkataraman, 2010). Thus, for each stock, we calculate $\sum_{j=1}^{k} \widehat{x}_{i,t}^{j} \eta_{i,t}$, and then we average this value across stocks. We use the absolute value of this average to ensure that the proxy for transaction costs is always positive. Note that Eq.(6) accounts for the possibility that rebalancing different characteristics can reduce transaction costs. For example, characteristics with positive and negative values can offset each other, reducing the value of the term $\sum_{j=1}^{k} \widehat{x}_{i,t}^{j} \eta_{i,t}$ (e.g., DeMiguel et al., 2020).[2]

Before discussing the main empirical results, two comments are in order. First, in addition to a calibration based on transaction costs, we also build upon the prior comparative statics and experiment with different values of ν to assess its implications on the number of firm characteristics selected and the amount of shrinkage. Second, one can envision a more data-driven approach to estimate ν based on the data. To this end, Appendix **??** sketches a potential strategy to estimate ν based on a prior $\nu \sim \mathcal{G}(a, b)$. The posterior distribution is not available in closed form, and one needs to resort to a Metropolis-Hastings algorithm. This could be an interesting development for future research. Although it may imply some rigidity, we show in the empirical analysis that embedding some economic rationale in the prior structure provides a more intuitive approach to factor in the role of transaction costs for the interplay between shrinkage and sparsity and, ultimately, on the parametric portfolio choice. The latter is a primary objective of this paper.

[2] They show that when including transaction costs in an otherwise conventional lasso-type penalty, the number of characteristics needed to explain the cross-section of stock returns increases as the individual trading costs cancel each other out when trading more characteristic-managed portfolios.

3 Empirical analysis

We collect firm characteristics from the www.openassetpricing.com website (see Chen and Zimmermann, 2021, for more details).[3] The initial data includes 212 US firm characteristics every month from January 1925 to November 2022. These characteristics are merged with monthly stock returns data from the Center for Research in Security Prices (CRSP). We filter out stocks with market capitalization below the 10th percentile, thus excluding very small and illiquid stocks. We also filter out those observations with missing size data and showing extreme returns above 250% or below -100%.

We refine the data by considering only continuous characteristics demonstrating at least some cross-sectional return predictability, according to Chen and Zimmermann (2021). We filter out those characteristics with more than 60% of missing values on average during our initial sample period. The remaining missing values are imputed using the cross-sectional median for each stock each month (see, e.g., Gu et al., 2020). Additionally, we standardize the predictors 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 data set comprises 145 characteristics covering an unbalanced panel of 7,675 stocks with a minimum of 1,347 and a maximum of 2,588 stocks per month, from January 1985 to November 2022.

Figure **??** in Appendix **??** reports the cross-sectional average of the sample skewness (left panel) and kurtosis (right panel) for each characteristic. The descriptive statistics show that most characteristics exhibit, on average across stocks, a positive and large skewness and kurtosis over the sample period. This provides prima facie evidence that a normal prior for the non-zero θ_j may be too restrictive to capture the full extent of the distribution of the "signal" embedded in a given firm characteristic. A prior that supports more extreme values

[3] We use version 1.3.0, released in August 2023.

in the characteristics, such as the one outlined in Eq.(4), may be less restrictive and, as such, more practically relevant.

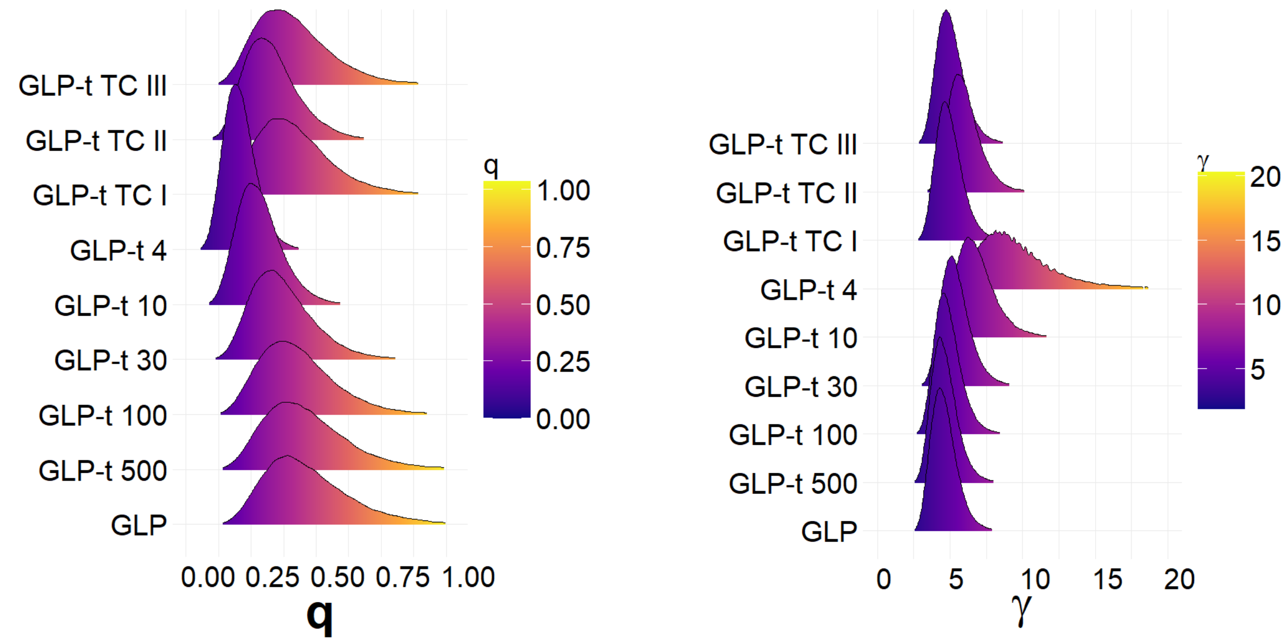
3.1 In-sample results

In this Section, we discuss the full sample estimates of the marginal posterior distributions for q and γ^2 for different values of ν. To this end, we consider the GLP prior and the GLP-t(ν) with $\nu = [4, 10, 30, 100, 500]$ and when $\nu = \sum_{t=1}^{T} TC_t$, with the transaction costs at time t approximates as in Eq.(6). Note the value of $TC_t << 1, \forall t$. As a result, a simple plug-in calibration would be unfeasible as none of the moments of the Student-t can be defined for $\nu < 1$. To address this issue, we consider different order-invariant rescaling strategies, which then redefine the calibrated degrees of freedom in the interval $[4, 500]$ (TC1), $[4, 100]$ (TC2), or multiplied by 1,000 (TC3).[4] Next, we will discuss the posterior inclusion probabilities for individual firm characteristics and the implications for the optimal portfolio choice. Finally, we will discuss the in-sample performance of different priors vs conventional benchmarks and alternative shrinkage priors.

Posterior evidence of sparsity vs shrinkage. We begin by investigating the marginal posteriors of q (sparsity) and γ^2 (shrinkage) parameters based on the full sample of observations. Figure 3 reports the results. Three facts emerge. First, a heavier-tailed prior produces stronger evidence in favour of aggregate sparsity; the smaller the ν, the more the posterior distribution of q is concentrated around a small value. Conversely, shrinkage is negatively correlated with the prior degrees of freedom, as highlighted by the posterior estimates of γ^2. These become more dispersed around larger values as the value of ν decreases.[5] Second, with the partial exception of $\nu = 4$, there is little evidence of a clear sparsity pattern in the set

[4]The rescaled value of ν is 68.85 for TC1, 16.55 for TC2, and 54.60 for TC3.
[5]This echoes the results in Fava and Lopes (2021) in the context of returns predictability.

of firm characteristics as the posterior distribution of q is far from zero. This echoes the intuition in Kozak et al. (2020); Haddad et al. (2020); Kelly et al. (2022), whereby a dense model may represent a better approximation of the stochastic discount factor in the presence of many risk factors. Third, a simple calibration of ν based on transaction costs does not support a heavier-tailed specification ($\nu = 4$), favouring sparsity over shrinkage and allowing the prior for the selected characteristics to be rather flat, i.e., less shrinkage. This is preliminary evidence that selecting many variables, i.e., high q, while shrinking their coefficients, i.e., low γ^2, may be more adequate to summarise the information in firm characteristics when transaction costs are considered (see DeMiguel et al., 2020).



(a) Posterior of q

(b) Posterior of γ^2

Figure 3: **In-sample posterior estimates of q and γ^2.** The figure shows the in-sample posterior estimates of q (left panel) and γ^2 (right panel) for the GLP and the GLP-t(ν) with $\nu = [4, 10, 30, 100, 500, TC1, TC2, TC3]$. The sample period is from January 1985 to November 2022.

The marginal posterior estimates of q and γ^2 suggest that for smaller values of ν, the evidence in favour of sparsity strengthens, whereas there is less shrinkage a priori, and vice versa.

A heavier-tailed prior has more aggressive sparsity-inducing properties, leaving the posterior estimate of θ_j relatively unconstrained for the selected characteristics. This is consistent with the results in Fava and Lopes (2021) in the context of stock returns predictability. Figure **??** in Appendix **??** provides further evidence by juxtaposing the joint posterior density of q and γ^2 with the joint prior for different values of ν. The posterior densities are much more concentrated than the corresponding prior, exhibiting an even sharper negative correlation: the lower (higher) the sparsity, the larger (smaller) the shrinkage imposed on θ_j's.

The important implication of the posterior estimates shown in Figure 3 is that focusing on sparsity and ignoring shrinkage as a viable tool to reduce overfitting and estimation error might lead to artificially recovering a sparse set of characteristics, which may not necessarily be best positioned to capture the cross-sectional variation in stock returns. Our findings persist for heavier-tailed prior specifications (see Figure **??**).

Firm characteristics and model uncertainty. The previous subsection has presented evidence about the *share* of relevant firm characteristics for the cross-section of stock returns; that is, the degree of average sparsity as proxied by the posterior density of q. We now ask whether the *identity* of these firm characteristics can be recovered for different levels of ν. To this end, we calculate the posterior inclusion probability of each firm characteristic in the optimal portfolio choice.[6] Figure 4 shows the results. Each horizontal stripe corresponds to a firm characteristic, and darker shades denote higher inclusion probabilities.[7] The estimates obtained from different prior specifications are labelled by column.

Except for GLP-t(4), there is no clear sparsity pattern in that none of the characteristics

[6] Appendix **??** provides a detailed derivation of how the probability of inclusion for each firm characteristic is calculated.

[7] 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.

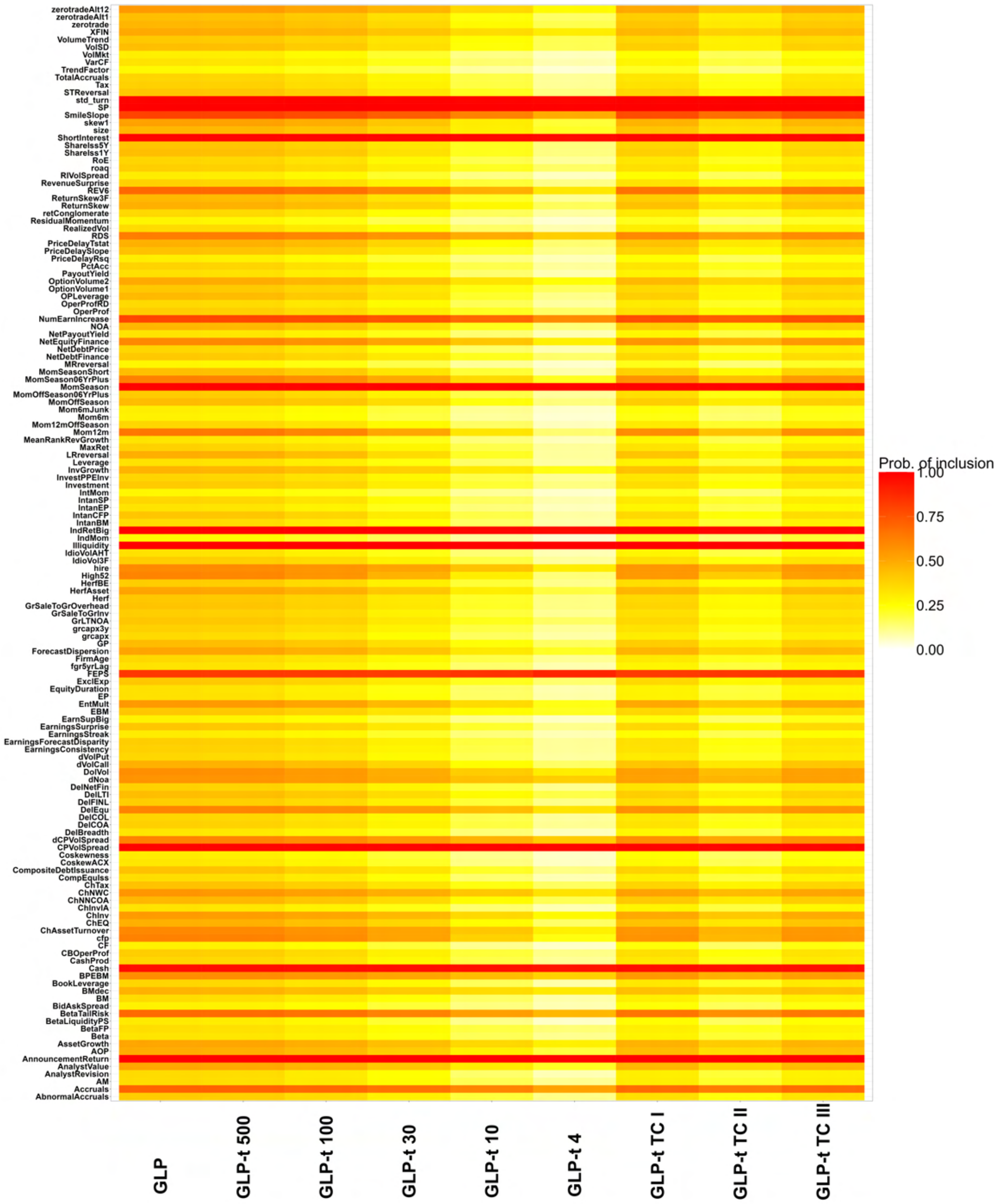


Figure 4: **In-sample posterior inclusion probabilities**. The figure shows the in-sample posterior probability of inclusion for each characteristic in the cross-section of stock returns for different levels of $\nu = [4, 10, 30, 100, 500, TCI, TCII, TCIII]$ for the GLP specification. The sample period is from January 1985 to November 2022.

can be unequivocally excluded from the parametric portfolio rule. This is despite the posterior distribution of q having most probability mass between 0.25 and 0.5. These values of q do not necessarily imply that the most accurate model includes between 25% to 50% of the firm characteristics and excludes all others. If this were the case, the inclusion probabilities for, e.g., GLP, GLP-t(ν) with $\nu = 100, 500$ or the GLP-t with ν calibrated based on transaction costs, would show many near-white stripes corresponding to the firm characteristics that are systematically excluded. Instead, there is substantial model uncertainty about whether certain firm characteristics should be used to predict the cross-section of stock returns, which results in their selection only in a subset of the posterior draws.

Perhaps not surprisingly (see Figure 3), model uncertainty is reduced for smaller values of ν, as the inclusion probability of most firm characteristic decreases to the point where it certainly can be excluded (see the white stripes for GLP-t(4). This reflects the intuition a heavier-tailed prior may be better equipped to learn whether to include a given characteristic or not (e.g. Fava and Lopes, 2021). Interestingly, a more sparsity-inducing prior does not change the likelihood of selecting strong characteristics but only reduces the probability of including those characteristics only mildly associated with the cross-section of stock returns, i.e., those with a low inclusion probability across priors.

In this respect, Table **??** in Appendix **??** shows that when applying a simple threshold of 0.5 (or 50%) to each posterior inclusion probability, earnings announcement return (Chan et al., 1996), cash holdings (Palazzo, 2012), illiquidity (Amihud, 2002), return seasonality (Heston and Sadka, 2008), short interest (Dechow et al., 2001), volatility spreads (Bali and Hovakimian, 2009), accruals (Sloan, 1996), and turnover volatility (Chordia et al., 2001) are firm characteristics that are always selected irrespective of the prior tails. Yet, the number of selected characteristics increases as we increase the value of ν.

Implications for optimal portfolio choice. The evidence reported so far suggests that different assumptions on the prior tails – meaning different values of ν – have implications for the number of firm characteristics that enter the parametric portfolio choice as well as the amount of shrinkage on the posterior distribution of θ_js itself. Figure 5 shows the posterior estimates of θ_j for some of the firm characteristics selected across all prior specifications based on a 0.5 threshold (see Table ??). For ease of exposition, each subplot reports the estimates from GLP, GLP-t(30), GLP-t(TC1), and GLP-t(4). Figure ?? in Appendix ?? also shows the posterior estimates of corresponding λ_j^2.

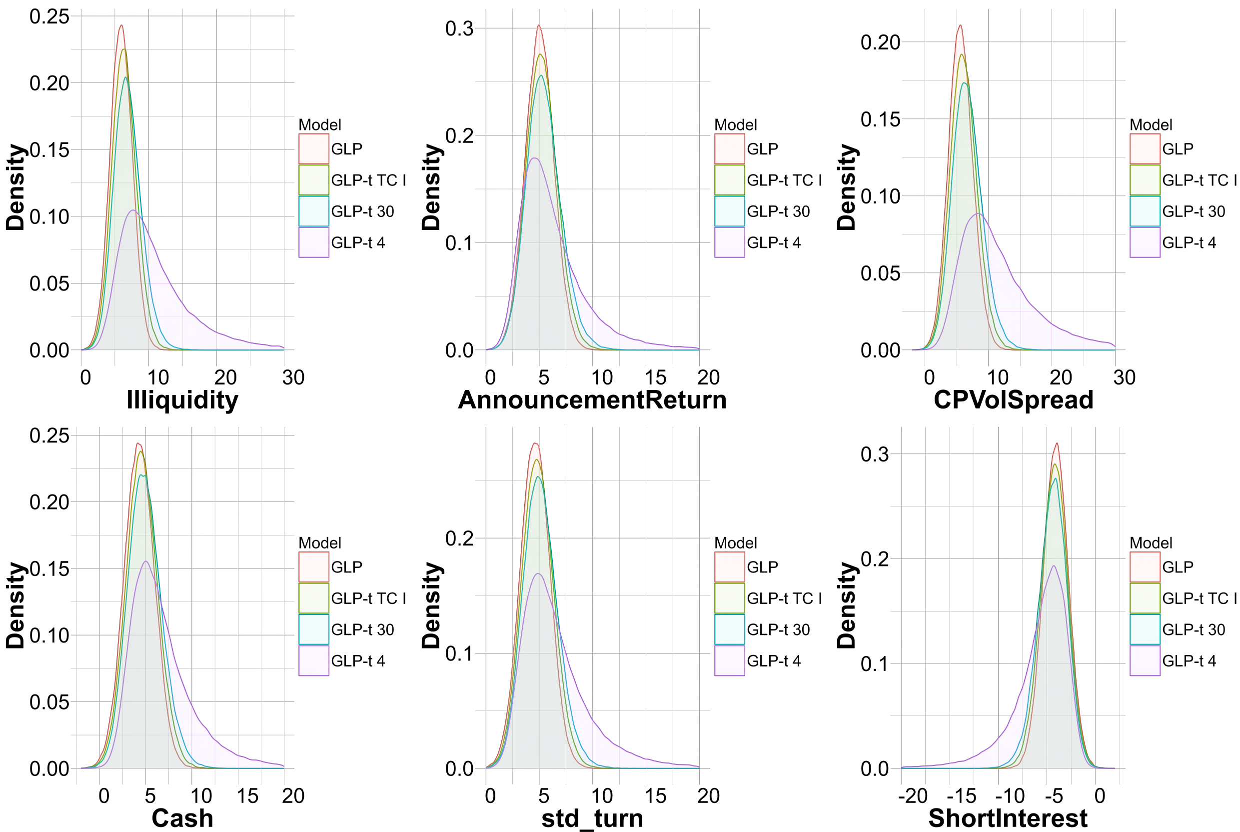


Figure 5: **Posterior estimates of θ_j.** The figure shows the posterior estimates for some firm characteristics selected by all spike-and-slab priors. The sample period is from January 1985 to November 2022.

A heavier-tailed prior translates into posterior distributions less concentrated around their mean. The lower the value of ν, the more dispersed the posterior of θ_j. This is particularly evident for characteristics such as the call-minus-put implied volatility spread (CPVolSpread), the Amihud (2002) illiquidity ratio, cash holdings (Cash), and turnover volatility (std_turn). Notably, when the tails of the prior are calibrated based on transaction costs (TC1), the posterior promotes a shrinkage that is not dissimilar from the GLP prior; that is, the probability mass on the tails is closer to a normal distribution.

The posterior estimates in Figures 4 and Figure 5 have first-order implications for optimal portfolio allocations. Recall the parametric portfolio choice for a given stock i is $w_{i,t} = w^b_{i,t} + \left(\widehat{x}^1_{i,t}\theta_1 + \ldots + \widehat{x}^k_{i,t}\theta_k\right), \forall i, t$. In the language of Brandt et al. (2009), selecting fewer characteristics but with a larger impact means that a representative investor trades firm characteristics less at the extensive margin but more at the intensive margin. This tradeoff can have important consequences for portfolio profitability net of transaction costs. For instance, as argued by DeMiguel et al. (2020), selecting more characteristics rather than less can ultimately improve portfolio performances as the trading costs cancel out across different characteristic-managed portfolios.

Therefore, while heavier tails may help reduce the uncertainty as to which firm characteristics matter for the cross-section of stock returns, the larger weight implied by those characteristics may hamper the portfolio profitability once transaction costs are factored in. Put differently, although inducing more sparsity may be beneficial from a pure signal extraction standpoint, it may be suboptimal from an economic perspective.

In the next section, we put this intuition to task and investigate the in-sample performance of the parametric portfolio choice in Eq.(1) based on different prior assumptions. A particular emphasis is given to the role of prior tails, i.e., ν, as this has been shown to affect the tradeoff between sparsity and shrinkage.

In-sample portfolio performance. In addition to the GLP and GLP-t priors, we consider two popular Bayesian shrinkage approaches such as the Bayesian lasso (Park and Casella, 2008) and the horseshoe (Carvalho et al., 2009). This allows us to underscore the role of regularisation methods that do not explicitly allow for sparsity in regression coefficients. Park and Casella (2008) built upon Tibshirani (1996) and proposed a prior of the form $\theta_j \sim \mathcal{N}\left(0, \sigma^2\lambda_j^2\right), \lambda_j^2 \sim \mathcal{E}\left(\gamma^2/2\right), \gamma^2 \sim \mathcal{IG}\left(a, b\right)$, where $\mathcal{E}\left(\gamma^2/2\right)$ represents an exponential distribution with rate parameter $\gamma^2/2$.[8] The latter is akin to the penalty term in a conventional lasso regression; the larger γ^2, the more concentrated the prior is around zero. The horseshoe prior proposed by Carvalho et al. (2009) has the form $\theta_j \sim \mathcal{N}\left(0, \sigma^2\gamma^2\lambda_j^2\right), \lambda_j^2 \sim \mathcal{C}^+\left(0,1\right), \gamma^2 \sim \mathcal{C}^+\left(0,1\right)$, where $\mathcal{C}^+\left(0,1\right)$ is the half-Cauchy distribution on the positive reals with scale parameter one. Appendix **??** provides more details on the posterior distributions from both approaches.

Table 1 reports the descriptive statistics on the optimal portfolio weights $\mathbf{w}_t = \mathbf{w}_t^b + \frac{1}{N_t}\widehat{\mathbf{X}}_t\widehat{\boldsymbol{\theta}}$ (Panel A) and the corresponding portfolio performance (Panel B). The estimates $\widehat{\boldsymbol{\theta}}$ are based on the full sample of stock returns and firm characteristics. In addition to the equal-weight portfolio, i.e., $1/N_t$, we compare each prior against the parametric portfolio choice in Brandt et al. (2009) (BSV henceforth). The latter includes size, book-to-market, and twelve-month momentum as firm characteristics.[9]

Few results emerge. The BSV approach entails less extreme weights, with a range between -1.22% and 2.27%. This compares to $[-11.9\%, 13.8\%]$ and $[-5\%, 11.5\%]$ for the Bayesian lasso and the horseshoe, respectively. This is perhaps surprising, given that BSV is based on a handful of firm characteristics. More importantly, a heavier-tailed prior implies more extreme positions in individual stocks compared to a more conservative calibration of ν; for

[8] Tibshirani (1996) first noted that the frequentist lasso estimate could be derived as a Bayes posterior mode under a Laplace prior.

[9] In this respect, the original implementation of Brandt et al. (2009) can be seen as inducing extreme sparsity in the set of available characteristics.

	EW	BSV	Blasso	HS	GLP	GLP-t(ν)						
						4	10	30	100	TC1	TC2	TC3
Panel A: Portfolio weights statistics												
Max w (%)	0.049	2.279	13.863	11.583	10.342	18.256	13.522	11.126	10.760	10.531	12.113	10.658
Min w (%)	0.049	-1.227	-11.947	-4.999	-5.863	-7.967	-6.540	-5.998	-5.712	-5.706	-6.226	-5.751
Mean |w| (%)	0.049	0.255	0.899	0.409	0.492	0.625	0.534	0.514	0.498	0.492	0.514	0.495
Mean (w < 0) (%)	0.000	-0.255	-0.844	-0.312	-0.410	-0.487	-0.418	-0.411	-0.400	-0.395	-0.404	-0.396
Prop w < 0	0.000	0.403	0.504	0.580	0.544	0.595	0.584	0.570	0.565	0.565	0.579	0.567
Herfindahl	0.000	0.022	0.429	0.112	0.141	0.259	0.175	0.149	0.143	0.139	0.157	0.141
Turnover	0.073	0.970	12.532	6.768	7.837	9.810	8.478	8.238	7.908	7.961	8.379	8.012
Panel B: Portfolio performance												
Mean	0.012	0.029	0.131	0.084	0.103	0.120	0.105	0.101	0.104	0.103	0.100	0.104
Sd	0.049	0.071	0.085	0.093	0.090	0.128	0.107	0.093	0.094	0.096	0.099	0.097
Skew	-0.805	-0.384	0.167	0.702	0.438	0.700	0.608	0.240	0.620	0.655	0.399	0.614
ES (5%)	-0.113	-0.130	-0.041	-0.101	-0.079	-0.133	-0.110	-0.097	-0.085	-0.089	-0.106	-0.091
SR	0.660	1.273	5.229	3.044	3.882	3.195	3.325	3.659	3.712	3.626	3.418	3.607
p-value ΔSR	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fees	-		0.096	0.045	0.065	0.055	0.056	0.061	0.063	0.062	0.057	0.062
Panel C: Portfolio performance with transaction costs												
Mean	0.012	0.024	0.070	0.051	0.064	0.071	0.063	0.061	0.065	0.064	0.058	0.064
Sd	0.048	0.071	0.081	0.089	0.085	0.121	0.101	0.090	0.089	0.091	0.094	0.092
Skew	-0.811	-0.441	0.084	0.464	0.257	0.471	0.391	0.067	0.383	0.408	0.188	0.379
ES(5%)	-0.113	-0.137	-0.099	-0.136	-0.114	-0.18	-0.152	-0.138	-0.123	-0.129	-0.147	-0.132
SR	0.648	1.032	2.914	1.898	2.521	1.968	2.067	2.251	2.411	2.343	2.063	2.321
p-value ΔSR	0.009		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fees	-		0.042	0.019	0.034	0.018	0.024	0.028	0.032	0.031	0.023	0.031

Table 1: **In-sample portfolio performance**. This table reports the in-sample portfolio weights statistics (Panel A), performance (Panel B), and performance with transaction costs (Panel C) from the parametric portfolio choice as in Eq.(1). Transaction costs are included in the performance as in Eq.(8). We report the implied fees f only if they are positive. In addition to the GLP and GLP-t priors with different ν calibrations, we also consider the benchmark portfolio, i.e., $1/N_t$, two alternative shrinkage priors such as the Horseshoe (HS) and the Bayesian lasso (Blasso), and the original implementation of Brandt et al. (2009) (BSV). The sample period is from January 1985 to November 2022.

instance, the GLP-t(4) produces portfolio weights in the range $[-7\%, 18\%]$, which compares to $[-5.7\%, 10.76\%]$ obtained from the GLP-t(100). The latter is almost equivalent to the GLP specification.

More extreme weights are coupled with less diversification, as proxied by the Herfindahl index (HI henceforth). The latter is calculated by squaring the weight $w_{i,t}$ allocated to each stock in the portfolio and then summing the resulting values. The higher the index, the higher the concentration of the weights in a few stocks and, as such, the less diversified the portfolio is, whereas a smaller value indicates the opposite. Panel A reports the average of the HI index over the sample period. A more sparsity-inducing specification with thick tails leads to a substantially higher portfolio concentration. For instance, the average HI for the GLP-t(4) is almost twice as large as the GLP-t(100) and the GLP prior. Overall, a heavier-tailed prior implies trading fewer stocks but more at the intensive margin, whereas a normal spike-and-slab encourages more evenly distributed portfolios. The last three columns of Panel A suggest that calibrating ν based on transaction costs produces less extreme weights and more diversified portfolios.

Panel B of Table 1 reports the in-sample portfolio performance without considering transaction costs. Interestingly, all priors outperform the BSV benchmark. For instance, the GLP prior achieves a 3.8 annualised Sharpe ratio, more than three times larger than BSV and more than six times the equal-weight portfolio. The null hypothesis that the Sharpe ratios are the same is strongly rejected based on p-values obtained from the bootstrap approach of Ledoit and Wolf (2008). In addition to the Sharpe ratios, we follow Della Corte et al. (2008) and quantify the fee f which would make an investor indifferent, in terms of utility, between a parametric portfolio constructed based on the BSV approach or any other prior specification. For mean-variance utility, the fee is calculated as,

$$\sum_{t=1}^{T}(r_{p,t} - f) - \frac{\delta}{2}(r_{p,t} - f)^2 = \sum_{t=1}^{T} r_{Bench,t} - \frac{\delta}{2} r_{Bench,t}^2, \tag{7}$$

where δ is the risk aversion parameter, set to $\delta = 5$ in our application, and $r_{Bench,t}$ represents the benchmark portfolio returns, set to BSV in our application. The results show that a mean-variance investor is willing to pay up to 6.3% monthly to access the portfolio allocation obtained from the GLP-t(100) or GLP instead of BSV. This compares to 5.5% from the GLP-t(4) or 4.5% from the horseshoe prior. Again, the larger the ν, the larger the fees an investor will pay to access the portfolio allocation obtained from the spike-and-slab prior.

Panel A suggests that the higher performance shown in Panel B comes at the cost of a larger turnover. The latter is calculated as $\sum |w_{i,t} - w^{+}_{i,t-1}|$, where $w_{i,t}$ represents the portfolio weight for asset i at time t, and $w^{+}_{i,t-1} \equiv w_{i,t-1}(1+r_{i,t})$ is the adjusted weight from the previous period (e.g. DeMiguel et al., 2009). For instance, the portfolio constructed using GLP-t(4) included 12 variables and had a turnover of 9.81 compared to the 37 characteristics selected by GLP-t(100) and GLP with a turnover of 7.9 and 7.84, respectively. As a result, it is worth investigating the performance of GLP and GLP-t once transaction costs are factored in. To this end, we calculate the dynamics of wealth obtained from each portfolio choice as

$$W_t = W_{t-1}(1 + r_{p,t}) \left(1 - \sum_{i=1}^{N_t} \eta_{i,t} |w_{i,t} - w^{+}_{i,t-1}| \right), \qquad t = 2, \ldots, T \tag{8}$$

where W_t denotes the wealth at period t, $r_{p,t}$ is the portfolio return, and $\eta_{j,t}$ is the half bid-ask spread for asset i at time t. The return net of transaction costs is thus given by $r^{Net}_{p,t} = \frac{W_t}{W_{t-1}} - 1$ (e.g., DeMiguel et al., 2009).

Panel C in Table 1 shows the results. Not surprisingly, the in-sample portfolio performance substantially deteriorates; the Sharpe ratio from GLP-t(100) (or GLP) goes from 3.7 to 2.4 (or from 3.8 to 2.5) annualised. The implied fees are also substantially lower; the GLP-t(100) goes from 6.5% without transaction costs to 3.4% with transaction costs. Yet, there is a clear inverse relationship between the assumption on the prior tails, meaning ν, and the portfolio performance. The heavier-tailed GLP-t(4) prior underperforms most other approaches except for the horseshoe.

Note that in Table 1 we selected firm characteristics based on a conventional 0.5 threshold, meaning we retained those firm characteristics with a posterior inclusion probability larger than 0.5 (see Table **??** for more details).[10] However, it is important to underscore that our findings are robust to different thresholds. In Appendix **??**, we consider a more restrictive approach whereby we exclude a given characteristic if its posterior inclusion probability is less than $1-\text{mode}(q)$. This implies that the number of characteristics that enter the portfolio rule decreases as sparsity increases. Table **??** shows that the results based on an alternative selecting threshold are similar to Table 1, and all the same conclusions hold.

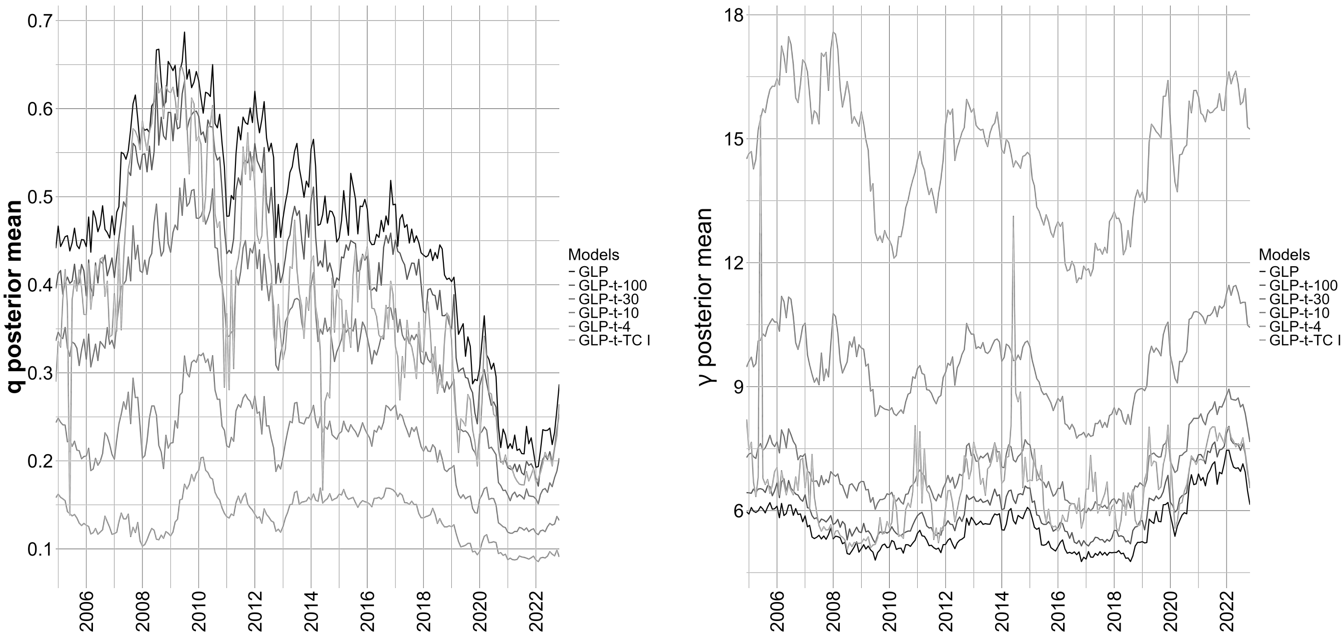
3.2 Out-of-sample results

The results reported so far are based on the full-sample posterior estimates of θ_j, q and γ^2. Therefore, the portfolio allocations are inherently in-sample and are of limited practical utility. We now take a more realistic approach and discuss the recursive estimates of the sparsity q and shrinkage γ^2 parameters – and the corresponding inclusion probabilities – as a function of ν. The recursive parameter estimates will later serve as the basis for a real-time implementation of the parametric portfolio choice introduced in Section 2.

Posterior estimate of q and γ^2 over time. We implement each estimate using a rolling window of 240 months of training data. We consider the GLP prior, GLP-t fixing $\nu = [4, 10, 30, 100, 500]$ and $\nu = TC_t$, where TC_t is set to the last observed value over the training period and thus will vary over time, reflecting the different transaction costs for each period. Similar to the full-sample estimation, TC_t is rescaled at each time t in the interval $[4, 500]$ (TC1), $[4, 100]$ (TC2), or multiplied by 1,000 (TC3).

Figure 6 reports the recursive posterior mean of q (left panel) and γ^2 (right panel). To increase readability, we report the estimates based on GLP, GLP-t(ν) with $\nu = [4, 10, 30, 100, TC1]$

[10] Barbieri and Berger (2004) show that the median probability model, meaning the model where only variables with probabilities larger than 0.5 are selected/retained, is optimal for prediction.



(a) Posterior mean of q

(b) Posterior mean of γ^2

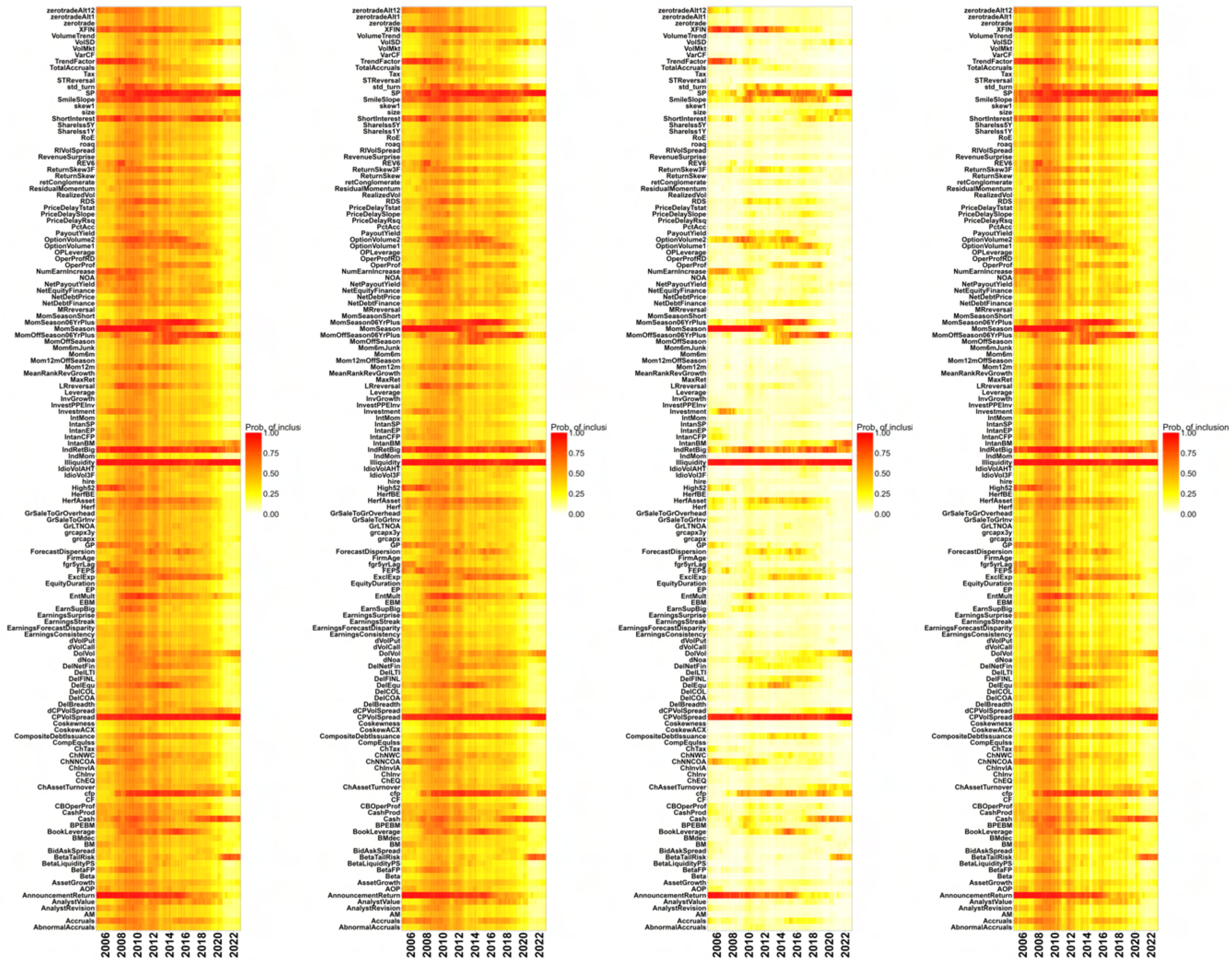
Figure 6: **Recursive estimates of q and γ^2.** The figure shows the posterior mean of q (left panel) and γ^2 (right panel) for different levels of $\nu = [4, 10, 30, 100, 500, TC1]$ and for the GLP specification. The estimates are based on a rolling window of 240 months. The sample period is from January 1985 to November 2022.

and leave to Figure **??** in Appendix **??** the remaining results. There is a considerable variation in the posterior estimates. For instance, the posterior mean of q for the GLP prior is as high as 0.65 in the aftermath of the great financial crisis and then drops to 0.25 throughout the COVID-19 pandemic at the end of the sample. GLP-t(ν) with $\nu = 100$ shows a similar trajectory. This suggests that a prior with less heavy tails consistently favours a dense model which selects many variables. On the other hand, with a more sparsity-inducing assumption on the prior tails, i.e., a smaller ν, the evidence in favour of sparsity is more convincing and stable over time, with the posterior mean of q consistently in the range of 0.1-0.2 for GLP-t(4) (0.15-0.3 for GLP-t(10)). Interestingly, when calibrating ν based on the latest estimates of transaction costs, the trajectory of the posterior mean of q follows a similar pattern of GLP but is slightly more erratic, reflecting the volatile nature of TC_t over the sample.

The right panel of Figure 6 shows the recursive posterior mean of γ^2. Consistent with the full-sample estimates in Figure 3, there is a negative relationship between sparsity and shrinkage, with the latter trending upward over the sample period for GLP, GLP-t with $\nu = 100, 30$. Such an inverse relationship is less evident for heavier-tailed Student-t priors, which show persistently larger values of γ^2 over time. The evidence on q and γ^2 together suggests that a heavier-tailed distribution tends to induce more sparsity over time while leaving the posterior estimate of θ_j relatively less constrained. Overall, the posterior mean of γ^2 tends to be smaller for smaller values of ν while the posterior mean of q is far from zero. This implies that shrinkage cannot be dismissed as a tool to summarise the information across firm characteristics (e.g. Kelly et al., 2022). The same intuition is confirmed when calibrating the prior tails based on time-varying transaction costs.

Time-varying inclusion probabilities. The previous subsection has presented evidence about the *share* of relevant firm characteristics over time. We now ask whether the *identity* of these firm characteristics is well identified over the sample period. To this end, Figure 7 plots the posterior probabilities of inclusion of each firm characteristic over time. To increase readability, we report only the results for the GLP prior, GLP-t(100), GLP-t(4), and GLP-t(TC1). Each horizontal stripe corresponds to a firm characteristic, with darker shades denoting higher inclusion probabilities. Time is reported on the x-axis.

Consistent with the posterior estimates of q, there is substantial uncertainty over time about whether certain firm characteristics should be used to predict the cross-section of stock returns. Such uncertainty is reduced for smaller values of ν to the point where a large set of firm characteristics can unequivocally be excluded (see the white stripes for GLP-t(4)); only illiquidity and volatility spreads turn out to show an inclusion probability of one for the whole sample, whereas other characteristics, such as earnings announcements, industry



(a) GLP (b) GLP-t(100) (c) GLP-t(4) (d) GLP-t(TC1)

Figure 7: **Recursive estimates of the probability of inclusion for each characteristic**. The figure shows the recursive estimate of the posterior probability of inclusion for each characteristic based on the GLP, the GLP-t(100), the GLP-t(4), and the GLP-t(TC1) priors. The sample period is from January 1985 to November 2022.

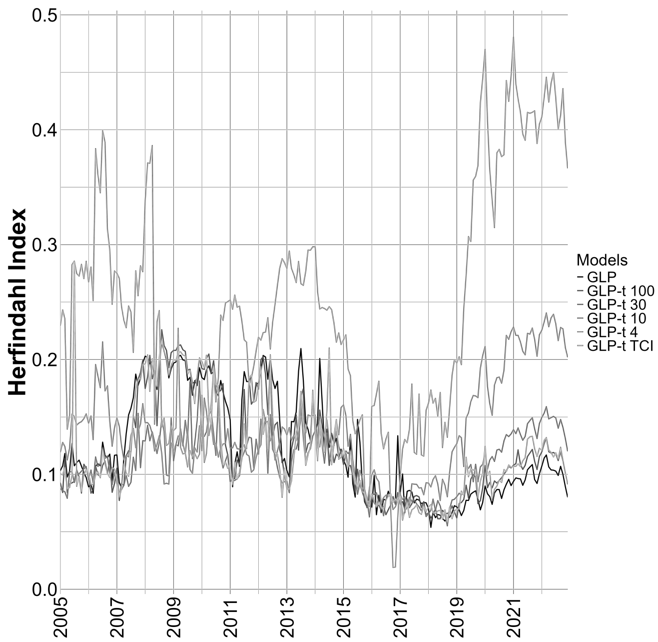
return of big firms (Hou, 2007), momentum seasonality, sales-to-price (Barbee Jr et al., 1996), and next external financing (Bradshaw et al., 2006), have an inclusion probability of one over different periods. This echoes the full-sample results whereby a heavier-tailed distribution reduces model uncertainty.

Yet, the GLP and the GLP-t(4) are similar in terms of which characteristic has a posterior inclusion probability equal to one. The main difference lies in the sparsity-inducing properties of those characteristics that appear systematically excluded, as highlighted by many near-white stripes for the GLP-t(4) prior. The right panel of Figure 7 shows that calibrating ν based on transaction costs the prior promotes a degree of sparsity and model uncertainty that is close to GLP.

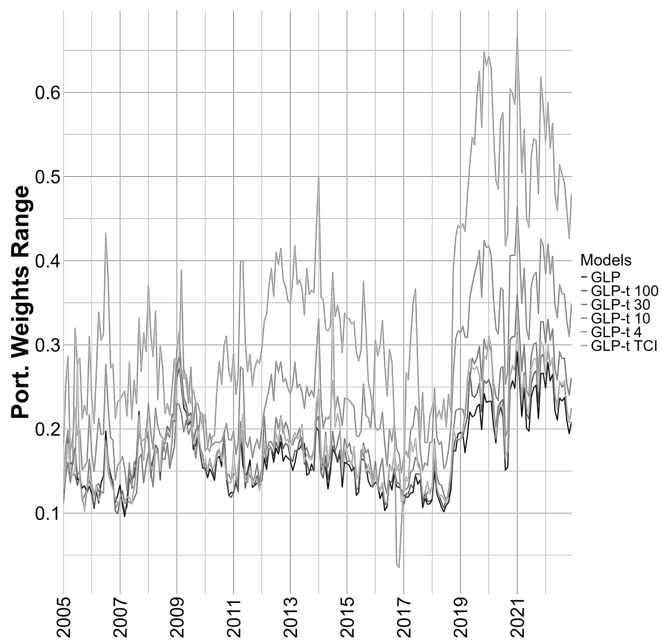
Implications for real-time portfolios. As shown by Figure 6 and Figure 7, different assumptions on the prior tails involve a different number of firm characteristics that likely enter the parametric portfolio choice and the amount of shrinkage imposed on each of the corresponding θ_js. In this section, we investigate the implications for the portfolio allocation at the extensive vs intensive margin. To this end, we begin by calculating the portfolios' Herfindhal (HI) index at each time t. The left panel of Figure 8 shows the HI value over the out-of-sample period.

A more sparsity-inducing prior with low ν translates into a less diversified portfolio. The HI index of the portfolio obtained from the GLP-t(4) is more than four times larger than that obtained from GLP in the lead-up and throughout the COVID-19 pandemic. With the partial exception of GLP-t(10), the level of portfolio diversification is fairly comparable for all other prior specifications. Consistent with the posterior mean estimates of q and γ^2 (see Figure 6), by calibrating ν based on transaction costs, the prior promotes a portfolio diversification that is, albeit more erratic, comparable with the GLP prior.

We also investigate the dynamic of the portfolio weights based on different prior assumptions. Specifically, we calculate the distance between the maximum and minimum portfolio weight $|\max(w_{i,t}) - \min(w_{i,t})|$ at each time t. We use this as a measure of intensive margin, as higher spreads between the long and short positions indicate higher liquidity needs (e.g., Patton and Weller, 2020). The right panel of Figure 8 shows that a heavier-tailed prior



(a) Portfolio Herfindhal index



(b) Portfolio weights range

Figure 8: **Portfolio diversification and extreme weights**. The figure shows the Herfindhal index (left panel) and the weights range (right panel) obtained for the recursive parametric portfolio allocation for different prior specifications. The value of the Herfindhal index is rescaled in the interval $[0, 1]$ to increase readability. The optimal allocation is based on a rolling window of 240 months. The sample period is from January 1985 to November 2022.

implies more extreme portfolio weights, with the spread between $\max(w_{i,t})$ and $\min(w_{i,t})$ as high as 0.6 (or 60%) during the COVID-19 pandemic. This is three times higher than a more restrictive GLP prior or a GLP-t(TC1) prior.

Overall, Figure 8 provides some interesting insight into the implications of different priors for real-time optimal portfolio choices. A less sparsity-inducing prior specification with no heavy tails generates a more diversified portfolio, i.e., more trading at the extensive margin, and less extreme portfolio weights, i.e., less trading at the intensive margin. The opposite holds for smaller values of ν. This provides direct evidence to support the intuition by DeMiguel et al. (2020) whereby when considering transaction cost, a model that considers many characteristics while shrinking their effect on the optimal portfolio choice may be preferable, from an economic standpoint to a more sparse representation of the cross-section of stock returns.

Out-of-sample portfolio performance. We now test the performance of real-time portfolios constructed based on the abovementioned recursive estimates. Similar to the in-sample results, we compare the spike-and-slab prior with the Bayesian lasso (Park and Casella, 2008) (Blasso), the horseshoe (Carvalho et al., 2009) (HS), and an equal-weight portfolio (EW). Each prior performance is benchmarked against the original Brandt et al. (2009) (BSV) approach based on value, size, and momentum as firm characteristics.

Table 2 shows the results. The out-of-sample performance of the portfolio allocations confirms some of the in-sample portfolio properties and the intuition highlighted in Figure 8. A heavier-tailed prior generates portfolios that, on average, are less stable, take more extreme positions, and are less diversified. Yet, calibrating the prior tails based on transaction costs produces portfolios closer to a more restrictive normal spike-and-slap prior. Alternative shrinkage approaches, such as the Blasso and the HS, generate more extreme portfolio weights than GLP and GLP-t(100), and in the case of the Blasso also a much higher turnover.

Unsurprisingly, the out-of-sample performance is lower than the in-sample. Panel B shows that the Sharpe ratio from the GLP-t(100) and the GLP priors are 2.3 annualised compared to the 3.8 in-sample. The performance fees are also substantially lower. For instance, an investor would pay a 2.9% monthly fee to access the real-time portfolio from the GLP-t(100) prior. This is compared to the 6.5% for the in-sample fee. Yet, the findings that a more sparsity-inducing prior does not maximise an investor's utility hold out-of-sample. The Sharpe ratio and the performance fees are substantially lower for the heavy-tailed prior GLP-t(4) compared to, for example, GLP-t(100) or GLP.

	EW	BSV	Blasso	HS	GLP	GLP-t(ν)						
						4	10	30	100	TC1	TC2	TC3
Panel A: Portfolio weights statistics												
Max w (%)	0.044	2.217	17.797	14.831	10.563	25.092	16.706	12.670	11.296	12.107	15.350	11.910
Min w (%)	0.044	-1.803	-16.527	-7.490	-6.784	-11.899	-8.681	-7.376	-6.995	-7.230	-8.225	-7.188
Mean |w| (%)	0.044	0.273	1.115	0.347	0.473	0.541	0.455	0.437	0.461	0.449	0.453	0.448
Mean (w < 0) (%)	0.000	-0.301	-1.156	-0.342	-0.490	-0.521	-0.437	-0.441	-0.473	-0.460	-0.437	-0.458
Prop w < 0	0.000	0.383	0.466	0.459	0.450	0.506	0.497	0.465	0.458	0.460	0.494	0.459
Herfindahl	0.000	0.030	0.675	0.122	0.148	0.329	0.189	0.148	0.148	0.149	0.176	0.147
Turnover	0.074	1.119	16.764	7.744	8.590	12.005	9.760	8.855	8.676	8.918	9.586	8.763
Panel B: Portfolio performance												
Mean	0.010	0.025	0.073	0.049	0.066	0.063	0.058	0.062	0.067	0.063	0.061	0.064
Sd	0.054	0.073	0.154	0.079	0.097	0.120	0.100	0.096	0.098	0.095	0.100	0.096
Skew	-0.433	0.329	-0.075	0.156	0.009	0.150	0.107	0.419	0.173	0.201	0.042	0.389
ES(5%)	-0.124	-0.127	-0.257	-0.120	-0.149	-0.192	-0.155	-0.134	-0.146	-0.137	-0.161	-0.131
SR	0.565	1.148	1.631	2.091	2.326	1.794	1.955	2.209	2.327	2.256	2.069	2.280
p-value ΔSR	0.005		0.164	0.001	0.000	0.055	0.009	0.001	0.000	0.001	0.003	0.000
Fees	-		-	0.021	0.029	0.010	0.018	0.026	0.029	0.027	0.022	0.028
Panel C: Portfolio performance with transaction costs												
Mean	0.010	0.020	-0.007	0.011	0.025	0.005	0.010	0.019	0.025	0.020	0.015	0.021
Sd	0.055	0.073	0.141	0.078	0.091	0.116	0.097	0.089	0.091	0.087	0.093	0.088
Skew	-0.439	0.228	-0.217	0.001	-0.241	-0.030	-0.086	-0.013	-0.132	-0.090	-0.202	0.096
ES(5%)	-0.124	-0.137	-0.318	-0.161	-0.192	-0.262	-0.210	-0.181	-0.190	-0.182	-0.209	-0.176
SR	0.572	0.882	-0.208	0.458	0.907	0.111	0.331	0.700	0.901	0.735	0.511	0.803
p-value ΔSR	0.113		0.002	0.119	0.936	0.016	0.057	0.530	0.950	0.629	0.204	0.791
Fees	-		-	-	0.001	-	-	-	0.001	-	-	-

Table 2: **Out-of-sample portfolio performance**. This table reports the out-of-sample performance of the parametric portfolio choice based on different prior specifications. The sample period is from January 1985 to November 2022. The portfolio is implemented in real time based on a 240-month rolling window. Panel A reports statistics for the portfolio weights, whereas Panel B and C report performance metrics with and without transaction costs. The latter are proxied based on half bid-ask spread for each asset in the portfolio. We report the implied fees f only if they are positive. In addition to the GLP and GLP-t priors with different ν calibrations, we also consider the benchmark portfolio, i.e., $1/N_t$, two alternative shrinkage priors such as the Horseshoe (HS) and the Bayesian lasso (Blasso), and the original implementation of Brandt et al. (2009) (BSV).

When factoring transaction costs into the real-time implementation, the out-of-sample performance deteriorates to the point that BSV becomes a challenging benchmark to beat. This is due to a substantially lower turnover. Although larger economically, the Sharpe ratio obtained from the GLP and GLP-t(100) priors are statistically indistinguishable from BSV. Nevertheless, GLP and GLP-t(100) can still generate a positive, albeit small, performance fee of 0.1% monthly. This is substantially better than a heavier-tailed spike-and-slab prior; the GLP-t(4) produces a rather dismal out-of-sample Sharpe ratio of 0.11 annualised, significantly smaller than the 0.88 from BSV, and a negative performance fee.

Panel A of Table 2 shows that some prior specifications imply a large spread between the largest long and short positions. This may limit their actual usefulness under common liquidity and diversification constraints (e.g., Patton and Weller, 2020). To mitigate this issue, we investigate the out-of-sample performance of each prior specification when capping the portfolio weights $w_{i,t} \in (-3\%, 3\%), \forall i, t$. Table 3 reports the results. Without transaction costs (Panel B), the GLP-t(4) prior produces considerably lower annualised risk-adjusted returns (SR=1.82) compared to GLP-t(100) (SR=2.38), GLP (SR=2.41), and all of the calibrations based on transaction costs (SR from 2.05 for GLP-t(TC2) to 2.32 for GLP-t(TC3)). The performance fees to access a spike-and-slab prior also favour more restrictive GLP and GLP-t(100) specifications.

Despite the weight cap, the BSV remains a competitive benchmark when considering transaction costs, as shown in Panel C. Only GLP produces higher risk-adjusted returns (SR=0.97) and positive performance fees (Fees=0.3% monthly) compared to BSV. The GLP-t(100) is on par with BSV as far as risk-adjusted returns are concerned. Heavier-tailed priors substantially underperform GLP with negative risk-adjusted returns (SR=-0.181 for GLP-t(4)) and negative performance fees. The Bayesian lasso produces the lowest performance with an out-of-sample SR of -0.374.

Appendix **??** reports the results of two additional exercises. Table **??** reports the portfolio performance with no-short sales constraints, i.e., $w_{i,t} \geq 0, \forall i, t$. The turnover is substantially smaller and comparable across all priors. As a result, performance differences net of transaction costs are rather flat. However, a heavy tail prior produces smaller risk-adjusted returns and performance fees than more restrictive GLP-t(100) and GLP priors. Table **??** reports

	EW	BSV	Blasso	HS	GLP	GLP-t(ν)						
						4	10	30	100	TC1	TC2	TC3
Panel A: Portfolio weights statistics												
Max w (%)	0.044	2.217	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000
Min w (%)	0.044	-1.803	-3.000	-3.000	-3.000	-3.000	-3.000	-3.000	-3.000	-3.000	-3.000	-3.000
Mean |w| (%)	0.044	0.273	0.975	0.330	0.458	0.490	0.428	0.419	0.445	0.432	0.430	0.431
Mean (w < 0) (%)	0.000	-0.301	-1.005	-0.323	-0.473	-0.468	-0.410	-0.422	-0.455	-0.441	-0.413	-0.440
Prop w < 0	0.000	0.383	0.464	0.437	0.444	0.473	0.479	0.456	0.450	0.451	0.477	0.450
Turnover	0.074	1.119	14.595	7.484	8.344	11.112	9.328	8.568	8.416	8.630	9.193	8.491
Panel B: Portfolio performance												
Mean	0.010	0.025	0.056	0.042	0.061	0.048	0.049	0.056	0.061	0.056	0.053	0.058
Sd	0.054	0.073	0.118	0.068	0.086	0.090	0.086	0.086	0.087	0.083	0.087	0.085
Skew	-0.433	0.329	0.070	0.037	-0.021	0.265	0.015	0.562	0.164	0.194	0.109	0.375
ES(5%)	-0.124	-0.127	-0.196	-0.107	-0.133	-0.131	-0.137	-0.123	-0.132	-0.124	-0.139	-0.119
SR	0.565	1.148	1.621	2.097	2.405	1.818	1.912	2.201	2.385	2.307	2.051	2.324
p-value ΔSR	0.005		0.142	0.001	0.000	0.032	0.008	0.000	0.000	0.000	0.002	0.000
Fees	-		-	0.021	0.029	0.010	0.018	0.026	0.029	0.027	0.022	0.028
Panel C: Portfolio performance with transaction costs												
Mean	0.010	0.020	-0.011	0.013	0.024	-0.004	0.005	0.015	0.021	0.016	0.017	0.018
Sd	0.055	0.073	0.108	0.078	0.082	0.088	0.084	0.079	0.080	0.076	0.094	0.077
Skew	-0.439	0.228	-0.264	0.028	-0.268	-0.078	-0.277	-0.048	-0.223	-0.182	-0.178	0.004
ES(5%)	-0.124	-0.137	-0.255	-0.159	-0.175	-0.202	-0.190	-0.167	-0.174	-0.168	-0.206	-0.164
SR	0.572	0.882	-0.374	0.541	0.966	-0.181	0.153	0.605	0.856	0.669	0.608	0.740
p-value ΔSR	0.113		0.000	0.209	0.777	0.000	0.006	0.304	0.928	0.452	0.349	0.614
Fees	-		-	-	0.003	-	-	-	-	-	-	-

Table 3: **Out-of-sample portfolio performance with capped weights**. This table reports the out-of-sample performance of the parametric portfolio choice *with capped weights*, i.e., $w_{i,t} \in (-3\%, 3\%) \forall i, t$, based on different prior specifications. The sample period is from January 1985 to November 2022. The portfolio is implemented in real time based on a 240-month rolling window. Panel A reports statistics for the portfolio weights, whereas Panel B and C report performance metrics with and without transaction costs. The latter are proxied based on half bid-ask spread for each asset in the portfolio. We report the implied fees f only if they are positive. In addition to the GLP and GLP-t priors with different ν calibrations, we also consider the benchmark portfolio, i.e., $1/N_t$, two alternative shrinkage priors such as the Horseshoe (HS) and the Bayesian lasso (Blasso), and the original implementation of Brandt et al. (2009) (BSV).

the results based on an alternative threshold whereby we exclude a given characteristic if its posterior inclusion probability is less than $1 - \text{mode}(q)$. This implies that the number of characteristics that enter the portfolio rule decreases as sparsity increases. The GLP is the only prior specification generating a higher risk-adjusted return than BSV – although the difference is not statistically significant – and positive performance fees (0.2% monthly).

Spanning regressions. We now further compare the economic value of different priors vs the baseline BSV approach based on spanning regressions of the form $r^{\mathcal{M}}_{t+1} = \alpha + \beta r^{BSV}_{t+1} + \epsilon_{t+1}$ where $r^{\mathcal{M}}_{t+1}$ is the return of a parametric portfolio based on the prior specification and r^{BSV}_{t+1} is the return on the BSV implementation (e.g., Moreira and Muir, 2017). The economic implication of $\alpha > 0$ is that a parametric portfolio based on a given prior specification expands the mean-variance efficient frontiers compared to a stand-alone investment in BSV characteristic-based portfolios (e.g., Gibbons et al., 1989). This follows from a direct link between spanning tests and mean-variance portfolio optimisation.

	Blasso	**HS**	**GLP**	**GLP-t(ν)**						
				4	10	30	100	TCI	TCII	TCIII
Panel A: Portfolio returns										
$\alpha(\%)$	7.474	3.911	5.919	5.837	4.728	5.190	5.910	5.468	5.064	5.603
p-value α	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
β	-0.053	0.379	0.287	0.189	0.407	0.402	0.317	0.315	0.406	0.322
p-value β	0.730	0.000	0.030	0.330	0.002	0.002	0.021	0.017	0.002	0.019
AR	0.487	0.530	0.624	0.491	0.496	0.572	0.620	0.598	0.530	0.604
$Adj R^2$	-0.004	0.119	0.042	0.009	0.084	0.090	0.051	0.055	0.083	0.056
Panel B: Portfolio returns with transaction costs										
$\alpha(\%)$	-0.651	0.401	1.935	0.095	0.255	0.788	1.875	1.368	0.705	1.549
p-value α	0.516	0.569	0.007	0.923	0.747	0.290	0.008	0.039	0.320	0.020
β	-0.050	0.372	0.274	0.191	0.396	0.355	0.300	0.300	0.394	0.305
p-value β	0.699	0.001	0.023	0.307	0.002	0.004	0.018	0.013	0.001	0.015
AR	-0.046	0.055	0.220	0.008	0.027	0.091	0.213	0.162	0.080	0.182
$Adj R^2$	-0.004	0.117	0.044	0.010	0.083	0.077	0.053	0.058	0.090	0.059

Table 4: **Spanning regressions based on out-of-sample portfolio returns**. This table reports the results of spanning regressions of the form $r^{\mathcal{M}}_{t+1} = \alpha + \beta r^{BSV}_{t+1} + \epsilon_{t+1}$ where $r^{\mathcal{M}}_{t+1}$ is the return of a parametric portfolio based on the prior specification and r^{BSV}_{t+1} is the return on the BSV implementation. The table reports the $\alpha(\%)$, the β and the corresponding p-values. In addition, the table reports the appraisal ratio α/σ_ϵ where σ_ϵ is the standard deviation of the estimated residual $\widehat{\epsilon}_{t+1}$. Panel A reports the results based on returns with no transaction costs, whereas Panel B reports the results with returns after transaction costs as in Eq.(8). In addition to the GLP and GLP-t priors with different ν calibrations, we also consider the two alternative shrinkage priors such as the Horseshoe (HS) and the Bayesian lasso (Blasso).

Table 4 reports the $\alpha(\%)$, the β, the appraisal ratio α/σ_ϵ, where σ_ϵ is the standard deviation of the estimated residual $\widehat{\epsilon}_{t+1}$, and the adjusted R^2 of the regressions. When considering out-of-sample returns with no transaction costs (Panel A), the spanning regressions suggest that Bayesian priors provide parametric portfolios that expand the mean-variance frontier compared to BSV. The alphas are large and statistically significant at a conventional 1% confidence level. The GLP-t(100) and GLP priors produce the highest appraisal ratios, meaning that higher performance does not come at the cost of a higher idiosyncratic risk.

Unsurprisingly, when considering transaction costs, the excess returns with respect to BSV of different priors tend to decrease (see Panel B). Yet, GLP, GLP-t(4), GLP-t(TC1), and GLP-t(TC2) can all generate positive and significant alphas from 1.4% when $\nu = TC1$ to 1.9% for GLP. On the other hand, heavier-tailed spike-and-slabs produce small and statistically insignificant alphas. For instance, GLP-t(4) generates a negligible 0.09% alpha with a p-value of 0.92. Similar to the returns with no transaction costs, GLP and GLP-t(100) also produce the highest appraisal ratios compared to BSV, with 0.22 and 0.21 ratios, respectively.

Overall, Table 4 confirms that more sparsity-inducing priors generate a rather discouraging performance, especially when transaction costs are considered. Instead, a more conservative prior with less heavy tails, i.e. higher values of ν, – perhaps calibrated based on transaction costs – produces significantly better mean-variance risk-adjusted portfolio returns.

4 Conclusions

We are interested in the role of firm characteristics in predicting the return variation for a large cross-section of stocks. To this end, we leverage the flexibility of a Bayesian variable selection prior and offer a nuanced understanding of how firm characteristics influence optimal portfolio choices by separately controlling for sparsity and shrinkage. The main results suggest that model uncertainty is pervasive, and there is little evidence in favour of sparsity from a purely portfolio allocation perspective.

Our empirical analysis reveals that the marginal posterior distributions of the sparsity and shrinkage parameters vary significantly based on the prior specification. Specifically, a heavier-tailed prior induces greater sparsity and diminishes uncertainty about which firm characteristics matter. However, this approach also leads to more concentrated portfolios with extreme weights and high turnover, raising concerns about the implications of sparsity for portfolio diversification and transaction costs. Our results suggest that sparsity can reduce model uncertainty, but it does not automatically translate into superior economic performance, especially when considering transaction costs.

In conclusion, we provide evidence that the choice between sparse and dense models has significant implications for predicting cross-sectional stock returns and portfolio allocation. Overall, our findings reveal that incorporating a broader set of stock characteristics enhances portfolio performance and diversification, particularly in the presence of transaction costs.

References

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.

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

Bali, T. G. and Hovakimian, A. (2009). Volatility spreads and expected stock returns. *Management Science*, 55(11):1797–1812.

Bańbura, 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.

Barbee Jr, W. C., Mukherji, S., and Raines, G. A. (1996). Do sales–price and debt–equity explain stock returns better than book–market and firm size? *Financial Analysts Journal*, 52(2):56–60.

Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, 32(3):870–897.

Bessembinder, H. and Venkataraman, K. (2010). Bid-ask spreads: Measuring trade execution costs in financial markets. *Encyclopedia of quantitative finance*, pages 184–190.

Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2015). Dirichlet–laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490.

Bradshaw, M. T., Richardson, S. A., and Sloan, R. G. (2006). The relation between corporate financing activities, analysts' forecasts and stock returns. *Journal of accounting and economics*, 42(1-2):53–85.

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.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671.

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.

Chan, L. K., Jegadeesh, N., and Lakonishok, J. (1996). Momentum strategies. *The Journal of finance*, 51(5):1681–1713.

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.

Chordia, T., Subrahmanyam, A., and Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of financial Economics*, 59(1):3–32.

Dechow, P. M., Hutton, A. P., Meulbroek, 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.

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.

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

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

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152.

Green, J., Hand, J. R. M., and Zhang, X. F. (2017). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies*, 30(12):4389–4436.

Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188.

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.

Heston, S. L. and Sadka, R. (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, 87(2):418–445.

Hjalmarsson, E. and Manchev, P. (2012). Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance*, 36(5):1392–1401.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *The review of financial studies*, 20(4):1113–1138.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

Ishwaran, H. and Rao, J. S. (2005). Spike and slab variable selection: Frequentist and bayesian strategies. *The Annals of Statistics*, 33(2):730–773.

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2022). Is there a replication crisis in finance? *The Journal of Finance*.

Kelly, B. T., Malamud, S., and Zhou, K. (2022). The virtue of complexity everywhere. *Available at SSRN 4166368*.

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.

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.

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612.

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, 72(4):1611–1644.

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics*, 104(1):162–185.

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.

Ročková, V. and George, E. I. (2014). Emvs: The em approach to bayesian variable selection. *Journal of the American Statistical Association*, 109(506):828–846.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting review*, pages 289–315.

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.