Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models*

Svetlana Bryzgalova† Jiantao Huang‡ Christian Julliard§

July 7, 2020

Abstract

We propose a novel framework for analyzing linear asset pricing models: simple, robust, and applicable to high dimensional problems. For a (potentially misspecified) standalone model, it provides reliable risk premia estimates of both tradable and non-tradable factors, and detects those weakly identified. For competing factors and (possibly non-nested) models, the method automatically selects the best specification – *if* a dominant one exists – or provides a model averaging, if there is no clear winner given the data. We analyze 2.25 quadrillion models generated by a large set of existing factors, and gain novel insights on the empirical drivers of asset returns.

Keywords: Cross-sectional asset pricing, factor models, model evaluation, multiple testing, data mining, p-hacking, Bayesian methods, shrinkage, SDF.
JEL codes: G12, C11, C12, C52, C58.

*Any errors or omissions are the responsibility of the authors. Christian Julliard thanks the Economic and Social Research Council (UK) [grant number: ES/K002309/1] for financial support. For helpful comments, discussions and suggestions, we thank Caio Almeida, Mikhail Chernov, Pierre Collin-Dufresne, Aureo de Paula, Marcelo Fernandes, Rodrigo Guimaraes, Raymond Kan, Frank Kleibergen, Lars Lochstoer, Albert Marcet, Alexander Michaelides, Olivier Scaillet, George Tauchen, Motohiro Yogo, and seminar and conference participants at the Fourth International Workshop in Financial Econometrics, ICEF Moscow, Goethe University Frankfurt, London Business School, London School of Economics, Second David Backus Memorial Conference on Macro-Finance, SoFiE Online seminar, University College London, and Princeton University.

†London Business School, sbryzgalova@london.edu

‡Department of Finance, London School of Economics, J.Huang27@lse.ac.uk

§Department of Finance, FMG, and SRC, London School of Economics, and CEPR; c.julliard@lse.ac.uk.

I Introduction

In the last decade or so, two observations have come to the forefront of the empirical asset pricing literature. First, at current production rates, in the near future we will have more sources of empirically "identified" risk than stock returns to price with these factors – the so-called factors zoo phenomenon (see, e.g., Harvey, Liu, and Zhu (2016)). Second, given the commonly used estimation methods in empirical asset pricing, useless factors (i.e., factors whose true covariance with asset returns is asymptotically zero) are not only likely to appear empirically relevant but also invalidate inference regarding the true sources of risk (see, e.g., Gospodinov, Kan, and Robotti (2019)). Nevertheless, to the best of our knowledge, no general method has been suggested to date that: *i*) is applicable to both tradable and non-tradable factors, *ii*) can handle the very large factor zoo, and *iii*) remains valid under model misspecification, while *iv*) being robust to the spurious inference problem. And that is exactly what we provide.

We develop a *unified framework* for tackling linear asset pricing models. In the case of standalone model estimation, our method provides reliable risk premia estimates (for both tradable and non-tradable factors), hypothesis testing and confidence intervals for these parameters, as well as confidence intervals for all the possible objects of interest, for example, alphas, measures of fit, and model-implied Sharpe ratios. The approach naturally extends to model comparison and factor selection, even when all models are misspecified and non-nested. Furthermore, it endogenously delivers a specification selection – *if* a dominant model exists – or model averaging, if there is no clear winner given the data at hand. The method is numerically simple, fast, extremely easy to use, and can be feasibly applied to a very large number (literally, quadrillions) of candidate factor models, while being robust to the common identification problems.

As stressed by Harvey (2017) in his AFA presidential address, the zoo of factors in the empirical literature naturally calls for a Bayesian solution – and we develop one. Furthermore, we show that factors proliferation and spurious inference are tightly connected problems, and a naïve Bayesian approach to model selection fails in the presence of spurious factors. Hence, we correct it, and apply our method to the zoo of traded and non-traded factors proposed in the literature, jointly evaluating 2.25 quadrillion models and gaining novel insights regarding the empirical drivers of asset returns. Our results are based on the beta-representation of linear factor models, but the method is straightforwardly extendable to the direct estimation of the stochastic discount factor (SDF).

We find that only a handful of factors proposed in the previous literature are robust explanators of the cross-section of asset returns, and a three *robust* factor model easily outperforms canonical models. Nevertheless, there is no clear "winner" across the whole space of potential models: hundreds of possible specifications, none of which has been examined in the previous literature, are virtually equally likely to price the cross-section of returns. Furthermore, we find that the "true" latent SDF is dense in the space of observable factors, that is, a large subset of the factors proposed in the literature is needed to fully capture its pricing implications. Nonetheless, the SDF-implied maximum Sharpe ratio in the economy is not unrealistically high, suggesting substantial commonality among the risks spanned by the factors in the zoo.

Our contribution is fourfold. First, we develop a very simple Bayesian version of the canonical Fama and MacBeth (1973) regression method that is applicable to both traded and non-traded factors. This approach makes useless factors easily detectable in finite sample, while delivering sharp posteriors for the strong factors' risk premia (i.e., leaving inference about them unaffected). The result is quite intuitive. Useless factors make frequentist inference unreliable since, when factor exposures go to zero, risk premia are no more identified. We show that exactly the same phenomenon causes the Bayesian posterior credible intervals of risk premia to become diffuse and centered at zero, which makes them easily detectable in empirical applications. This robust inference approach is as easy to implement as the canonical Shanken (1992) correction of the standard errors.

Second, the main intent of this paper is to provide a method for handling inference on the *entirety* of the factor zoo at once. This naturally calls for the use of model (and factor) posterior probabilities. However, as we show, model and factor selection based on marginal likelihoods (i.e., on posterior probabilities or Bayes factors) is unreliable under flat priors for risk premia: asymptotically, weakly identified factors get selected with probability one even if they do not command any risk premia. This is due to the fact that lack of identification generates an unbounded manifold for the risk premia parameters, over which the likelihood surface is totally flat.[1] Hence, integrating such a likelihood under a flat prior produces improper marginals that select useless factors with probability tending to one. As a result, in the presence of identification failure, naïve Bayesian inference has the same weakness as the frequentist one. This observation, however, not only illustrates the nature of the problem; it also suggests how to restore inference: use suitable, non-informative – but yet non-flat – priors.

Third, building upon the literature on predictor selection (see, e.g., Ishwaran, Rao, et al. (2005) and Giannone, Lenza, and Primiceri (2018)), we provide a novel (continuous) "spike-and-slab" prior that restores the validity of model and factor selection based on posterior model probabilities and Bayes factors. The prior is economically motivated – it has a direct mapping to beliefs about the Sharpe ratio of the risk factors. It is uninformative (the "slab") for strong factors, but shrinks away (the "spike") useless factors. This approach is

[1]This is similar to the effect of "weak instruments" in IV estimations, as discussed in Sims (2007).

similar in spirit to a ridge regression and acts as a (Tikhonov-Phillips) regularization of the likelihood function of the cross-sectional regression needed to estimate factor risk premia. A distinguishing feature of our prior is that the prior variance of a factor's risk premium is proportional to its correlation with the test asset returns. Hence, when a useless factor is present, the prior variance of its risk premium converges to zero, so the shrinkage dominates and forces its posterior distribution to concentrate around zero. Not only does this prior restore integrability, but it also: *i*) makes it computationally feasible to analyze quadrillions of alternative factor models; *ii*) allows the researcher to encode prior beliefs about the sparsity of the true SDF without imposing hard thresholds; *iii*) restores the validity of hypothesis testing; *iv*) shrinks the estimate of useless factors' risk premia toward zero. We regard this novel spike-and-slab prior approach as a solution for the high-dimensional inference problem generated by the factor zoo.

Our method is easy to implement and, in all of our simulations, has good finite sample properties, even when the cross-section of test assets is large. We investigate its performance for risk premia estimation, model evaluation, and factor selection, in a range of simulation designs that mimic the stylized features of returns. Our simulations account for potential model misspecification and the presence of either strong or useless factors in the model. The use of posterior sampling naturally allows to build credible confidence intervals not only for risk premia, but also other statistics of interest, such as the cross-sectional R^2, which is notoriously hard to estimate precisely (Lewellen, Nagel, and Shanken (2010)). We show that whenever risk premia are well identified, both our method and the frequentist approach provide valid confidence intervals for model parameters, with empirical coverage being close to its nominal size. However, in the presence of useless factors, canonical frequentist inference becomes unreliable. Instead, the posterior distributions of useless factors' risk premia are reliably centered around zero, which quickly reveals them even in a relatively short sample. We find that the Bayesian estimation of strong factors is largely unaffected by the identification failure, with posterior-based confidence intervals corresponding to their nominal size. In other words, our Bayesian approach provides sound inference on model parameters. Furthermore, we also illustrate the factor (and model) selection pitfalls generated by flat priors for risk premia, and show that our spike-and-slab prior successfully eliminates spurious factors, while retaining the true sources of risk, even in a relatively short sample.

Fourth, our results have important empirical implications for the estimation of popular linear factor models and their comparison. We examine 51 factors proposed in the previous literature, yielding a total of 2.25 quadrillion possible models to analyze, and we find that only a handful of variables are robust explanators of the cross-section of asset returns (the Fama and French (1992) "high-minus-low" proxy for the value premium, as well the adjusted versions of both market and "small-minus-big" size factors of Daniel, Mota, Rottke, and

Santos (2020)).

Jointly, the three robust factors provide a model that is, compared to the previous empirical literature, one order of magnitude more likely to have generated the observed asset returns: its posterior probability is about 85-88%, while the most likely model among the ones previously proposed (the five-factor model of Fama and French (2016)) has a posterior probability of about 2 – 6%. However, when considering the *whole* space of potential models, there is no clear best model: hundreds of possible specifications, none of which has been proposed before, are equally likely to price the cross-section of returns. This is due to the fact that, as we show, the "true" latent SDF is dense in the space of factors proposed in the empirical literature: capturing the characteristics of the SDF requires the use of 24-25 observable factors. Nevertheless, the SDF-implied maximum Sharpe ratio is not excessive, indicating a high degree of commonality, in terms of captured risks, among the factors proposed in the empirical literature.

Furthermore, we apply our useless factors detection method to a selection of popular linear SDF models. We find that a variety of models with both tradable and non-tradable factors are only weakly identified at best and are characterized by a substantial degree of model misspecification and uncertainty.

While most of our results are obtained on the joint cross-section of 25 Fama-French portfolios, sorted by size and value, and 30 industry portfolios, we have also analyzed other test assets used in the empirical literature. In aggregating estimation output across these different portfolios, we largely rely on a "*revealed preferences*" approach. That is, we carefully survey test assets used by the existing literature and focus on the most popular (and, arguably, most salient) cross-sections used in the papers. Based on the empirical frequency of particular portfolios used in the literature, we build a set of 25 composite cross-sections and average our findings across all of them, with weights proportional to the frequency of their empirical use. Despite using different test assets, we still find that HML and the adjusted version of the market factor by Daniel, Mota, Rottke, and Santos (2020) are robust explanators of the cross-section of returns.

Finally, since traditional cross-sectional asset pricing is often criticized on the grounds of relying on only in-sample estimates and model evaluation, we also conduct out-of-sample analysis. We split our sample in two parts, estimate the model in each of them, and evaluate the forecast for the cross-sectional spread of returns on the half of the data not used for estimation. We find that our model out-of-sample performance is remarkably stable for a wide range of parameters. Furthermore, we show that the key to this finding is a combination of the overall degree of factors shrinkage, and the ability to successfully separate strong and weak factors, that our prior design generates.

The remainder of the paper is organized as follows. First, we focus on the related literature and our contribution to it, correspondingly. In Section II we proceed to outline the Bayesian Fama-MacBeth estimation and its properties for individual factor inference, and selection/averaging. Section III describes the simulation designs, highlighting both small and large-sample behavior of our method and explaining how it relates to the traditional frequentist approach. Section IV presents most of the empirical results for standalone popular linear models (Section IV.1), and provides further insights from sampling the whole candidate model space (Sections IV.2 – IV.5). We analyze cross-section uncertainty and out-of-sample performance in Sections IV.6 and IV.7, respectively. Finally, we discuss potential extensions to our procedure in Section V, and Section VI concludes.[2]

I.1 Closely related literature

There are numerous contributions to the literature that rely on the use of Bayesian tools in finance, especially in the areas of asset allocation (for an excellent overview, see Fabozzi, Huang, and Zhou (2010), and Avramov and Zhou (2010)), model selection (e.g., Barillas and Shanken (2018)), and performance evaluation (Baks, Metrick, and Wachter (2001), Pástor and Stambaugh (2002), Jones and Shanken (2005), Harvey and Liu (2019)). Therefore, we aim to provide only an overview of the literature that is most closely related to our paper.

While there are multiple papers in the literature that adopt a Bayesian approach to analyze linear factor models and portfolio choice, most of them focus on the time series regressions, where the intercepts, thanks to factors being directly traded (or using their mimicking portfolios) can be interpreted as the vector of pricing errors – the $\boldsymbol{\alpha}$'s. In this case, the use of test assets actually becomes irrelevant, since the problem of comparing model performance is reduced to the spanning tests of one set of factors by another (e.g., Barillas and Shanken (2018)).

Our paper instead develops a method that can be applied to both tradable and non-tradable factors. As a result, we focus on the general pricing performance in the cross-section of asset returns (which is no longer irrelevant) and show that there is a tight link between the use of the most popular, diffuse, priors for the risk premia, and the failure of the standard estimation techniques in the presence of useless factors (e.g., Kan and Zhang (1999a)).

Shanken (1987) and Harvey and Zhou (1990) were probably the first to contribute to the literature that adapted the Bayesian framework to the analysis of portfolio choice, and developed GRS-type tests (cf. Gibbons, Ross, and Shanken (1989)) for mean-variance efficiency. While Shanken (1987) was the first to examine the posterior odds ratio for portfolio alphas in the linear factor model, Harvey and Zhou (1990) set the benchmark by imposing

[2]Additional results are reported in the Online Appendix available at: `https://ssrn.com/abstract=3627010`

the priors, both diffuse and informative, directly on the deep model parameters.

Pástor and Stambaugh (2000) and Pástor (2000) directly assign a prior distribution to the vector of pricing errors $\boldsymbol{\alpha}$, $\boldsymbol{\alpha} \sim \mathcal{N}(0, \kappa\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix of returns and $\kappa \in \mathbb{R}_{+}$, and apply it to the Bayesian portfolio choice problem. The intuition behind their prior is that it imposes a degree of shrinkage on the alphas, so whenever factor models are misspecified, the pricing errors cannot be too large a priori, hence, placing a bound on the Sharpe ratio achievable in this economy. Therefore, they argue, a diffuse prior for the pricing errors $\boldsymbol{\alpha}$ in general should be avoided.

Barillas and Shanken (2018) extend the aforementioned prior to derive a closed-form solution for the Bayes' factor in a setting in which all risk factors are tradable, and use it to compare different linear factor models exploiting the time series dimension of the data. Chib, Zeng, and Zhao (2020) show that the *improper* prior specification of Barillas and Shanken (2018) is problematic and propose a new class of priors that leads to valid comparison for traded factor models. In a recent paper, Goyal, He, and Huh (2018) also extend the notion of distance between alternative model specifications and highlight the tension between the power of the GRS-type tests and the absolute return-based measures of mispricing.

Last, but not least, there is a general close connection between the Bayesian approach to model selection or parameter estimation and the shrinkage-based one. Garlappi, Uppal, and Wang (2007) impose a set of different priors on expected returns and the variance-covariance matrix and find that the shrinkage-based analogue leads to superior empirical performance. Anderson and Cheng (2016) develop a Bayesian model-averaging approach to portfolio choice, with model uncertainty being one of the key ingredients that yields robust asset allocation, and superior out-of-sample performance of the strategies. Finally, the shrinkage-based approach to recovering the SDF of Kozak, Nagel, and Santosh (2019) can also be interpreted from a Bayesian perspective. Within a universe of characteristic-managed portfolios, the authors assign prior distributions to expected returns,[3] and their posterior maximum likelihood estimators resemble a ridge regression. Instead, we work directly with tradable and non-tradable factors and consider (endogenously) heterogenous priors for factor risk premia, $\boldsymbol{\lambda}$. The dispersion of our prior for each λ directly depends on the correlation between test assets and the factor, so that it mimics the strength of the identification of the factor risk premium.

Naturally, our paper also contributes to the very active (and growing) body of work that critically evaluates existing findings in the empirical asset pricing literature and tries to develop a robust methodology. There is ample empirical evidence that most linear asset

[3]Or, equivalently, the coefficients vector $\boldsymbol{b}$ when the linear stochastic discount factor is represented as $m_t = 1 - (\boldsymbol{f_t} - \mathbb{E}[\boldsymbol{f}_t])^{\top}\boldsymbol{b}$, where $\boldsymbol{f}_t$ and $\mathbb{E}$ denote, respectively, a vector of factors and the unconditional expectation operator.

pricing models are misspecified (e.g. Chernov, Lochstoer, and Lundeby (2019), He, Huang, and Zhou (2018)). Gospodinov, Kan, and Robotti (2014) develop a general approach for misspecification-robust inference that provides valid confidence interval for the pseudo-true values of the risk premia. Giglio and Xiu (2018) exploit the invariance principle of the PCA and recover the risk premium of a given factor from the projection on the span of latent factors driving a cross-section of asset returns. Uppal, Zaffaroni, and Zviadadze (2018) adopt a similar approach by recovering latent factors from the residuals of the asset pricing model, effectively completing the span of the SDF. Daniel, Mota, Rottke, and Santos (2020) instead focus on the construction of cross-sectional factors and note that many well-established tradable portfolios, such as HML and SMB, can be substantially improved in asset pricing tests by hedging their unpriced component (which does not carry a risk premium). We do not take a stand on the origin of the factors or the completion of the model space. Instead, we consider the whole universe of potential models that can be created from the set of observable candidates factors proposed in the empirical literature. As such, our analysis explicitly takes into account both standard specifications that have been successfully used in numerous papers (e.g., Fama-French three-factor model, or nondurable consumption growth), as well as the combinations of the factors that have never been explicitly tested.

Following Harvey, Liu, and Zhu (2016), a large body of literature has tried to understand which of the existing factors (or their combinations) drive the cross-section of asset returns. Giglio, Feng, and Xiu (2019) combine cross-sectional asset pricing regressions with the double-selection LASSO of Belloni, Chernozhukov, and Hansen (2014) to provide valid uniform inference on the selected sources of risk. Huang, Li, and Zhou (2018) use a reduced rank approach to select among not only the observable factors, but their total span, effectively allowing for sparsity not necessarily in the observable set of factors, but their combinations as well. Kelly, Pruitt, and Su (2019) build a latent factor model for stock returns, with factor loadings being a linear function of the company characteristics, and find that only a small subset of the latter provide substantial independent information relevant for asset pricing. Our approach does not take a stand on whether there exists a single combination of factors that substantially outperforms other model specifications. Instead, we let the data speak, and find out that the cross-sectional likelihood across the many models is rather flat, meaning the data is not informative enough to reliably indicate that there is a single dominant specification.

Finally, our paper naturally contributes to the literature on weak identification in asset pricing. Starting from the seminal papers of Kan and Zhang (1999a,b), identification of risk premia has been shown to be challenging for traditional estimation procedures. Kleibergen (2009) demonstrates that the two-pass regression of Fama-MacBeth lead to biased estimates of the risk premia and spuriously high significance levels. Moreover, useless factors often crowd out the impact of the true sources of risk in the model and lead to seemingly high levels of cross-sectional fit (Kleibergen and Zhan (2015)). Gospodinov, Kan, and Robotti (2014, 2019) demonstrate that most of the estimation techniques used to recover risk premia in the cross-section are invalidated by the presence of useless factors, and they propose alternative procedures that effectively eliminate the impact of these factors. We build upon the intuition developed in these papers and formulate the Bayesian solution to the problem by providing a prior that directly reflects the strength of the factor. Whenever the vector of correlation coefficients between asset returns and a factor is close to zero, the prior variance of λ for this specific factor also goes to zero, and the penalty for the risk premium converges to infinity, effectively shrinking the posterior of the useless factors' risk premia toward zero. Therefore, our priors are particularly robust to the presence of spurious factors. Conversely, they are very diffuse for strong factors, with the posterior reflecting the full impact of the likelihood.

II Inference in Linear Factor Models

This section introduces the notation and reviews the main results of the Fama-MacBeth (FM) regression method (see Fama and MacBeth (1973)). We focus on classic linear factor models for cross-sectional asset returns. Suppose that there are K factors, $\boldsymbol{f}_t = (f_{1t} \dots f_{Kt})^\top$, $t = 1, \dots T$, which could be either tradable or non-tradable. To simplify exposition, but without loss of generality, we consider demeaned factors so that we have both $\mathbb{E}[\boldsymbol{f}_t] = \mathbf{0}_{\boldsymbol{K}}$ and $\bar{\boldsymbol{f}} = \mathbf{0}_{\boldsymbol{K}}$, where $\mathbb{E}[.]$ denotes the unconditional expectation and the upper bar denotes the sample mean operator. The returns of N test assets, in excess of the risk-free rate, are denoted by $\boldsymbol{R}_t = (R_{1t} \dots R_{Nt})^\top$.

In the FM procedure, the factor exposures of asset returns, $\boldsymbol{\beta_f} \in \mathbb{R}^{N\times K}$, are recovered from the following linear regression:

$$\boldsymbol{R}_t = \boldsymbol{a} + \boldsymbol{\beta_f} \boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \tag{1}$$

where $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_T \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}_{\boldsymbol{N}}, \boldsymbol{\Sigma})$ and $\boldsymbol{a} \in \mathbb{R}^N$. Given the mean normalization of $\boldsymbol{f_t}$ we have $\mathbb{E}[\boldsymbol{R_t}] = \boldsymbol{a}$.

The risk premia associated with the factors, $\boldsymbol{\lambda_f} \in \mathbb{R}^K$, are then estimated from the cross-sectional regression:

$$\bar{\boldsymbol{R}} = \lambda_c \mathbf{1}_{\boldsymbol{N}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{f}} \boldsymbol{\lambda_f} + \boldsymbol{\alpha}, \tag{2}$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{f}}$ denotes the time series estimates, λ_c is a scalar average mispricing that should be equal to zero under the null of the model being correctly specified, $\mathbf{1}_{\boldsymbol{N}}$ denotes an N-dimensional vector of ones, and $\boldsymbol{\alpha} \in \mathbb{R}^N$ is the vector of pricing errors in excess of λ_c. If the model is correctly specified, it implies the parameter restriction: $\boldsymbol{a} = \mathbb{E}[\boldsymbol{R}_t] = \lambda_c \mathbf{1}_{\boldsymbol{N}} + \boldsymbol{\beta}_{\boldsymbol{f}} \boldsymbol{\lambda}_{\boldsymbol{f}}$. Therefore, we can rewrite the two-step FM regression into one equation as

$$\boldsymbol{R}_t = \lambda_c \mathbf{1}_{\boldsymbol{N}} + \boldsymbol{\beta}_{\boldsymbol{f}} \boldsymbol{\lambda}_{\boldsymbol{f}} + \boldsymbol{\beta}_{\boldsymbol{f}} \boldsymbol{f}_t + \boldsymbol{\epsilon}_t. \tag{3}$$

Equation (3) is particularly useful in our simulation study. Note that the intercept λ_c is included in (2) and (3) in order to separately evaluate the ability of the model to explain the average level of the equity premium and the cross-sectional variation of asset returns.

Let $\boldsymbol{B}^\top = (\boldsymbol{a}, \boldsymbol{\beta}_{\boldsymbol{f}})$ and $\boldsymbol{F}_t^\top = (1, \boldsymbol{f}_t^\top)$, and consider the matrices of stacked time series observations, $\boldsymbol{R} = (\boldsymbol{R}_{\mathbf{1}}, \ldots, \boldsymbol{R}_{\boldsymbol{T}})^\top$, $\boldsymbol{F} = (\boldsymbol{F}_{\mathbf{1}}, \ldots, \boldsymbol{F}_{\boldsymbol{T}})^\top$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_{\mathbf{1}}, \ldots, \boldsymbol{\epsilon}_{\boldsymbol{T}})^\top$. The regression in (1) can then be rewritten as $\boldsymbol{R} = \boldsymbol{F}\boldsymbol{B} + \boldsymbol{\epsilon}$, yielding the time series estimates of $(\boldsymbol{a}, \boldsymbol{\beta}_{\boldsymbol{f}})$ and $\boldsymbol{\Sigma}$ as follows:

$$\widehat{\boldsymbol{B}} = \left(\widehat{\boldsymbol{a}},\ \widehat{\boldsymbol{\beta}}_{\boldsymbol{f}}\right)^\top = (\boldsymbol{F}^\top \boldsymbol{F})^{-1} \boldsymbol{F}^\top \boldsymbol{R}, \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{T} (\boldsymbol{R} - \boldsymbol{F}\widehat{\boldsymbol{B}})^\top (\boldsymbol{R} - \boldsymbol{F}\widehat{\boldsymbol{B}}).$$

In the second step, the OLS estimates of the factor risk premia are

$$\widehat{\boldsymbol{\lambda}} = (\widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\beta}})^{-1} \widehat{\boldsymbol{\beta}}^\top \bar{\boldsymbol{R}}, \tag{4}$$

where $\widehat{\boldsymbol{\beta}} = (\mathbf{1}_{\boldsymbol{N}}\ \widehat{\boldsymbol{\beta}}_{\boldsymbol{f}})$ and $\boldsymbol{\lambda}^\top = (\lambda_c\ \boldsymbol{\lambda}_{\boldsymbol{f}}^\top)$. The canonical Shanken (1992) corrected covariance matrix of the estimated risk premia is[4]

$$\widehat{\boldsymbol{\sigma}}^2(\widehat{\boldsymbol{\lambda}}) = \frac{1}{T} \left[(\widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\beta}})^{-1} \widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\beta}} (\widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\beta}})^{-1} (1 + \widehat{\boldsymbol{\lambda}}_{\boldsymbol{f}}^\top \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^{-1} \widehat{\boldsymbol{\lambda}}_{\boldsymbol{f}}) + \widehat{\boldsymbol{V}}_{\boldsymbol{f}} \right], \widehat{\boldsymbol{V}}_{\boldsymbol{f}} = \begin{pmatrix} 0 & \mathbf{0}_{\boldsymbol{K}}^\top \\ \mathbf{0}_{\boldsymbol{K}} & \widehat{\Sigma}_f \end{pmatrix} \tag{5}$$

where $\widehat{\boldsymbol{\Sigma}}_f$ is the sample estimate of the variance-covariance matrix of the factors $\boldsymbol{f}_{\boldsymbol{t}}$. There are two sources of estimation uncertainty in the OLS estimates of $\boldsymbol{\lambda}$. First, we do not know the test assets' expected returns but instead estimate them as sample means, $\bar{\boldsymbol{R}}$. According to the time series regression, $\bar{\boldsymbol{R}} \sim \mathcal{N}(\boldsymbol{a}, \frac{1}{T}\boldsymbol{\Sigma})$ asymptotically. Second, if $\boldsymbol{\beta}$ is known, the asymptotic covariance matrix of $\widehat{\boldsymbol{\lambda}}$ is simply $\frac{1}{T}(\boldsymbol{\beta}^\top \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top \widehat{\boldsymbol{\Sigma}} \boldsymbol{\beta} (\boldsymbol{\beta}^\top \boldsymbol{\beta})^{-1}$. The extra term $(1 + \boldsymbol{\lambda}_{\boldsymbol{f}}^\top \boldsymbol{\Sigma}_{\boldsymbol{f}}^{-1} \boldsymbol{\lambda}_{\boldsymbol{f}})$ is included to account for the fact that $\boldsymbol{\beta}_{\boldsymbol{f}}$ is estimated.

Alternatively, we can run a (feasible) GLS regression in the second stage, obtaining the estimates

$$\widehat{\boldsymbol{\lambda}} = (\widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\beta}})^{-1} \widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \bar{\boldsymbol{R}}, \tag{6}$$

where $\widehat{\Sigma} = \frac{1}{T} \widehat{\boldsymbol{\epsilon}}^\top \widehat{\boldsymbol{\epsilon}}$ and $\widehat{\boldsymbol{\epsilon}}$ denotes the OLS residuals, and with the associated covariance matrix

[4] An alternative way (see e.g. Cochrane (2005), page 242) to account for the uncertainty from "generated regressors," such as $\widehat{\boldsymbol{\beta}}_{\boldsymbol{f}}$, is to estimate the whole system via GMM.

of the estimates

$$\widehat{\boldsymbol{\sigma}}^2(\widehat{\boldsymbol{\lambda}}) = \frac{1}{T}\left[(\widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\beta}})^{-1}(1+\widehat{\boldsymbol{\lambda}}_f^\top \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^{-1}\widehat{\boldsymbol{\lambda}}_{\boldsymbol{f}}) + \widehat{\boldsymbol{V}}_{\boldsymbol{f}}\right]. \tag{7}$$

Equations (4) and (6) make it clear that in the presence of a spurious (or useless) factor, that is, such that $\boldsymbol{\beta}_j = \frac{C}{\sqrt{T}}$, $C \in \mathbb{R}^N$, risk premia are no longer identified. Furthermore, their estimates diverge (i.e., $\widehat{\boldsymbol{\lambda}}_j \not\to 0$ as $T \to \infty$), leading to inference problems for both the useless and the strong factors (see, e.g., Kan and Zhang (1999b)). In the presence of such an identification failure, the cross-sectional R^2 also becomes untrustworthy. If a useless factor is included into the two-pass regression, the OLS R^2 tends to be highly inflated (although the GLS R^2 is less affected).[5]

This problem arises not only when using the Fama-MacBeth two-step procedure. Kan and Zhang (1999a) point out that the identification condition in the GMM test of linear stochastic discount factor models fails when a useless factor is included. Moreover, this leads to overrejection of the hypothesis of a zero risk premium for the useless factor under the Wald test, and the power of the over-identifying restriction test decreases. Gospodinov, Kan, and Robotti (2019) document similar problems within the maximum likelihood estimation and testing framework.

Consequently, several papers have attempted to develop alternative statistical procedures that are robust to the presence of useless factors. Kleibergen (2009) proposes several novel statistics whose large sample distributions are unaffected by the failure of the identification condition. Gospodinov, Kan, and Robotti (2014) derive robust standard errors for the GMM estimates of factor risk premia in the linear stochastic factor framework, and prove that t-statistics calculated using their standard errors are robust even when the model is misspecified and a useless factor is included. Bryzgalova (2015) introduces a LASSO-like penalty term in the cross-sectional regression to shrink the risk premium of the useless factor toward zero.

In this paper, we provide a Bayesian inference and model selection framework that *i*) can be easily used for robust inference in the presence, and detection, of useless factors (section II.1) and *ii*) can be used for both model selection, and model averaging, even in the presence of a very large number of candidate (traded or non-traded, and possibly useless) risk factors – that is, the entire factor zoo.

[5] For example, Kleibergen and Zhan (2015) derive the asymptotic distribution of the R^2 under the assumption that a few unknown factors are able to explain expected asset returns, and show that, in the presence of a useless factor, the OLS R^2 is more likely to be inflated than its GLS counterpart.

II.1 Bayesian Fama-MacBeth

This section introduces our hierarchical Bayesian Fama-MacBeth (BFM) estimation method. A formal derivation is presented in Appendix A.1.1. To start with, let's consider the time series regression. We assume that the time series error terms follow an iid multivariate Gaussian distribution (the approach, at the cost of analytical solutions, could be generalized to accommodate different distributional assumptions), that is $\boldsymbol{\epsilon} \sim \mathcal{MVN}(\mathbf{0}_{\boldsymbol{T}\times\boldsymbol{N}}, \boldsymbol{\Sigma} \otimes \boldsymbol{I_T})$. The time series likelihood of the data $(\boldsymbol{R}, \boldsymbol{F})$ is then

$$p(data|\boldsymbol{B}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{NT}{2}} |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2} tr \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{B})^\top (\boldsymbol{R} - \boldsymbol{F}\boldsymbol{B})\right]\right\}.$$

The time series regression is always valid even in the presence of a spurious factor. For simplicity, we choose the non-informative Jeffreys' prior for $(\boldsymbol{B}, \boldsymbol{\Sigma})$: $\pi(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{N+1}{2}}$. Note that this prior is flat in the $\boldsymbol{B}$ dimension. The posterior distribution of $(\boldsymbol{B}, \boldsymbol{\Sigma})$ is, therefore,

$$\boldsymbol{B}|\boldsymbol{\Sigma}, data \sim \mathcal{MVN}\left(\widehat{\boldsymbol{B}}_{ols}, \boldsymbol{\Sigma} \otimes (\boldsymbol{F}^\top \boldsymbol{F})^{-1}\right) \quad \text{and} \tag{8}$$

$$\boldsymbol{\Sigma}|data \sim \mathcal{W}^{-1}\left(T - K - 1, T\hat{\boldsymbol{\Sigma}}\right), \tag{9}$$

where $\widehat{\boldsymbol{B}}_{ols}$ and $\hat{\boldsymbol{\Sigma}}$ denote the canonical OLS based estimates, and $\mathcal{W}^{-1}$ is the inverse-Wishart distribution (a multivariate generalization of the inverse-gamma distribution). From the above, we can sample the posterior distribution of the parameters $(\boldsymbol{B}, \boldsymbol{\Sigma})$ by first drawing the covariance matrix $\boldsymbol{\Sigma}$ from the inverse-Wishart distribution conditional on the data, and then drawing $\boldsymbol{B}$ from a multivariate normal distribution conditional on the data and the draw of $\boldsymbol{\Sigma}$.

If the model is correctly specified, in the sense that all true factors are included, expected returns of the assets should be fully explained by their risk exposure, $\boldsymbol{\beta}$, and the prices of risk $\boldsymbol{\lambda}$, that is, $\mathbb{E}[\boldsymbol{R}_t] = \boldsymbol{\beta}\boldsymbol{\lambda}$. But since, given our mean normalization of the factors, $\mathbb{E}[\boldsymbol{R}_t] = \boldsymbol{a}$, we have the least square estimate $(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}$. Therefore, we can define our first estimator.

Definition 1 (Bayesian Fama-MacBeth (BFM)) *The posterior distribution of $\boldsymbol{\lambda}$ conditional on $\boldsymbol{B}$, $\boldsymbol{\Sigma}$ and the data, is a Dirac distribution at $(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}$. A draw $(\boldsymbol{\lambda}_{(j)})$ from the posterior distribution of $\boldsymbol{\lambda}$ conditional on the data only is obtained by drawing $\boldsymbol{B}_{(j)}$ and $\boldsymbol{\Sigma}_{(j)}$ from the Normal-inverse-Wishart (8)–(9), and computing $(\boldsymbol{\beta}_{(j)}^\top\boldsymbol{\beta}_{(j)})^{-1}\boldsymbol{\beta}_{(j)}^\top\boldsymbol{a}_{(j)}$.*

The posterior distribution of $\boldsymbol{\lambda}$ defined above accounts both for the uncertainty about the expected returns (via the sampling of $\boldsymbol{a}$) and the uncertainty about the factor loadings (via the sampling of $\boldsymbol{\beta}$). Note that, differently from the frequentist case in equation (5), there is no "extra term" $(1 + \boldsymbol{\lambda}_{\boldsymbol{f}}^\top \boldsymbol{\Sigma}_{\boldsymbol{f}}^{-1} \boldsymbol{\lambda}_{\boldsymbol{f}})$ to account for the fact that $\boldsymbol{\beta}_{\boldsymbol{f}}$ is estimated. The reason is that it is unnecessary to explicitly adjust standard errors of $\boldsymbol{\lambda}$ in the Bayesian approach, since we keep updating $\boldsymbol{\beta_f}$ in each simulation step, automatically incorporating the uncertainty about $\boldsymbol{\beta_f}$ into the posterior distribution of $\boldsymbol{\lambda}$. Furthermore, it is quite intuitive, from the definition above of the BFM estimator, that we expect posterior inference to detect weak and spurious factors in finite sample. For such factors, the near singularity of $(\boldsymbol{\beta}_{(j)}^\top \boldsymbol{\beta}_{(j)})^{-1}$ will cause the draws for $\boldsymbol{\lambda}_{(j)}$ to diverge, as in the frequentist case. Nevertheless, the posterior uncertainty about factor loadings and risk premia will cause $\boldsymbol{\beta}_{(j)}^\top \boldsymbol{a}_{(j)}$ to switch sign across draws, causing the posterior distribution of $\boldsymbol{\lambda}$ to put substantial probability mass on both values above and below zero. Hence, centered posterior credible intervals will tend to include zero with high probability.

In addition to the price of risk $\boldsymbol{\lambda}$, we are also interested in estimating the cross-sectional fit of the model, that is, the cross-sectional R^2. Once we obtain the posterior draws of the parameters, we can easily obtain the posterior distribution of the cross-sectional R^2, defined as

$$R^2_{ols} = 1 - \frac{(\boldsymbol{a} - \boldsymbol{\beta\lambda})^\top (\boldsymbol{a} - \boldsymbol{\beta\lambda})}{(\boldsymbol{a} - \bar{a}\mathbf{1_N})^\top (\boldsymbol{a} - \bar{a}\mathbf{1_N})}, \tag{10}$$

where $\bar{a} = \frac{1}{N}\sum_i^N a_i$. That is, for each posterior draw of $(\boldsymbol{a},\ \boldsymbol{\beta},\ \boldsymbol{\lambda})$, we can construct the corresponding draw for the R^2 from equation (10), hence tracing out its posterior distribution. We can think of equation (10) as the population R^2, where $\boldsymbol{a}$, $\boldsymbol{\beta}$, and $\boldsymbol{\lambda}$ are unknown. After observing the data, we infer the posterior distribution of $\boldsymbol{a}$, $\boldsymbol{\beta}$, and $\boldsymbol{\lambda}$, and from these we can recover the distribution of the R^2.

However, realistically, the models are rarely true. Therefore, one might want to allow for the presence of pricing errors, $\boldsymbol{\alpha}$, in the cross-sectional regression.[6] This can be easily accommodated within our Bayesian framework since in this case the data-generating process in the second stage becomes $\boldsymbol{a} = \boldsymbol{\beta\lambda} + \boldsymbol{\alpha}$. If we further assume that pricing error α_i follows an independent and identical normal distribution $\mathcal{N}(0, \sigma^2)$, the cross-sectional likelihood function in the second step becomes[7]

$$p(data|\boldsymbol{\lambda}, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{a} - \boldsymbol{\beta\lambda})^\top (\boldsymbol{a} - \boldsymbol{\beta\lambda})\right\}. \tag{11}$$

In the cross-sectional regression the "data" are the expected risk premia, $\boldsymbol{a}$, and the factor loadings, $\boldsymbol{\beta}$, albeit these quantities are not directly observable to the researcher. Hence, in the above, we are conditioning on the knowledge of these quantities, which can be sampled from the first step Normal-inverse-Wishart posterior distribution (8)–(9). Conceptually, this

[6]As we will show in the next section, this is essential for model selection.

[7]We derive a formulation with non-spherical cross-sectional pricing errors, which leads to a GLS type estimator, in Online Appendix OA.A.1.

is not very different from the Bayesian modeling of latent variables. In the benchmark case, we assume a Jeffreys' diffuse prior[8] for $(\boldsymbol{\lambda}, \sigma^2)$: $\pi(\boldsymbol{\lambda}, \sigma^2) \propto \sigma^{-2}$. In Appendix A.1.1, we show that the posterior distribution of $(\boldsymbol{\lambda}, \sigma^2)$ is then

$$\boldsymbol{\lambda}|\sigma^2, \boldsymbol{B}, \boldsymbol{\Sigma}, data \sim \mathcal{N}\left(\underbrace{(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}}_{\widehat{\boldsymbol{\lambda}}}, \underbrace{\sigma^2(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}}_{\boldsymbol{\Sigma_\lambda}}\right), \tag{12}$$

$$\sigma^2|\boldsymbol{B}, \boldsymbol{\Sigma}, data \sim \mathcal{IG}\left(\frac{N-K-1}{2}, \frac{(\boldsymbol{a}-\boldsymbol{\beta}\hat{\boldsymbol{\lambda}})^\top(\boldsymbol{a}-\boldsymbol{\beta}\hat{\boldsymbol{\lambda}})}{2}\right), \tag{13}$$

where $\mathcal{IG}$ denotes the inverse-Gamma distribution. The conditional distribution in equation (12) makes it clear that the posterior takes into account both the uncertainty about the market price of risk stemming from the first stage uncertainty about the $\boldsymbol{\beta}$ and $\boldsymbol{a}$ (that are drawn from the Normal-inverse-Wishart posterior in equations (8)-(9)), and the random pricing errors $\boldsymbol{\alpha}$ that have the conditional posterior variance in equation (13). If test assets' expected excess returns are fully explained by $\boldsymbol{\beta}$, there are no pricing errors and $\sigma^2(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}$ converges to zero; otherwise, this layer of uncertainty always exists.

Note also that we can think of the posterior distribution of $(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}$ as a Bayesian decision-maker's belief about the dispersion of the Fama-MacBeth OLS estimates after observing the data $\{\boldsymbol{R_t}, \boldsymbol{f_t}\}_{t=1}^T$. Alternatively, when pricing errors $\boldsymbol{\alpha}$ are assumed to be zero under the null hypothesis, the posterior distribution of $\boldsymbol{\lambda}$ in equation (12) collapses to a degenerate distribution, where $\boldsymbol{\lambda}$ equals $(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}$ with probability one.

Often the cross-sectional step of the FM estimation is performed via GLS rather than least squares. In our setting, under the null of the model, this leads to $\hat{\boldsymbol{\lambda}} = (\boldsymbol{\beta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}$. Therefore, we define the following GLS estimator.

Definition 2 (Bayesian Fama-MacBeth GLS (BFM-GLS)) *The posterior distribution of $\boldsymbol{\lambda}$ conditional on $\boldsymbol{B}$, $\boldsymbol{\Sigma}$ and the data, is a Dirac distribution at $(\boldsymbol{\beta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}$. A draw ($\boldsymbol{\lambda}_{(j)}$) from the posterior distribution of $\boldsymbol{\lambda}$ conditional on the data only is obtained by drawing $\boldsymbol{B}_{(j)}$ and $\boldsymbol{\Sigma}_{(j)}$ from the Normal-inverse-Wishart in equations (8)–(9) and computing $(\boldsymbol{\beta}_{(j)}^\top\boldsymbol{\Sigma}_{(j)}^{-1}\boldsymbol{\beta}_{(j)})^{-1}\boldsymbol{\beta}_{(j)}^\top\boldsymbol{\Sigma}_{(j)}^{-1}\boldsymbol{a}_{(j)}$.*

From the posterior sampling of the parameters in the definition above, we can also obtain the posterior distribution of the cross-sectional GLS R^2 defined as

$$R_{gls}^2 = 1 - \frac{(\boldsymbol{a}-\boldsymbol{\beta}\boldsymbol{\lambda_{gls}})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{a}-\boldsymbol{\beta}\boldsymbol{\lambda_{gls}})}{(\boldsymbol{a}-\bar{a}\boldsymbol{1_N})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{a}-\bar{a}\boldsymbol{1_N})}. \tag{14}$$

[8]As shown in the next subsection, in the presence of useless factors, such prior is not appropriate for model selection based on Bayes factors and posterior probabilities, since it does not lead to proper marginal likelihoods. Therefore, we introduce therein a novel prior for model selection.

Once again, we can think of equation (14) as the population GLS R^2, which is a function of the unknown quantities $\boldsymbol{a}$, $\boldsymbol{\beta}$, and $\boldsymbol{\lambda}$. But after observing the data, we infer the posterior distribution of the parameters, and from these we recover the posterior distribution of the R^2_{gls}.

Remark 1 (Generated factors) *Often factors are estimated, as, e.g., in the case of principal components (PCs) and factor mimicking portfolios (albeit the latter are not needed in our setting). This generates an additional layer of uncertainty normally ignored in empirical analysis due to the associated asymptotic complexities. Nevertheless, it is relatively easy to adjust the above defined Bayesian estimators of risk premia to account for this uncertainty. In the case of a mimicking portfolio, under a diffuse prior and Normal errors, the posterior distribution of the portfolio weights follow the standard Normal-inverse-Gamma of Gaussian linear regression models (see, e.g., Lancaster (2004)). Similarly, in the case of principal components as factors, under a diffuse prior, the covariance matrix from which the PCs are constructed follows an inverse-Wishart distribution.*[9] *Hence, the posterior distributions in Definitions 1 and 2 can account for the generated factors uncertainty by first drawing from an inverse-Wishart the covariance matrix from which PCs are constructed, or from the Normal-inverse-Gamma posterior of the mimicking portfolios coefficients, and then sampling the remaining parameters as explained in the above Definitions.*

Note that while we focus on the two-pass procedure, our method can be easily extended to the estimation of linear SDF models.

II.2 Model selection

In the previous subsection we have derived simple Bayesian estimators that deliver, in a finite sample, credible intervals robust to the presence of spurious factors, and avoid over-rejecting the null hypothesis of zero risk premia for such factors.

However, given the plethora of risk factors that have been proposed in the literature, a robust approach for models selection across non-necessarily nested models, and that can handle potentially a very large number of possible models as well as both traded and non-traded factors, is of paramount importance for empirical asset pricing. The canonical way of selecting models, and testing hypothesis, within the Bayesian framework, is through Bayes' factors and posterior probabilities, and that is the approach we present in this section. This is, for instance, the approach suggested by Barillas and Shanken (2018) for tradable factors. The key elements of novelty of the proposed method are that: i) our procedure is robust

[9] Based on these two observations, Allena (2019) proposes a generalization of Barillas and Shanken (2018) model comparison approach for these type of factors.

to the presence of spurious and weak factors, ii) it is directly applicable to both traded and non-traded factors, and iii) it selects models based on their cross-sectional performance (rather than the time series one), that is on the basis of the risk premia that the factors command.

In this subsection, we show first that flat priors for risk premia are not suitable for model selection in the presence of spurious factors. Given the close analogy between frequentist testing and Bayesian inference with flat priors, this is not too surprising. But the novel insight is that the problem arises exactly because of the use of flat priors and can therefore be fixed by using non-flat, yet non-informative, priors. Second, we introduce "spike-and-slab" priors that are robust to the presence of spurious factors, and particularly powerful in high-dimensional model selection, that is, when one wants, as in our empirical application, to test all factors in the zoo.

II.2.1 Pitfalls of Flat Priors for Risk Premia

We start this section by discussing why flat priors for risk premia are not desirable in model selection. Since we want to focus on, and select models based on the cross-sectional asset pricing properties of the factors, for simplicity we retain Jeffreys' priors for the time series parameter $(\boldsymbol{a}, \boldsymbol{\beta_f}, \boldsymbol{\Sigma})$ of the first-step regression.

In order to perform model selection, we relax the (null) hypothesis that models are correctly specified and allow instead for the presence of cross-sectional pricing errors. That is, we consider the cross-sectional regression $\boldsymbol{a} = \boldsymbol{\beta}\boldsymbol{\lambda} + \boldsymbol{\alpha}$. For illustrative purposes, we focus on spherical errors, but all the results in this and the following subsections can be generalized to the non-spherical error setting.[10]

Similar to many Bayesian variable selection problems, we introduce a vector of binary latent variables $\boldsymbol{\gamma}^\top = (\gamma_0, \gamma_1, \ldots, \gamma_K)$, where $\gamma_j \in \{0, 1\}$. When $\gamma_j = 1$, it indicates that the factor j (with associated loadings $\boldsymbol{\beta_j}$) should be included into the model, and vice versa. The number of included factors is simply given by $p_\gamma := \sum_{j=0}^{K} \gamma_j$. Note that we do not shrink the intercept, so γ_0 is always equal to 1 (as the common intercept plays the role of the first "factor"). The notation $\boldsymbol{\beta_\gamma} = [\boldsymbol{\beta}_j]_{\gamma_j=1}$ represents a p_γ-columns sub-matrix of $\boldsymbol{\beta}$.

When testing whether the risk premium of factor j is zero, the null hypothesis is $H_0 : \lambda_j = 0$. In our notation, this null hypothesis can be expressed as $H_0 : \gamma_j = 0$, while the alternative is $H_1 : \gamma_j = 1$. This is a small, but important, difference relative to the canonical frequentist testing approach: for useless factors, the risk premium is not identified, hence, testing whether it is equal to any given value is per se problematic. Nevertheless, as we show in the next section, with appropriate priors, whether a factor should be included or not is a

[10]See Online Appendix OA.A.1.

well-defined question even in the presence of useless factors.

In the Bayesian framework, the prior distribution of parameters under the alternative hypothesis should be carefully specified. Generally speaking, the priors for nuisance parameters, such as $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\Sigma}$, do not greatly influence the cross-sectional inference. But, as we are about to show, this is not the case for the priors about risk premia.

Recall that when considering multiple models, say, without loss of generality model $\boldsymbol{\gamma}$ and model $\boldsymbol{\gamma}'$, by Bayes' theorem we have that the posterior probability of model $\boldsymbol{\gamma}$ is

$$\Pr(\boldsymbol{\gamma}|data) = \frac{p(data|\boldsymbol{\gamma})}{p(data|\boldsymbol{\gamma}) + p(data|\boldsymbol{\gamma}')},$$

where we have given equal prior probability to each model and $p(data|\boldsymbol{\gamma})$ denotes the marginal likelihood of the model indexed by $\boldsymbol{\gamma}$. In Appendix A.1.2 we show that, when using a flat prior for $\boldsymbol{\lambda}$), the marginal likelihood is

$$p(data|\boldsymbol{\gamma}) \propto (2\pi)^{\frac{p_\gamma+1}{2}} |\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}}|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{N-p_\gamma+1}{2}\right)}{\left(\frac{N\hat{\sigma}_\gamma^2}{2}\right)^{\frac{N-p_\gamma+1}{2}}}\ , \tag{15}$$

where $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}} = (\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{a}$, $\hat{\sigma}_\gamma^2 = \frac{(\boldsymbol{a}-\boldsymbol{\beta}_{\boldsymbol{\gamma}}\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}})^\top(\boldsymbol{a}-\boldsymbol{\beta}_{\boldsymbol{\gamma}}\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}})}{N}$, and Γ denotes the Gamma function.

Therefore, if model $\boldsymbol{\gamma}$ includes a useless factor (whose $\boldsymbol{\beta}$ asymptotically converges to zero), the matrix $\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is nearly singular and its determinant goes to zero, sending the marginal likelihood in (15) to infinity. As a result, the posterior probability of the model containing the spurious factor goes to one.[11] Consequently, under a flat prior for risk premia, the model containing a useless factor will always be selected asymptotically. However, the posterior distribution of $\boldsymbol{\lambda}$ for the spurious factor is robust, and particularly disperse, in any finite sample.

Moreover, it is highly likely that conclusions based on the posterior coverage of $\boldsymbol{\lambda}$ contradict those arising from Bayes' factors. When the prior distribution of λ_j is too diffuse under the alternative hypothesis H_1, the Bayes' factor tends to favor H_0 over H_1, even though the estimate of λ_j is far from 0. The reason is that even though H_0 seems quite unlikely based on posterior coverages, the data is even more unlikely under H_1 than under H_0. Therefore, a disperse prior for λ_j may push the posterior probabilities to favor H_0 and make it fail to identify true factors. This phenomenon is the so-called "Bartlett Paradox" (see Bartlett (1957)).

Note also that flat, hence, improper, priors for the risk premia are not legitimate since

[11]Note that a similar problem also arises when performing time series regressions with a mimicking portfolio for weak factors.

they render the posterior model probabilities arbitrary. Suppose that we are testing the null $H_0 : \lambda_j = 0$. Under the null hypothesis, the prior for (λ, σ^2) is $\lambda_j = 0$ and $\pi(\lambda_{-j}, \sigma^2) \propto \frac{1}{\sigma^2}$. However, the prior under the alternative hypothesis is $\pi(\lambda_j, \lambda_{-j}, \sigma^2) \propto \frac{1}{\sigma^2}$. Since the marginal likelihoods of data, $p(data|H_0)$ and $p(data|H_1)$, are both undetermined, we cannot define the Bayes' factor $\frac{p(data|H_1)}{p(data|H_0)}$ (see, e.g., Chib, Zeng, and Zhao (2020)). In contrast, for nuisance parameters such as σ^2, we can continue to assign improper priors. Since both hypotheses H_0 and H_1 include σ^2, the prior for it will be offset in the Bayes' factor and in the posterior probabilities. Therefore, we can only assign improper priors for common parameters.[12] Similarly, we can still assign improper priors for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in the first time series step.

The final reason why it might be undesirable to use a flat prior in the second step is that it does not impose any shrinkage on the parameters. This is problematic given the large number of members of the factor zoo, while we have only limited time series observations of both factors and test asset returns.

In the next subsection, we propose an appropriate prior for risk premia that is both robust to spurious factors and can be used for model selection even when dealing with a very large number of potential models.

II.2.2 Spike-and-slab prior for risk premia

In order to make sure that the integration of the marginal likelihood is well-behaved, we propose a novel prior specification for the factors' risk premia $\boldsymbol{\lambda}_{\boldsymbol{f}}^{\top} = (\lambda_1, ..., \lambda_K)$.[13] Since the inference in time series regression is always valid, we only modify the priors of the cross-sectional regression parameters.

The prior that we propose belongs to the so-called *spike-and-slab* family. For exemplifying purposes, in this section we introduce a Dirac spike, so that we can easily illustrate its implications for model selection. In the next subsection we generalize the approach to a "continuous spike" prior, and study its finite sample performance in our simulation setup.

In particular, we model the uncertainty underlying the model selection problem with a mixture prior, $\pi(\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma}) \propto \pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2)\pi(\boldsymbol{\gamma})$, for the risk premium of the j-th factor. When $\gamma_j = 1$, and, hence, the factor should be included in the model, the prior follows a normal distribution given by $\lambda_j|\sigma^2, \gamma_j = 1 \sim \mathcal{N}(0, \sigma^2\psi_j)$, where ψ_j is a quantity that we will be defining below. When instead $\gamma_j = 0$, and the corresponding risk factor should not be included in the model, the prior is a Dirac distribution at zero. For the cross-sectional variance of the pricing errors we keep the same prior that would arise with Jeffreys' approach[14]: $\pi(\sigma^2) \propto \sigma^{-2}$.

[12] See Kass and Raftery (1995) (and also Cremers (2002)) for a more detailed discussion.

[13] We do not shrink the intercept λ_c.

[14] Note that since the parameter σ is common across models and has the same support in each model,

Let $\boldsymbol{D}$ denote a diagonal matrix with elements $c, \psi_1^{-1}, \cdots \psi_K^{-1}$, and $\boldsymbol{D}_{\boldsymbol{\gamma}}$ the sub-matrix of $\boldsymbol{D}$ corresponding to model $\boldsymbol{\gamma}$. We can then express the prior for the risk factors, $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$, of model $\boldsymbol{\gamma}$ as

$$\boldsymbol{\lambda}_{\boldsymbol{\gamma}} | \sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{D}_{\boldsymbol{\gamma}}^{-1}).$$

Note that c is a small positive number, since we do not shrink the common intercept, λ_c, of the cross-sectional regression.

Given the above prior specification, we sample the posterior distribution by sequentially drawing from the conditional distributions of the parameters (i.e., we use a Gibbs sampling algorithm). The crucial sampling steps of the cross-sectional parameters are given by the following proposition.

Proposition 2 (Posterior of Risk Premia with Dirac Spike-and-Slab) *The posterior distribution of* $(\lambda_\gamma, \sigma^2, \boldsymbol{\gamma})$ *under the assumption of Dirac spike-and-slab prior, conditional on the draws of the time series parameters from equations (8)–(9), is characterized via the following conditional distributions:*

$$\boldsymbol{\lambda}_{\boldsymbol{\gamma}} | data, \sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}_\gamma, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma)\right), \tag{16}$$

$$\sigma^2 | data, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N}{2}, \frac{SSR_\gamma}{2}\right), \tag{17}$$

$$p(\boldsymbol{\gamma} \mid data) = \propto \frac{|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}}{|\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}} \frac{1}{(SSR_\gamma/2)^{\frac{N}{2}}}, \tag{18}$$

where $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}} = (\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{a}$, $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma) = \sigma^2 (\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})^{-1}$, *and* $SSR_\gamma = \boldsymbol{a}^\top \boldsymbol{a} - \boldsymbol{a}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}} (\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{a} = \min_{\boldsymbol{\lambda}_{\boldsymbol{\gamma}}} \{(\boldsymbol{a} - \boldsymbol{\beta}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}_{\boldsymbol{\gamma}})^\top (\boldsymbol{a} - \boldsymbol{\beta}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}_{\boldsymbol{\gamma}}) + \boldsymbol{\lambda}_{\boldsymbol{\gamma}}^\top \boldsymbol{D}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}_{\boldsymbol{\gamma}}\}$ *and* $\mathcal{IG}$ *denotes the inverse-Gamma distribution.*

The result above is proved in Appendix A.1.3.

Note that SSR_γ is the minimized sum of squared errors with generalized ridge regression penalty term $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}^\top \boldsymbol{D}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}_{\boldsymbol{\gamma}}$. That is, our prior modeling is analogous to introducing a Tikhonov-Phillips regularization (see Tikhonov, Goncharsky, Stepanov, and Yagola (1995) and Phillips (1962)) in the cross-sectional regression step, and has the same rationale: delivering a well defined marginal likelihood in the presence of rank deficiency (which, in our settings, arises in the presence of useless factors). However, in our setting the shrinkage applied to the factors is heterogeneous, since we rely on the partial correlation between factors and test assets to set ψ_j as

$$\psi_j = \psi \times \boldsymbol{\rho}_{\boldsymbol{j}}^\top \boldsymbol{\rho}_{\boldsymbol{j}}, \tag{19}$$

the marginal likelihoods obtained under this improper prior are valid and comparable (see Proposition 1 of Chib, Zeng, and Zhao (2020)).

where $\boldsymbol{\rho_j}$ is an $N \times 1$ vector of correlation coefficients between factor j and the test assets, and $\psi \in \mathbb{R}_+$ is a tuning parameter that controls the shrinkage over all the factors.[15] When the correlation between f_{jt} and $\boldsymbol{R_t}$ is very low, as in the case of a useless factor, the penalty for λ_j, which is the reciprocal of $\psi \boldsymbol{\rho_j}^\top \boldsymbol{\rho_j}$, is very large and dominates the sum of squared errors.

Equation (16) makes clear why this Bayesian formulation is robust to spurious factors. When $\boldsymbol{\beta}$ converges to zero, $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})$ is dominated by $\boldsymbol{D}_{\boldsymbol{\gamma}}$, so the identification condition for the risk premia no longer fails. When a factor is spurious, its correlation with test assets converges to zero, hence, the penalty for this factor, ψ_j^{-1}, goes to infinity. As a result, the posterior mean of $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}} = (\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{a}$, is shrunk toward zero, and the posterior variance term $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}})$ approaches $\sigma^2 \boldsymbol{D}_{\boldsymbol{\gamma}}^{-1}$. Consequently, the posterior distribution of $\boldsymbol{\lambda}$ for a spurious factor is nearly the same as its prior. In contrast, for a normal factor that has non-zero covariance with test assets, the information contained in $\boldsymbol{\beta}$ dominates the prior information, since in this case the absolute size of $\boldsymbol{D}_{\boldsymbol{\gamma}}$ is small relative to $\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}}$.

When comparing two models, using posterior model probabilities is equivalent to simply using the ratio of the marginal likelihoods, that is, the Bayes factor, which is defined as

$$BF_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} = p(data|\boldsymbol{\gamma})/p(data|\boldsymbol{\gamma}'),$$

where we have given equal prior probability to model $\boldsymbol{\gamma}$ and model $\boldsymbol{\gamma'}$.

Corollary 1 (Variable Selection via the Bayes Factor) *Consider two nested linear factor models, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$. The only difference between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ is γ_p: γ_p equals 1 in model $\boldsymbol{\gamma}$ but 0 in model $\boldsymbol{\gamma}'$. Let $\boldsymbol{\gamma}_{-p}$ denote a $(K-1)\times 1$ vector of model index excluding γ_p: $\boldsymbol{\gamma}^\top = (\boldsymbol{\gamma}_{-p}^\top, 1)$ and $\boldsymbol{\gamma}'^\top = (\boldsymbol{\gamma}_{-p}^\top, 0)$ where, without loss of generality, we have assumed that the factor p is ordered last. The Bayes factor is then*

$$BF_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} = \left(\frac{SSR_{\gamma'}}{SSR_{\gamma}}\right)^{\frac{N}{2}} \left(1 + \psi_p \boldsymbol{\beta_p}^\top \left[\boldsymbol{I_N} - \boldsymbol{\beta}_{\boldsymbol{\gamma}'} (\boldsymbol{\beta}_{\boldsymbol{\gamma}'}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'})^{-1} \boldsymbol{\beta}_{\boldsymbol{\gamma}'}^\top\right] \boldsymbol{\beta_p}\right)^{-\frac{1}{2}}. \tag{20}$$

The result above is proved in Appendix A.1.4.

Since $\boldsymbol{\beta_p}^\top [\boldsymbol{I_N} - \boldsymbol{\beta}_{\boldsymbol{\gamma}'} (\boldsymbol{\beta}_{\boldsymbol{\gamma}'}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'})^{-1} \boldsymbol{\beta}_{\boldsymbol{\gamma}'}^\top] \boldsymbol{\beta_p}$ is always positive, ψ_p plays an important role in variable selection. For a strong and useful factor that can substantially reduce pricing errors, the first term in equation (20) dominates, and the Bayes factor will be much greater than 1, hence, providing evidence in favor of model $\boldsymbol{\gamma}$.

Remember that $SSR_\gamma = \min_{\boldsymbol{\lambda}_{\boldsymbol{\gamma}}} \{(\boldsymbol{a} - \boldsymbol{\beta}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}_{\boldsymbol{\gamma}})^\top (\boldsymbol{a} - \boldsymbol{\beta}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}_{\boldsymbol{\gamma}}) + \boldsymbol{\lambda}_{\boldsymbol{\gamma}}^\top \boldsymbol{D}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}_{\boldsymbol{\gamma}}\}$, hence, we always have $SSR_\gamma \leq SSR_{\gamma'}$ in sample. There are two effects of increasing ψ_p: i) when ψ_p is large,

[15] Alternatively, we could have set $\psi_j = \psi \times \boldsymbol{\beta_j}^\top \boldsymbol{\beta_j}$, where $\boldsymbol{\beta_j}$ is an $N \times 1$ vector. However, $\boldsymbol{\rho_j}$ has the advantage of being invariant to the units in which the factors are measured.

the penalty for λ_p is small, hence, it is easier to minimize SSR_γ, and $SSR_{\gamma'}/SSR_\gamma$ becomes much larger than 1; ii) large ψ_p decreases the second term in equation (20), lowering the Bayes' factor, and acting as a penalty for dimensionality.

A particularly interesting case is when the factor is useless: $\boldsymbol{\beta_p}$ converges to zero, but the penalty term $1/\psi_p \propto 1/\boldsymbol{\rho_p}^\top \boldsymbol{\rho_p}$ goes to infinity. On the one hand, the first term in equation (20) will converge to 1; on the other hand, since $\psi_p \approx 0$ in large sample, the second term in equation (20) will also be around 1. Therefore, the Bayes factor for a useless factor will go to 1 asymptotically.[16] In contrast, a useful factor should be able to greatly reduce the sum of squared errors SSR_γ, so the Bayes' factor will be dominated by SSR_γ, yielding a value substantially above 1.

Note that since our prior restores the validity of the marginal likelihood, *any* hypothesis on the parameters (e.g., whether the pricing errors are jointly zero) can be tested via posterior probabilities or, equivalently, Bayesian p-values. In particular, we obtain closed-form solutions for testing hypothesis about risk premia by centering the Dirac spike at the null value rather than at zero.

Corollary 2 (Testing Risk Premia) *Let* $\boldsymbol{\lambda}_{-\boldsymbol{\gamma}} = \tilde{\boldsymbol{\lambda}}_{-\boldsymbol{\gamma}}$ *and* $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}|\sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{D}_{\boldsymbol{\gamma}}^{-1})$ *in model* $\boldsymbol{\gamma}$*, Proposition 2 still applies with* $SSR_{\boldsymbol{\gamma}}$ *replaced by:*

$$\begin{aligned}\widetilde{SSR}_{\boldsymbol{\gamma}} &= (\boldsymbol{a} - \boldsymbol{\beta}_{-\boldsymbol{\gamma}}\tilde{\boldsymbol{\lambda}}_{-\boldsymbol{\gamma}})^\top(\boldsymbol{a} - \boldsymbol{\beta}_{-\boldsymbol{\gamma}}\tilde{\boldsymbol{\lambda}}_{-\boldsymbol{\gamma}}) - (\boldsymbol{a} - \boldsymbol{\beta}_{-\boldsymbol{\gamma}}\tilde{\boldsymbol{\lambda}}_{-\boldsymbol{\gamma}})^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top(\boldsymbol{a} - \boldsymbol{\beta}_{-\boldsymbol{\gamma}}\tilde{\boldsymbol{\lambda}}_{-\boldsymbol{\gamma}})\\ &= \min_{\boldsymbol{\lambda}_{\boldsymbol{\gamma}}}\{(\tilde{\boldsymbol{a}} - \boldsymbol{\beta}_{\boldsymbol{\gamma}}\boldsymbol{\lambda}_{\boldsymbol{\gamma}})^\top(\tilde{\boldsymbol{a}} - \boldsymbol{\beta}_{\boldsymbol{\gamma}}\boldsymbol{\lambda}_{\boldsymbol{\gamma}}) + \boldsymbol{\lambda}_{\boldsymbol{\gamma}}^\top \boldsymbol{D}_{\boldsymbol{\gamma}}\boldsymbol{\lambda}_{\boldsymbol{\gamma}},\end{aligned}$$

where $\tilde{\boldsymbol{a}} \equiv \boldsymbol{a} - \boldsymbol{\beta}_{-\boldsymbol{\gamma}}\tilde{\boldsymbol{\lambda}}_{-\boldsymbol{\gamma}}$ *denotes the vector of cross-sectional residual expected returns that are unexplained by factors* $f_{-\boldsymbol{\gamma}}$ *with risk premia* $\tilde{\boldsymbol{\lambda}}_{-\boldsymbol{\gamma}}$*. A Bayesian p-value for the null hypothesis can then be constructed by integrating* $1 - p(\boldsymbol{\gamma} \mid data)$ *in equation (18) with respect to the Normal-inverse-Wishart in equations (8)–(9).*

The Corollary is established following the same steps as in the proof of Proposition 2 in Appendix A.1.3.

The result above can be used for joint hypothesis testing within the Bayesian framework (e.g., building confidence intervals), and it is very similar in spirit to the standard frequentist identification-robust inference.

Remark 3 (Level Factors) *Identification failure of factors' risk premia can arise in the presence of "level factors," exposure to which is non-zero, but lacks cross-sectional spread i.e.* $\boldsymbol{\beta}_j \to c\mathbf{1}_{\boldsymbol{N}}$ *with* $c \neq 0$*. These factors help explain the average level of returns, but not the*

[16] But in finite sample it may deviate from its asymptotic value, so we should not use 1 as a threshold when testing the null hypothesis $H_0 : \gamma_p = 0$.

their cross-sectional dispersion, and, hence, are collinear with the common cross-sectional intercept. Our approach can handle this case by using variance standardized variables in the estimation and replacing the penalty in (19) with

$$\psi_j = \psi \times \widetilde{\boldsymbol{\rho}}_j^\top \widetilde{\boldsymbol{\rho}}_j, \tag{21}$$

where $\widetilde{\boldsymbol{\rho}}_j$ *is the cross-sectionally demeaned vector of correlations with asset returns, for example,* $\widetilde{\boldsymbol{\rho}}_j = \boldsymbol{\rho}_j - \left(\frac{1}{N}\sum_{i=1}^{N} \rho_{j,i}\right) \times \mathbf{1}_{\boldsymbol{N}}$.

II.2.3 Continuous Spike

We extend the Dirac spike-and-slab prior by encoding a continuous spike for λ_j, when γ_j equals 0. Following the literature on Bayesian variable selection (see, e.g., George and McCulloch (1993, 1997) and Ishwaran, Rao, et al. (2005)), we model the uncertainty underlying model selection with a mixture prior $\pi(\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega}) = \pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2)\pi(\boldsymbol{\gamma}|\boldsymbol{\omega})\pi(\boldsymbol{\omega})$, which is specified as follows:

$$\lambda_j|\gamma_j, \sigma^2 \sim \mathcal{N}\left(0, r(\gamma_j)\psi_j\sigma^2\right). \tag{22}$$

Note the introduction of a new element, $r(\gamma_j)$, in the prior, and where $r(1) = 1$ and $r(0) = r \ll 1$. As we explain below, the additional parameter vector $\boldsymbol{\omega}$ encodes our prior beliefs about the sparsity of the true model.

Redefine $\boldsymbol{D}$ as a diagonal matrix with elements c, $(r(\gamma_1)\psi_1)^{-1}, \ldots, (r(\gamma_K)\psi_K)^{-1}$, where ψ_j is given as before by equation (19). In matrix notation, the prior for $\boldsymbol{\lambda}$ is: $\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2\boldsymbol{D}^{-1})$. The term $r(\gamma_j)\psi_j$ in $\boldsymbol{D}^{-1}$ is set to be small or large, depending on whether $\gamma_j = 0$ or $\gamma_j = 1$. In the empirical implementation, we set r to a value much smaller than 1 since we intend to shrink λ_j toward zero when γ_j is 0.[17] Hence the spike component concentrates the mass of $\boldsymbol{\lambda}$ toward zero, whereas the slab component allows $\boldsymbol{\lambda}$ to take values over a much wider range. Therefore, the posterior distribution of $\boldsymbol{\lambda}$ is very similar to the case of a Dirac spike in section II.2.2.

A desirable feature of the prior in equation (22) is that it encodes beliefs about quantities that are salient to, and observed by, practitioners and researchers: Sharpe ratios, excess returns, and their volatilities. For instance, in the empirical applications below we will consider a value of ψ in the 10–20 range as a reasonable benchmark since it is equivalent to a prior standard deviation for the (annualized) risk premia of about 9.1%–12.7% for the typical strong factor, implying that (annualized) factor Sharpe ratios as large as 0.92–2.62 are within the centered 95% prior coverage.

[17]We set $r = 0.0001$. In our framework, r is essentially a tuning parameter, hence, we need to choose a reasonable value such that we can identify useful factor but exclude spurious ones.

Furthermore, the prior encodes our expectation about the contribution of the factors to the squared Sharpe ratio of the test assets relative to the contribution coming from the pricing errors. To see this, consider the case in which (as in our empirical applications) both factors and returns are standardized.[18] It then follows that

$$\frac{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \boldsymbol{\gamma}, \sigma^2]}{\mathbb{E}_\pi[SR^2_{\boldsymbol{\alpha}} \mid \sigma^2]} = \frac{\sum_{k=1}^K r(\gamma_k)\psi_k}{N} = \frac{\psi \sum_{k=1}^K r(\gamma_k)\tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k}{N}.$$

where $SR_{\boldsymbol{f}}$ and $SR_{\boldsymbol{\alpha}}$ denote, respectively, the Sharpe ratios of all factors ($\boldsymbol{f_t}$) and of the pricing errors of all assets ($\boldsymbol{\alpha}$), and $\mathbb{E}_\pi$ denotes prior expectations. In the baseline sample of our empirical applications, $\sum_{k=1}^K \tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k / N \simeq 0.51$. Hence, for ψ in the 10–20 range, if, say, 50% of the factors are selected, our prior expectation is that the factors should explain about 71%–83% of the squared Sharpe ratio of test assets.

Even though the prior on model index $\boldsymbol{\gamma}$ could be simply set to be $\pi(\boldsymbol{\gamma}) = 1/2^K$, we decide instead to encode our a priori belief about the sparsity of the true model using the prior distribution $\pi(\gamma_j = 1|\omega_j) = \omega_j$. As in the literature on predictors selection, we assign the following prior distribution to $(\boldsymbol{\gamma}, \boldsymbol{\omega})$:

$$\pi(\gamma_j = 1|\omega_j) = \omega_j, \quad \omega_j \sim Beta\,(a_\omega, b_\omega)\,.$$

Different hyper-parameters a_ω and b_ω determine whether we a priori favor more parsimonious models or not, since the prior expected probability of selecting a factor is simply $\frac{a_\omega}{a_\omega + b_\omega}$.[19] Furthermore, a_ω and b_ω can be chosen to encode prior beliefs about the Sharpe ratio achievable in the economy since $\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2] = \frac{a_\omega}{a_\omega + b_\omega}\psi\sigma^2 \sum_{k=1}^K \tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k$ as $r \to 0$.

The considerations above imply that an agent's expectations about the Sharpe ratio achievable *i*) with only *one* factor, *ii*) with *all* the factors jointly, as well as *iii*) the sparsity of the "true" model, uniquely determine the parameters ψ, a_ω, b_ω.

When ω_j is constant and equal to 0.5 and r converges to 0, the continuous spike-and-slab prior is equivalent to the one with a Dirac spike in Section II.2.2. However, treating instead ω_j, and consequently γ_j, as a parameter to be drawn is particularly useful in the high dimensional case. Imagine that there are 30 candidate factors in the factor zoo. In the Dirac spike-and-slab prior case we have to calculate the posterior model probabilities for 2^{30} different models. Given that we update $(\boldsymbol{a}, \boldsymbol{\beta})$ in each sampling round, posterior probabilities for all models are necessarily re-computed for every new draw of these quantities, rendering the computational cost very large. In contrast, with the above approach we can simply use

[18]So that λ_k is the Sharpe ratio of factor k and α_n is the Sharpe ratio of the pricing error of asset n.

[19]We set $a_\omega = b_\omega = 2$ in the benchmark case, that is, each factor has an ex ante expected probability of being selected equal to 50%. However, we could assign $a_\omega = 1$ and $b_\omega = 2$ in order to select a sparser model.

the posterior mean of γ_j to approximate the posterior marginal probability of the j-th factor.

Similarly to the Dirac spike-and-slab case, we use Gibbs sampling to draw the posterior distribution of the parameters $(\boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2)$ and, most importantly, $\boldsymbol{\gamma}$.

Proposition 4 (Posterior of Risk Premia with Continuous Spike-and-Slab) *The posterior distribution of $(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \sigma^2)$ under the assumption of continuous spike-and-slab prior, conditional on the draws of the time series parameters from equations (8)–(9), is characterized via the following conditional distributions:*

$$\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})\right), \tag{23}$$

$$\frac{p(\gamma_j = 1|data, \boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2, \boldsymbol{\gamma_{-j}})}{p(\gamma_j = 0|data, \boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2, \boldsymbol{\gamma_{-j}})} = \frac{\omega_j}{1-\omega_j}\frac{p(\lambda_j|\gamma_j = 1, \sigma^2)}{p(\lambda_j|\gamma_j = 0, \sigma^2)}, \tag{24}$$

$$\omega_j|data, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^2 \sim Beta\left(\gamma_j + a_\omega, 1 - \gamma_j + b_\omega\right), \text{ and} \tag{25}$$

$$\sigma^2|data, \boldsymbol{\omega}, \boldsymbol{\lambda}, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N+K+1}{2}, \frac{(\boldsymbol{a}-\boldsymbol{\beta}\boldsymbol{\lambda})^\top(\boldsymbol{a}-\boldsymbol{\beta}\boldsymbol{\lambda}) + \boldsymbol{\lambda}^\top \boldsymbol{D}\boldsymbol{\lambda}}{2}\right), \tag{26}$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{\beta}^\top\boldsymbol{\beta} + \boldsymbol{D})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}) = \sigma^2(\boldsymbol{\beta}^\top\boldsymbol{\beta} + \boldsymbol{D})^{-1}$.

The result above is proved in Appendix A.1.5.

III Simulation

We build a simple setting for a linear factor model that includes both strong and irrelevant factors and allows for potential model misspecification.

The cross-section of asset returns mimics the empirical properties of 25 Fama-French portfolios sorted by size and value. We generate both factors and test asset returns from normal distributions, assuming that HML is the only useful factor. A misspecified model also includes pricing errors from the two-step FM procedure, which makes the vector of simulated expected returns equal to their sample mean estimates of 25 Fama-French portfolios. Finally, a spurious factor is simulated from an independent normal distribution with mean zero and standard deviation 1%. In summary,

$$f_{t,useless} \overset{\text{iid}}{\sim} \mathcal{N}(0, (1\%)^2), \quad \tilde{f}_{t,HML} \overset{\text{iid}}{\sim} \mathcal{N}(\bar{r}_{HML}, \hat{\sigma}^2_{HML}), \quad \bar{f}_{t,HML} = \tilde{f}_{t,HML} - \bar{\tilde{f}}_{t,HML}$$
$$\boldsymbol{R_t}|\bar{f}_{t,HML} \overset{\text{iid}}{\sim} \begin{cases} \mathcal{N}\left(\hat{\lambda}_c \boldsymbol{1_N} + \hat{\boldsymbol{\beta}}\left(\hat{\lambda}_{HML} + \bar{f}_{t,HML}\right),\ \hat{\boldsymbol{\Sigma}}\right), & \text{if the model is correct, and} \\ \mathcal{N}\left(\bar{\boldsymbol{R}} + \hat{\boldsymbol{\beta}}\bar{f}_{t,HML},\ \hat{\boldsymbol{\Sigma}}\right), & \text{if the model is misspecified,} \end{cases}$$

where factor loadings, risk premia, and variance-covariance matrix of returns are equal to their OLS sample estimates from the time series and cross-sectional regressions of the two-pass FM procedure, applied to 25 size-and-value portfolios and HML as a factor. All the model parameters are estimated on monthly data from July 1963 to December 2017.

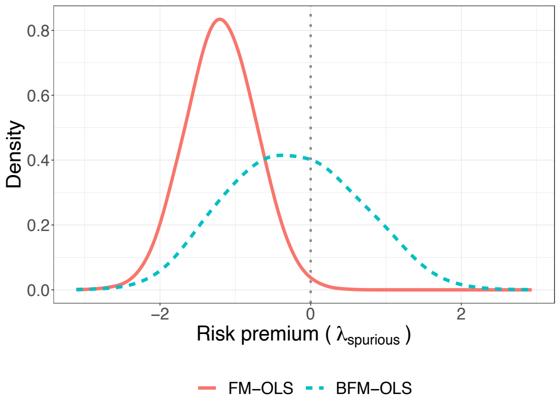
To illustrate the properties of the frequentist and Bayesian approaches, we consider 3 estimation setups: (a) the model includes only a strong factor (HML); (b) the model includes only a useless factor; and (c) the model includes both strong and useless factors. Each setting can be correctly or incorrectly specified, with the following sample sizes: T = 100, 200, 600, and 1,000. We compare the performance of the OLS/GLS standard frequentist and Bayesian Fama-MacBeth estimators (FM and BFM, correspondingly) with the focus on risk premia recovery, testing, and identification of strong and useless factors for model comparison.

III.1 Estimating risk premia via Bayesian Fama-MacBeth

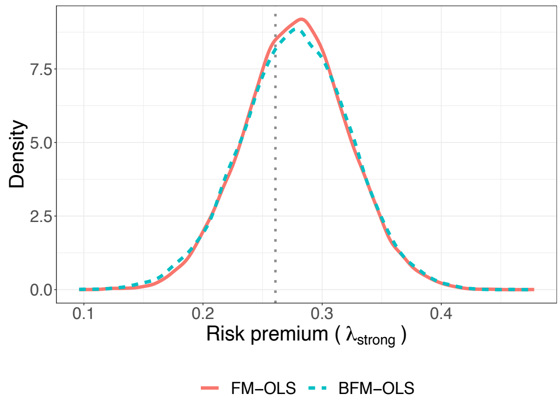
Since it is unlikely that in most empirical settings a linear factor model is correctly specified, we focus our discussion on the case that allows for model misspecification. Furthermore, we focus on the most realistic (and challenging) model setup, which includes both useless and strong factors.

Table 1 compares the performance of frequentist and Bayesian Fama-MacBeth estimators and reports the size of the tests for risk premia and confidence intervals for cross-sectional R^2. Since the model is misspecified, cross-sectional R^2 never reaches 100% (with the population value of 31% (82%) for OLS (GLS)). In the case of the standard FM approach, tests are constructed using standard t-statistics, adjusted for Shanken correction, and in the case of the BFM and BFM-GLS we rely on the quantiles of the posterior distribution to form the credible confidence intervals for parameters. The last two columns also report the quantiles of the mode of the posterior distribution of R^2 across the simulations. As expected, in the conventional case of frequentist Fama-MacBeth estimation, the useless factor is often found to be a significant predictor of the asset returns: its OLS (GLS) t-statistic would be above a 5%-critical value in more than 60% (80%) of the simulations. On the contrary, the Bayesian confidence intervals have approximately the right coverage and reject the null of no risk premia attached to the spurious factor with frequency asymptotically approaching the size of the tests.

The crowding out of the true factors by the useless ones could also be an important empirical concern. When the model is misspecified, the presence of spurious factors can also bias the risk premia estimates for the strong ones, and often leads to their *crowding out* of the model. Panel A in Table 1 serves as a good illustration of this possibility, with risk premia estimates for the strong factor clearly biased in the frequentist estimation by the identification failure in case of the frequentist approach. Again, in this case BFM provides reliable, albeit conservative, confidence bounds for model parameters.



(a) Model with a spurious factor



(b) Model with a strong factor

Figure 1: Distribution of risk premia estimates.
Posterior distribution of risk premia (blue dashed line) from BFM-OLS estimation of a misspecified one-factor model based on a single simulation with $T = 1000$, and asymptotic distribution of the frequentist FM estimate (red solid line). The dotted line corresponds to the pseudo-true value of the parameter (defined to be 0 for a useless factor). Panel (a): case of a single spurious factor included into the model. Panel (b): relies on a strong, well-identified factor, included in a (nevertheless) misspecified model.

In the Online Appendix OA.B.1 we report additional results for a wide range of alternative simulation settings, also considering correctly specified models and cross-sections of different dimensions. In all cases BFM and BFM-GLS perform very well in both detecting the spurious factor and retaining the strong factor, hence, confirming the soundness of the proposed method.

Why does the Bayesian approach work when the frequentist fails? The argument is probably best summarized by Figure 1a, which plots a posterior distribution of $\hat{\lambda}_{useless}$ for BFM from one of the simulations, along with the pseudo-true value of the risk premium, defined as 0 in this case. In this particular simulation, Fama-MacBeth OLS estimate of $\lambda_{useless}$ is -1.19%, with Shanken-corrected t-statistics equal to -2.55, so according to traditional hypothesis testing, we would reject the null of $\lambda_{useless} = 0$ even at 1%. The posterior distribution of BFM estimates of the risk premium (the blue line in Figure 1a) behaves rather differently: it is centered around 0 and overall more spread out, with a confidence interval $(-1.603\%, 1.201\%)$. Intuitively, the main driving force behind it is the fact that in BFM, $\boldsymbol{\beta}$ is updated continuously: when $\hat{\boldsymbol{\beta}}$ is close to zero, the posterior draws of $\boldsymbol{\beta}$ will be positive or negative randomly, which implies that the conditional expectation of $\boldsymbol{\lambda}$ in equation 12 will also switch sign, depending on the draw. As a result, the posterior distribution of $\lambda_{useless}$ is centered around 0, and so is the confidence interval. The same logic applies to the case of BFM-GLS. Note that the Bayesian prior does not have any significant impact on the risk premia estimation of strong factors: In the case of well-identified sources of risk (see, e.g., Figure 1b), the Bayesian and frequentist approach give virtually identical results.

Note also that restoring the validity of the marginal likelihood using the spike-and-slab

Table 1: Tests of risk premia in a misspecified model with useless and strong factors

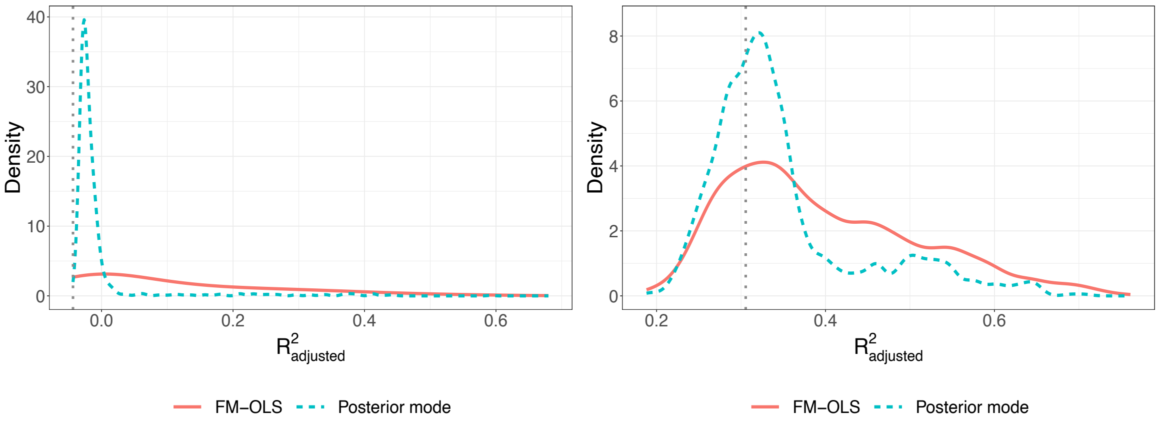
	T	λ_c 10%	λ_c 5%	λ_c 1%	λ_{strong} 10%	λ_{strong} 5%	λ_{strong} 1%	$\lambda_{useless}$ 10%	$\lambda_{useless}$ 5%	$\lambda_{useless}$ 1%	R^2_{adj} 5th	R^2_{adj} 95th
Panel A: OLS												
FM	100	0.082	0.039	0.008	0.121	0.067	0.016	0.099	0.023	0.001	-5.13%	56.63%
	200	0.096	0.044	0.005	0.157	0.100	0.034	0.129	0.039	0.005	1.27%	61.90%
	600	0.093	0.034	0.014	0.212	0.147	0.071	0.264	0.129	0.022	8.40%	61.78%
	1000	0.102	0.046	0.010	0.261	0.194	0.098	0.380	0.199	0.056	11.84%	62.48%
	20000	0.114	0.054	0.009	0.289	0.229	0.152	0.848	0.633	0.240	25.07%	60.76%
BFM	100	0.035	0.012	0.001	0.028	0.007	0.001	0.004	0.001	0.000	-2.11%	40.33%
	200	0.049	0.017	0.001	0.067	0.031	0.004	0.011	0.003	0.000	-1.75%	48.28%
	600	0.05	0.018	0.004	0.099	0.047	0.005	0.047	0.014	0.002	10.20%	55.72%
	1000	0.041	0.021	0.003	0.102	0.048	0.011	0.071	0.035	0.004	14.87%	56.95%
	20000	0.017	0.007	0.000	0.087	0.033	0.007	0.099	0.055	0.012	24.80%	54.66%
Panel B: GLS												
FM	100	0.219	0.155	0.057	0.224	0.135	0.066	0.303	0.198	0.064	19.11%	77.75%
	200	0.155	0.092	0.028	0.149	0.090	0.024	0.263	0.183	0.061	55.37%	81.71%
	600	0.121	0.068	0.015	0.116	0.064	0.016	0.391	0.293	0.134	69.48%	84.33%
	1000	0.115	0.061	0.013	0.115	0.057	0.012	0.487	0.387	0.216	73.05%	84.74%
	20000	0.084	0.050	0.009	0.100	0.041	0.005	0.864	0.836	0.757	79.79%	84.24%
BFM	100	0.122	0.069	0.016	0.129	0.070	0.017	0.046	0.017	0.002	32.43%	68.69%
	200	0.112	0.056	0.012	0.099	0.048	0.012	0.031	0.012	0.000	48.44%	73.55%
	600	0.096	0.049	0.011	0.086	0.045	0.009	0.049	0.016	0.002	65.76%	80.30%
	1000	0.081	0.036	0.007	0.073	0.032	0.006	0.058	0.030	0.003	70.64%	81.54%
	20000	0.027	0.005	0.000	0.022	0.007	0.000	0.098	0.047	0.013	79.74%	82.59%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_c and λ_{strong}, $\lambda^*_{useless} \equiv 0$ in a misspecified model with intercept, a strong, and a useless factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 1,000 simulations, evaluated at the point estimates for FM and at the posterior mode for BFM. The true value of cross-sectional R^2_{adj} is 30.55% (81.75%) for OLS (GLS) estimation.

prior of Section II.2.2 allows for valid hypothesis testing, even as $T \to \infty$, via posterior probabilities and Bayes factors as per Corollary 2. We report corresponding simulation results for the Bayesian p-value in Online Appendix OA.B.3, and show that spurious factors are easily detected, while true sources of risk are retained. Furthermore, the presence of a spurious factor leaves both power and size of tests of the strong factor virtually unaffected.

III.1.1 Evaluating cross-sectional fit

In addition to risk premia estimates, it is often useful to understand the quality of cross-sectional fit of the model. Indeed, the increase in cross-sectional R^2 is often interpreted as measuring the *economic* importance of the predictor, contrary to the statistical one implied by the risk premia significance. It is well-known, however, that the average values of R^2 are not always informative about the true model performance: its sample distribution often suffers from a large estimation uncertainty (see, e.g., Stock (1991) and Lewellen, Nagel, and Shanken (2010)), and has a non-standard distribution when the matrix of $\boldsymbol{\beta}$ has reduced rank (see Kleibergen and Zhan (2015) and Gospodinov, Kan, and Robotti (2019)). In this section we further investigate the properties of cross-sectional R^2 in the frequentist and Bayesian



(a) Model with a spurious factor **(b)** Model with a strong and useless factors

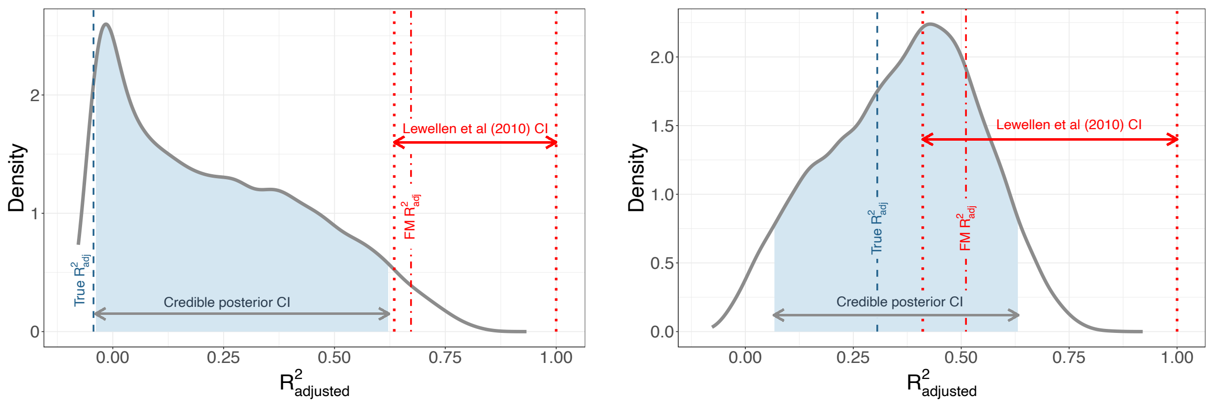
Figure 2: Cross-sectional distribution of R^2_{adj} in models with strong and useless factors.

Asymptotic distribution of cross-sectional R^2 under different model specifications across 1,000 simulations of sample size $T = 20,000$. Blue dashed lines correspond to the distribution of the posterior mode for R^2_{adj}, while red solid lines depict the pointwise sample distribution of R^2_{adj} evaluated at the frequentist FM estimates. Grey dotted line stands for the true value of R^2_{adj}.

FM regressions.

Figure 2 shows the distribution of cross-sectional OLS R^2 across a large number of simulations for the asymptotic case of $T = 20,000$ and a misspecified process for returns. Since when the model is strongly identified, the distribution of posterior BFM estimates of the risk premia coincides almost exactly with that of the standard FM procedure (see Figure 1b), cross-sectional fit would have been the same as well. The major difference emerges whenever a useless factor is included into the candidate set of variables. Indeed, it is well-known that in this case the distribution of conventional measures of fit is non-standard and often inflated (Kleibergen and Zhan (2015)). This is further confirmed in Figure 2, which shows that under the presence of spurious factors, conventional FM R^2 has an extremely spreadout right tail of the distribution, which makes it easy to find a substantial increase of fit whenever the model is simply not identified. This unfortunate property of the frequentist approach is not shared by the inference with BFM. Indeed, the mode of the posterior distribution of R^2 is generally tightly centered around the true values. The slight bump to the right tail of the distribution comes from the fact that whenever a spurious factor is included into the model with a small probability (based on t-statistic cut-off, this is equal to the size of the test; see, e.g., Table 1, Panel A), its fit will be similar to that of the frequentist estimation.

However, the pointwise distribution of cross-sectional R^2 across the simulations is only part of the story, as it does not reveal the in-sample estimation uncertainty and whether the confidence intervals are credible in reflecting it. While BFM incorporates this uncertainty directly into the shape of its posterior distribution, one needs to rely on bootstrap-like



(a) Model with a spurious factor **(b)** Model with a strong and useless factors

Figure 3: The estimation uncertainty of cross-sectional R^2.

Posterior densities of cross-sectional R^2_{adj} in one representative simulation with centered 90% confidence interval (shaded area). Blue dashed line denotes the true R^2_{adj}. Red dashed-dotted line depicts FM R^2_{adj} estimate with 90% Lewellen, Nagel, and Shanken (2010) confidence intervals (red dotted lines).

algorithms to build a similar analogue in the frequentist case. As a frequentist benchmark, we use the approach of Lewellen, Nagel, and Shanken (2010) to construct the confidence interval[20] for R^2.

Figure 3 presents the posterior distribution of cross-sectional R^2 for a model that contains a useless factor (and, potentially, a strong one, too) and contrasts it with a frequentist value and the confidence interval around it. Consider, for example, Figure 3a. The fact that the in-sample FM estimate of cross-sectional fit (51%) is substantially higher than the mode of the posterior distribution (–2%, which is close to the true value of R^2_{adj}, about −4%) is not surprising, given the previous results on the pointwise distribution of the estimates. What is quite interesting, however, is the coverage of the confidence interval constructed via the simulation-based approach of Lewellen, Nagel, and Shanken (2010). Not only does it not include true value of the cross-sectional fit, but, in fact, in this particular simulation, it suggests that R^2_{adj} should be between 42% and 100%. A similar mismatch between the seemingly high levels of cross-sectional fit produced by a frequentist approach and their true values can also be observed in Figure 3b for the case of including both strong and a useless factors.

The BFM estimator performed well in a wide range of additional simulations we have conducted. For example, in Section OA.B.2 of the Online Appendix we show that it can easily be applied even in the case of a large cross-section of test assets, that is, all of the nice properties of the estimator discussed above hold in a large-N setting as well.

[20]Details on this procedure can be found in the Online Appendix OA.A.2

III.2 The Bayes factors

How well do flat and spike-and-slab priors work empirically in selecting relevant and detecting spurious factors in the cross-section of asset returns? We revisit the theoretical results from Section II.1 using the same simulation design employed to evaluate the estimation of risk premia.

Table 2: The probability of retaining risk factors using Bayes factors

		T	55%	57%	59%	61%	63%	65%
Panel A: strong factors								
Flat Prior	f_{strong}	200	0.813	0.784	0.758	0.722	0.693	0.662
		600	0.929	0.915	0.896	0.876	0.851	0.834
		1,000	0.972	0.963	0.957	0.951	0.937	0.924
Spike-and-Slab Prior	f_{strong}	200	0.917	0.902	0.884	0.868	0.837	0.813
		600	0.998	0.998	0.998	0.996	0.993	0.991
		1,000	1.000	1.000	1.000	1.000	1.000	0.999
Panel B: useless factors								
Flat Prior	$f_{useless}$	200	1.000	0.996	0.988	0.967	0.919	0.822
		600	0.998	0.998	0.995	0.988	0.977	0.943
		1,000	1.000	1.000	1.000	0.994	0.983	0.965
Spike-and-Slab Prior	$f_{useless}$	200	0.022	0.004	0.001	0.001	0.000	0.000
		600	0.000	0.000	0.000	0.000	0.000	0.000
		1000	0.000	0.000	0.000	0.000	0.000	0.000
Panel C: strong and useless factors								
Flat Prior	f_{strong}	200	0.924	0.897	0.874	0.848	0.821	0.799
		600	0.988	0.985	0.976	0.974	0.965	0.958
		1,000	0.998	0.996	0.996	0.995	0.992	0.987
	$f_{useless}$	200	0.984	0.960	0.910	0.811	0.702	0.584
		600	0.999	0.993	0.985	0.954	0.913	0.854
		1,000	1.000	1.000	0.995	0.986	0.966	0.945
Spike-and-Slab Prior	f_{strong}	200	0.916	0.901	0.877	0.861	0.837	0.816
		600	0.998	0.998	0.998	0.996	0.994	0.991
		1,000	1.000	1.000	1.000	1.000	1.000	0.999
	$f_{useless}$	200	0.005	0.001	0.000	0.000	0.000	0.000
		600	0.000	0.000	0.000	0.000	0.000	0.000
		1,000	0.000	0.000	0.000	0.000	0.000	0.000

Frequency of retaining risk factors for different choice sets across 1,000 simulations of different size (T=200, 600, and 1,000). A factor is retained if its posterior probability, $\Pr(\gamma_i = 1|data)$, is greater than a certain threshold: 55%, 57%, 59%, 61%, 63% and 65%. In Panel A, the candidate risk factor is truly cross-sectionally priced and strongly identified, while in Panel B it is not. Panel C reports the case of both strong and useless candidate factors in the model. In the estimation with spike-and-slab prior, we standardize both returns and factors. In addition, we choose hyper-parameters $\psi = 20$ and $r = 0.0001$, and the prior variance of risk premium λ_f is proportional to its demeaned-correlation with test asset excess returns as in remark 3.

Consider a cross-section of 25 portfolios that is actually loading on two systematic sources of risk, with the econometrician potentially observing at most only one of them,[21] a strong (and priced) f_t. However, there is also a second candidate factor available, which is orthogonal to asset returns and essentially useless. We compute Bayes' factors, corresponding to

[21] That is, we focus on the empirically relevant case of the model being always misspecified. We report very similar results results for the case of correct specification in Section OA.B.4 of the Online Appendix.

each of the potential sources of risk, and document the empirical probability of retaining the variable in the model across 1,000 simulations. Again, we consider models that contain either strong or useless factors, or a combination of both, and different sample sizes ($T = 200$, 600, and 1,000). In each case we run the Gibbs sampling algorithm derived using the continuous spike-and-slab prior and then approximate the marginal probability of each factor by the posterior mean of γ_j. The decision rule is based on a range of critical values, 55%–65%, such that whenever the posterior mean of γ_j is above a particular threshold, we retain the factor. Finally, we also compute the probability of retaining a factor under a flat prior, which would be the standard in the literature.

Table 2 summarizes our findings. When only a true risk factor is included in the candidate set (Panel A), both flat and spike-and-slab priors successfully identify it with a high probability, especially in large sample. But the spike-and-slab prior is characterized by higher power in retaining the strong factor for all specifications and sample sizes considered.

The difference between the two priors becomes drastic whenever useless factors are included in the model (Panels B and C in Table 2). As discussed in Section II.2.1, since in this case the matrix $\boldsymbol{\beta}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is nearly singular and its determinant goes to zero, under a flat prior for risk premia the posterior probability of including a spurious factor in the model converges to 1 asymptotically. For example, the probability of misidentifying a spurious factor as being the true source of risk is almost 1 under flat prior, even for a very short sample. This in turn makes the overall process of model selection invalid.

Overall, we find the behavior of the spike-and-slab prior very encouraging for variable and model selection: it successfully eliminates the impact of the spurious factors from the model and identifies the true sources of risk.

IV Empirical Applications

In this section we apply our Bayesian approach to a large set of factors proposed in the previous literature. First, we use the Bayesian Fama-MacBeth method to analyze several notable factor models (subsection IV.1). Second, we consider 51 tradable and non-tradable factors, yielding more than two quadrillion possible models, and employ our spike-and-slab priors to compute factors' posterior probabilities and implied risk premia (subsections IV.2 and IV.3). Third, we compare the performance of a (low-dimensional) robust model, constructed with only the factors that have high posterior probability, to the one of several notable factor models (subsection IV.4). Fourth, we estimate the degree of sparsity (in terms of linear factors) of the true, latent SDF, as well as the SDF-implied maximum Sharpe ratio (subsection IV.5). Fifth, we evaluate the uncertainty arising from the choice of the cross-section test assets (subsection IV.6). Sixth, we analyze the out-of-sample performance of our method (subsection IV.7).

IV.1 Some notable factor models

In this section we illustrate the differences between the frequentist and Bayesian FM estimation (both OLS and GLS) for several candidate models. In particular, we estimate a set of linear factor models on the returns of the standard 25 Fama-French portfolios, sorted by size and value, using frequentist and Bayesian FM estimators. We use monthly data over the 1970:01–2017:12 sample for tradable factors and, whenever possible, non-tradables. For factors available only at quarterly frequency, the sample is 1952:Q1–2017:Q3 (whenever possible). A full description of the data and models used, as well as additional empirical results, can be found in Section OA.C.2 of the Online Appendix.

Tables 3 and 4 summarize the performance of several leading factor models. For the classical FM approach, we report point estimates of risk premia with their Shanken-corrected t-statistics, and the cross-sectional R^2, along with its 90% confidence interval (constructed following the methodology of Lewellen, Nagel, and Shanken (2010)). For BFM, we report the posterior mean of risk premia estimates, and the posterior median and mode of R^2, along with the centered 90% posterior coverage. We report both median and mode of the cross-sectional fit because its posterior distribution is often heavily skewed.

Carhart (1997) four-factor model: OLS and GLS Fama-MacBeth estimates of risk premia indicate that size, value, and momentum (SMB, HML, and UMD, respectively) are significant drivers of the cross-section of test assets. The market factor does not command a significant risk premium, which is a typical finding for this model. Cross-sectional fit seems to be high, with R^2 over 70%, even though it comes with rather wide confidence bounds, according to the Lewellen, Nagel, and Shanken (2010) approach. The Bayesian estimation indicates that part of the model success is due to the fact that this cross-section of test assets does not have much exposure to momentum, especially after one controls for the conventional Fama-French factors. While still marginally significant, its risk premium is substantially lower under both BFM and BFM-GLS estimations, with tighter bounds for R^2 as well. On the contrary, both HML and SMB have virtually identical risk prices under both FM and Bayesian estimations.

The *Hou, Xue, and Zhang (2014) q-factor model* emphasizes the role of investment (IA) and profitability (ROE) in matching the cross-section of equity returns, and we find these factors significantly priced using the frequentist inference. The Bayesian estimation delivers very similar risk premia estimates for most factors, but it finds weaker support for ROE being a significant explanator of the cross-section of returns, as well as lower values and tighter bounds, for the measures of fit.

The *Liquidity-Adjusted CAPM of Pastor and Stambaugh (2003)* seems to suffer from

Table 3: Tradable factors and 25 Fama-French portfolios, sorted by size and value

Model	Factors	FM $\hat{\lambda}_j$	FM R^2_{adj}	BFM $\bar{\lambda}_j$	BFM $R^2_{adj,mode}$	BFM $R^2_{adj,median}$
Panel A: OLS						
Carhart (1997)	Intercept	0.489	70.63	0.703*	64.32	63.29
		[-0.244, 1.222]	[31.60, 94.00]	[-0.061, 1.426]	[48.26, 76.46]	
	MKT	0.120		-0.101		
		[-0.631, 0.870]		[-0.822, 0.683]		
	SMB	0.171***		0.164***		
		[0.100, 0.241]		[0.089, 0.232]		
	HML	0.404***		0.396***		
		[0.331, 0.477]		[0.330, 0.466]		
	UMD	2.445***		1.806**		
		[0.955, 3.936]		[0.259, 3.328]		
q-factor model	Intercept	0.912***	65.67	0.922***	60.62	61.23
Hou, Xue, and Zhang (2014)		[0.286, 1.539]	[30.40, 86.80]	[0.276, 1.560]	[41.31, 76.40]	
	ROE	0.394**		0.377*		
		[0.016, 0.771]		[-0.020, 0.789]		
	IA	0.387***		0.385***		
		[0.203, 0.571]		[0.208, 0.580]		
	ME	0.274***		0.268***		
		[0.169, 0.379]		[0.158, 0.376]		
	MKT	-0.371		-0.378		
		[-0.995, 0.252]		[-1.005, 0.272]		
Liquidity-CAPM	Intercept	0.973*	36.24	1.162**	34.09	30.27
Pastor and Stambaugh (2000)		[-0.084, 2.030]	[-9.09, 100.00]	[0.175, 2.120]	[-2.39, 61.46]	
	LIQ	3.057**		1.785		
		[0.727, 5.388]		[-1.237, 4.150]		
	MKT	-0.281		-0.449		
		[-1.350, 0.788]		[-1.371, 0.509]		
Panel A: GLS						
Carhart (1997)	Intercept	1.017***	89.64	1.083***	85.87	86.3
		[0.389, 1.645]	[82.00, 97.60]	[0.458, 1.717]	[80.85, 91.05]	
	MKT	-0.434		-0.504		
		[-1.065, 0.196]		[-1.150, 0.122]		
	SMB	0.191***		0.189***		
		[0.150, 0.233]		[0.150, 0.230]		
	HML	0.356***		0.356***		
		[0.313, 0.400]		[0.316, 0.395]		
	UMD	1.626***		1.264**		
		[0.479, 2.772]		[0.077, 2.401]		
q-factor model	Intercept	1.305***	55.03	1.277***	47.28	48.54
Hou, Xue, and Zhang (2014)		[0.779, 1.831]	[24.40, 96.40]	[0.702, 1.879]	[32.45, 64.19]	
	ROE	0.295*		0.266		
		[-0.026, 0.615]		[-0.087, 0.640]		
	IA	0.270***		0.265***		
		[0.104, 0.437]		[0.093, 0.450]		
	ME	0.251***		0.246***		
		[0.161, 0.341]		[0.144, 0.345]		
	MKT	-0.749***		-0.720**		
		[-1.268, -0.229]		[-1.292, -0.156]		
Liquidity-CAPM	Intercept	1.244***	49.38	1.256***	52.98	43.17
Pastor and Stambaugh (2000)		[0.664, 1.824]	[26.91, 98.91]	[0.738, 1.749]	[12.50, 66.53]	
	LIQ	1.141		0.775		
		[-0.232, 2.514]		[-0.450, 2.116]		
	MKT	-0.664**		-0.678***		
		[-1.242, -0.086]		[-1.176, -0.162]		

Risk premia estimates and cross-sectional fit for a selection of models with tradable risk factors on a cross-section of 25 Fama-French monthly excess returns. Each model is estimated via OLS and GLS. We report point estimates and 95% confidence intervals for risk premia, which are constructed based on the asymptotic normal distribution, and cross-sectional R^2 and its (5%, 95%) confidence level constructed as in Lewellen, Nagel, and Shanken (2010) for FM estimation. For BFM estimation we report the posterior mean of λ ($\bar{\lambda}_j$), its (2.5%, 97.5%) credible interval, posterior mode and median of cross-sectional R^2_{adj}, and centered 90% credible intervals. *, ** and *** denote, respectively, 90%, 95% and 99% levels of significance.

Table 4: Non-tradable factors and 25 Fama-French portfolios, sorted by size and value

		FM		BFM		
Model	Factors	$\hat{\lambda}_j$	R^2_{adj}	$\bar{\lambda}_j$	$R^2_{adj,mode}$	$R^2_{adj,median}$
		Panel A: OLS				
Scaled CCAPM	Intercept	1.046	25.67	1.791**	34.36	29.19
Lettau and Ludvigson (2001)		[-0.848, 2.940]	[-14.29, 100.00]	[0.001, 3.723]	[-4.76, 62.07]	
	cay	1.817		0.791		
		[-0.653, 4.288]		[-1.347, 2.686]		
	ΔC_{nd}	0.713*		0.303		
		[-0.030, 1.456]		[-0.462, 0.951]		
	$\Delta C_{nd} \times cay$	0.804		0.301		
		[-1.645, 3.253]		[-1.911, 2.270]		
HC-CAPM	Intercept	3.243***	-1.22	3.090**	3.54	9.57
Jagannathan and Wang (1996)		[1.228, 5.257]	[-9.09, 33.45]	[0.790, 5.259]	[-7.48, 44.31]	
	ΔY	0.464		0.085		
		[-0.213, 1.140]		[-1.119, 1.058]		
	MKT	-0.719		-0.656		
		[-2.680, 1.242]		[-2.859, 1.558]		
Durable CCAPM	Intercept	2.214	52.38	2.780*	47.1	40.78
Yogo (2006)		[-1.037, 5.465]	[28.00, 100.00]	[-0.184, 5.751]	[1.20, 69.91]	
	ΔC_{nd}	0.743*		0.357		
		[-0.025, 1.511]		[-0.207, 0.832]		
	ΔC_d	-0.057		0.014		
		[-0.719, 0.605]		[-0.668, 0.693]		
	MKT	0.083		-0.495		
		[-3.322, 3.489]		[-3.395, 2.555]		
		Panel B: GLS				
Scaled CCAPM	Intercept	2.180***	-10.24	2.257***	-6.58	-3.13
Lettau and Ludvigson (2001)		[0.825, 3.536]	[-14.29, 64.57]	[1.221, 3.258]	[-11.87, 15.17]	
	cay	0.435		0.256		
		[-0.774, 1.643]		[-0.688, 1.217]		
	ΔC_{nd}	0.118		0.089		
		[-0.266, 0.502]		[-0.214, 0.407]		
	$\Delta C_{nd} \times cay$	0.141		0.063		
		[-1.005, 1.286]		[-0.845, 0.938]		
HC-CAPM	Intercept	2.730***	56.36	2.759***	58.24	49.26
Jagannathan and Wang (1996)		[1.458, 4.002]	[30.18, 83.64]	[1.379, 4.095]	[9.67, 75.07]	
	ΔY	-0.421**		-0.241		
		[-0.742, -0.099]		[-0.598, 0.114]		
	MKT	-0.717		-0.740		
		[-1.979, 0.545]		[-2.073, 0.622]		
Durable CCAPM	Intercept	2.960**	44.54	2.841***	54.74	40.99
Yogo (2006)		[0.547, 5.374]	[2.86, 78.29]	[1.102, 4.558]	[-2.41, 72.15]	
	ΔC_{nd}	0.105		0.052		
		[-0.265, 0.475]		[-0.201, 0.311]		
	ΔC_d	0.055		0.025		
		[-0.390, 0.501]		[-0.286, 0.327]		
	MKT	-0.941		-0.822		
		[-3.314, 1.432]		[-2.528, 0.895]		

Risk premia estimates and cross-sectional fit for a selection of models with non-tradable risk factors on a cross-section of 25 Fama-French monthly excess returns. Each model is estimated via OLS and GLS. We report point estimates and 95% confidence intervals for risk premia, which are constructed based on the asymptotic normal distribution, and cross-sectional R^2 and its (5%, 95%) confidence level constructed as in Lewellen, Nagel, and Shanken (2010) for FM estimation. For the BFM estimation we report the posterior mean of λ ($\bar{\lambda}_j$), its (2.5%, 97.5%) credible interval, posterior mode and median of cross-sectional R^2_{adj}, and centered 90% credible intervals. *, ** and *** denote, respectively, 90%, 95% and 99% levels of significance.

identification failure, as the risk premium on the liquidity factor is substantially reduced and no more significant when BFM is used in estimation. Wide confidence bounds and uncertain cross-sectional fit provide a stark difference to the pointwise estimates and their seemingly high significance levels under the standard frequentist approach.

The *Conditional CCAPM of Lettau and Ludvigson (2001)* appears weakly identified at best. Unlike the basic FM estimation, which indicates a relative empirical success of the model, the Bayesian approach reveals most risk premia to be substantially lower, losing all the accompanying statistical and economic significance. This is particularly pronounced in the BFM-GLS, which delivers both risk premia and cross-sectional R^2 close to zero.

The *Labour-Adjusted CAPM of Jagannathan and Wang (1996)* extends the classic CAPM framework by introducing a proxy for human capital and finds it strongly priced in the cross-sections of stocks returns. The BFM estimates of risk premia are substantially lower and no longer significant, with the same patterns observed under both OLS and GLS procedures.

The *Durable CCAPM of Yogo (2006)* in the linearized version, included the durable consumption factor and found that its impact is priced in a number of cross-sections sorted by size and value, past betas, and other characteristics. Even though the Lewellen, Nagel, and Shanken (2010) approach indicates a really wide support for the cross-sectional R^2, the model found empirical support in the data. We find that both durable and nondurable consumption are weak predictors of the cross-section of returns, as the magnitude of their risk premia substantially declines and is no longer significant. The model is still characterized by a wide confidence interval for R^2, but overall its pricing ability is questionable at best.

The Online Appendix (Tables OA14–OA17) provides additional empirical results on the performance of both frequentist and BFM estimators applied to notable factor models. In many cases, when the models are well specified and strongly identified in the data, there is almost no distinction between the two approaches. One notable difference, however, are the confidence intervals of the R^2, which are often notoriously wide in the frequentist case. There are also cases, however, in which the difference in model performance becomes large, affecting both risk premia estimates and measures of cross-sectional fit. Similar to Gospodinov, Kan, and Robotti (2019), we caution the reader against blindly relying on the estimates produced by conventional Fama-MacBeth procedure, and we advocate a robust approach to inference.

IV.2 Sampling two quadrillion models

We now turn our attention to a large cross-section of candidate asset pricing factors. In particular, we focus on 51 (both tradable and non-tradable) monthly factors available from October 1973 to December 2016 (i.e. $T \simeq 600$). Factors are described in Table A1 in the Appendix, with additional details available in Table OA13 of the Online Appendix. In choosing the cross-section of assets to price, we follow Lewellen, Nagel, and Shanken (2010) and employ 25 Fama-French size and book-to-market portfolios plus 30 industry portfolios (i.e., $N = 55$).[22] Since we do not restrict the maximum number of factors to be included, all the possible combinations of factors give us a total of 2^{51} possible specifications, that is 2.25 quadrillion models. Note that each model involves 55 time series regressions and one cross-section regression, that is, we jointly evaluate the equivalent of 126 quadrillion regressions.

We employ the continuous spike-and-slab approach of Section II.2.3, since it is the most suited for handling a very large number of possible models, and report both the posterior probability (given the data) of each factor (that is, $\mathbb{E}[\gamma_j|\text{data}], \forall j$) as well as the posterior means of the factors' risk premia (that is, $\mathbb{E}[\lambda_j|\text{data}], \forall j$) computed as the Bayesian Model Average (BMA)[23] across all the models considered. We use the formulation of the penalty term ψ_j in equation (21) in order to handle also identification failures of factors' risk premia caused by level factors (see Remark 3).[24]

The posterior evaluation is performed and reported over a wide range for the parameter (ψ in equation (19)) that controls the degree of shrinkage of potentially useless factors' risk premia: from $\psi = 1$ (that is, very strong shrinkage) to $\psi = 100$ (making the shrinkage virtually irrelevant). As discussed in Section II.2.3, we consider a value of ψ in the 10–20 range as a reasonable benchmark.

The prior probability for each factor inclusion is drawn from a $Beta(1,1)$ (that is, a uniform on $[0,1]$), yielding a prior expectation for γ_j equal to 50%, that is, a priori we have maximum uncertainty about whether a factor should be included or not.[25]

Figure 4 plots the posterior probabilities of the 51 factors as a function of the parameter

[22] In Section IV.6 below we extend our analysis to 24 additional cross-sections of test assets based on the portfolios most commonly used in the empirical literature.

[23] If we are interested in some quantity Δ that is well-defined for every model $m = 1, ..., M$ (e.g., risk premia, maximum Sharpe ratio), from the Bayes theorem, we have

$$\mathbb{E}[\Delta|\text{data}] = \sum_{m=0}^{M} \mathbb{E}[\Delta|\text{data}, \text{model} = m] \Pr(\text{model} = m|\text{data}),$$

where $\mathbb{E}[\Delta|\text{data}, \text{model} = m] = \lim_{L\to\infty} \frac{1}{L}\sum_{l=1}^{L} \Delta(\theta_l^{(m)})$ and $\left\{\theta_l^{(m)}\right\}_{l=1}^{L}$ denote L draws from the posterior distribution of the parameters of model m. That is, the (BMA) expectation of Δ, conditional on only the data, is simply the weighted average of the expectation in every model, with weights equal to the models' posterior probabilities. See, e.g., Raftery, Madigan, and Hoeting (1997), Hoeting, Madigan, Raftery, and Volinsky (1999).

[24] In Online Appendix OA.C.3 we report results based on the formulation in equation (21)) as well as the Fisher transformation of the correlation coefficients. The findings therein are very similar to the ones discussed below.

[25] Using a $Beta(2,2)$, which still implies a prior probability of factor inclusion of 50%, but lower probabilities for very dense and very sparse models, we obtain virtually identical results. Furthermore, using a prior in favor of more sparse factor models (that is, a $Beta(2,8)$), the empirical findings are very similar to the ones reported. These additional results are reported in Section OA.C.3 of the Online Appendix.

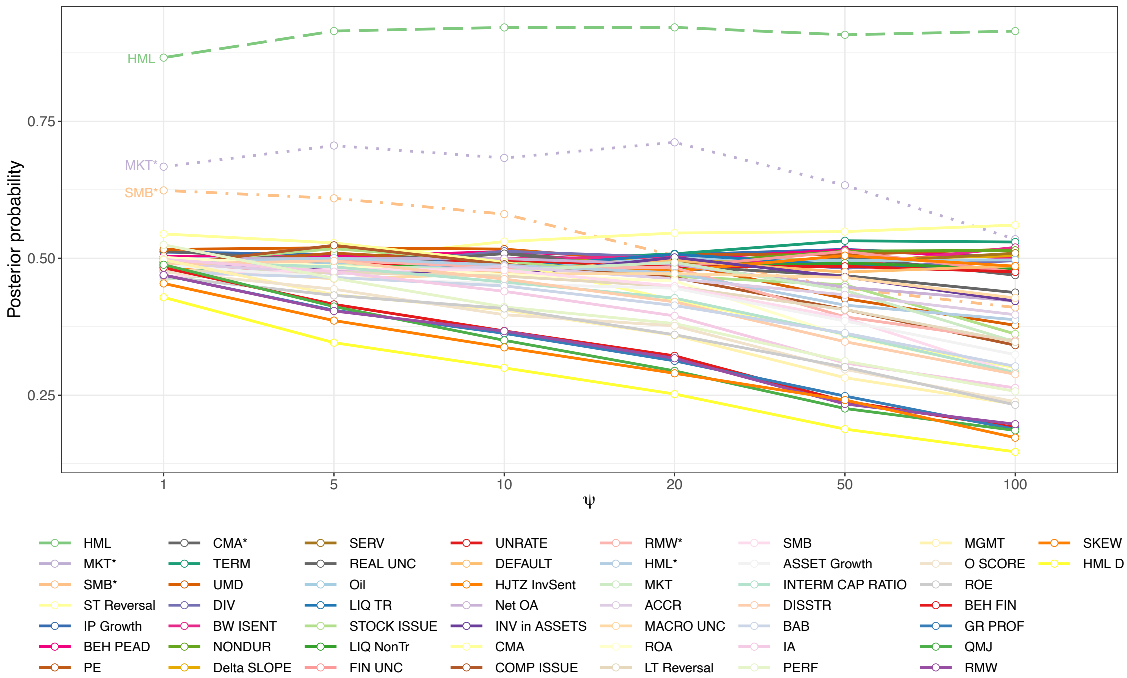


Figure 4: Posterior factor probabilities

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$ computed using the continuous spike-and-slab approach of Section II.2.3 and 51 factors described in Table A1 of the Appendix. Sample: 1973:10–2016:12. Test assets: 25 Fama-French size-and-book-to-market and 30 industry portfolios. Prior distribution for the j-th factor inclusion is a $Beta(1, 1)$, yielding a 0.5 prior expectation for γ_j. Posterior probabilities are plotted for $\psi \in [1,\ 100]$.

ψ. The corresponding values are reported in Table 5. Overall, the inclusion of only three factors finds substantial support in our empirical analysis. First, the celebrated Fama-French HML (high-minus-low), designed to capture the so-called "value premium", is a strong determinant of the cross-section of asset returns. For $\psi = 10$ (a reasonable benchmark), its posterior probability is about 92.1%, and only for very strong shrinkage ($\psi = 1$) the posterior probability gets reduced to 86.6%. Second, the market factor, in the version of Daniel, Mota, Rottke, and Santos (2020) (MKT*, which is meant to have hedged out the unpriced risk contained in the market index), has also high posterior probability (68.3% for $\psi = 10$). Instead, the simple market factor (MKT) seems to be driven out by MKT*. Third, albeit to a lesser extent, SMB*, the Daniel, Mota, Rottke, and Santos (2020) version of the small-minus-big Fama-French factor (meant to capture the so-called "size" premium), seems also to contain relevant information for pricing the cross-section of asset returns, with a posterior probability in the 51%–62% range for small values of ψ. Beside the ones mentioned above, all other factors have posterior probabilities of about 50% or less for all values of ψ. Interestingly, the results are not very sensitive to the choice of ψ.

In addition to the posterior probabilities of the factors, Table 5 reports the posterior means of the factor risk premia computed as Bayesian Model Average (BMA), that is,

Table 5: Posterior factor probabilities, $\mathbb{E}[\gamma_j|\text{data}]$, and risk premia. 2.25 quadrillion models

	$\mathbb{E}[\gamma_j\|\text{data}]$ ψ:						$\mathbb{E}[\lambda_j\|\text{data}]$ ψ:						
Factors:	1	5	10	20	50	100	1	5	10	20	50	100	$\bar{F}$
HML	0.866	0.915	0.921	0.921	0.908	0.915	0.173	0.263	0.281	0.290	0.292	0.300	0.377
MKT*	0.667	0.706	0.683	0.712	0.633	0.535	0.074	0.170	0.207	0.259	0.268	0.229	0.514
SMB*	0.624	0.609	0.581	0.505	0.446	0.410	0.057	0.105	0.115	0.105	0.104	0.108	0.215
STRev	0.513	0.498	0.530	0.546	0.549	0.561	0.003	0.013	0.025	0.047	0.095	0.149	0.438
IPGrowth	0.511	0.507	0.488	0.506	0.516	0.502	0.000	-0.001	-0.001	-0.002	-0.005	-0.008	0.097*
BEH_PEAD	0.503	0.503	0.512	0.499	0.515	0.500	0.003	0.010	0.016	0.025	0.048	0.070	0.619
PE	0.486	0.509	0.494	0.508	0.507	0.517	-0.001	-0.003	-0.004	-0.005	-0.011	-0.019	6.770*
CMA*	0.513	0.495	0.509	0.486	0.468	0.437	0.001	0.000	-0.002	-0.004	-0.008	-0.010	0.242
TERM	0.477	0.478	0.494	0.508	0.532	0.530	0.001	0.003	0.006	0.011	0.024	0.038	0.962*
UMD	0.516	0.519	0.517	0.492	0.426	0.378	0.019	0.050	0.067	0.082	0.091	0.098	0.646
DIV	0.491	0.484	0.513	0.502	0.482	0.496	0.000	0.000	-0.001	-0.001	-0.003	-0.005	0.926*
BW_ISENT	0.494	0.500	0.500	0.502	0.487	0.520	0.000	0.002	0.003	0.005	0.009	0.014	0.101*
NONDUR	0.487	0.478	0.494	0.490	0.513	0.515	0.000	0.002	0.003	0.005	0.011	0.019	0.151*
DeltaSLOPE	0.488	0.491	0.494	0.497	0.498	0.505	0.000	0.000	-0.001	-0.001	-0.003	-0.006	0.059*
SERV	0.493	0.494	0.489	0.491	0.491	0.509	0.000	0.000	0.000	0.000	-0.001	-0.001	0.045*
REAL_UNC	0.499	0.484	0.509	0.476	0.503	0.469	0.000	0.000	0.000	0.000	0.000	0.000	0.046*
Oil	0.491	0.491	0.482	0.500	0.496	0.497	0.002	0.011	0.020	0.037	0.074	0.126	0.740*
LIQ_TR	0.484	0.496	0.483	0.507	0.488	0.481	0.000	0.003	0.007	0.015	0.033	0.055	0.438
STOCK_ISS	0.491	0.517	0.494	0.467	0.452	0.362	-0.024	-0.062	-0.076	-0.088	-0.103	-0.096	0.515
LIQ_NT	0.493	0.482	0.497	0.483	0.491	0.481	-0.002	-0.002	-0.001	0.003	0.014	0.015	0.428*
FIN_UNC	0.484	0.479	0.483	0.484	0.513	0.475	0.000	0.000	0.000	0.000	0.000	-0.001	0.103*
UNRATE	0.488	0.487	0.497	0.484	0.485	0.475	0.000	-0.001	-0.002	-0.003	-0.006	-0.007	1.157*
DEFAULT	0.501	0.477	0.476	0.496	0.475	0.486	0.000	0.000	0.000	0.000	0.000	0.000	0.333*
HJTZ_ISENT	0.481	0.499	0.485	0.477	0.505	0.486	0.000	-0.001	-0.001	-0.001	-0.002	-0.002	0.242*
NetOA	0.499	0.499	0.499	0.489	0.447	0.422	0.006	0.018	0.028	0.040	0.052	0.056	0.544
INV_IN_ASSETS	0.492	0.480	0.468	0.502	0.466	0.422	0.001	0.003	0.005	0.010	0.017	0.024	0.549
CMA	0.544	0.528	0.493	0.424	0.362	0.300	0.026	0.050	0.057	0.054	0.051	0.044	0.351
COMP_ISSUE	0.487	0.524	0.489	0.465	0.407	0.341	0.028	0.071	0.084	0.094	0.098	0.094	0.497
RMW*	0.487	0.496	0.483	0.483	0.393	0.350	0.003	0.012	0.018	0.024	0.025	0.024	0.219
HML*	0.492	0.500	0.487	0.473	0.414	0.388	-0.001	0.002	0.006	0.011	0.015	0.021	0.251
MKT	0.461	0.488	0.478	0.492	0.441	0.347	0.019	0.075	0.107	0.144	0.158	0.126	0.563
ACCR	0.492	0.472	0.485	0.462	0.433	0.397	-0.005	-0.005	-0.001	0.006	0.014	0.008	0.343
MACRO_UNC	0.481	0.463	0.471	0.472	0.465	0.427	0.000	0.000	0.000	0.000	0.000	0.000	0.078*
ROA	0.502	0.486	0.471	0.459	0.358	0.293	0.033	0.082	0.105	0.124	0.108	0.086	0.551
LTRev	0.476	0.474	0.468	0.448	0.406	0.348	-0.005	-0.019	-0.024	-0.029	-0.027	-0.024	0.252
SMB	0.466	0.480	0.478	0.449	0.391	0.290	0.045	0.067	0.073	0.074	0.068	0.052	0.257
ASSET_Growth	0.488	0.465	0.453	0.445	0.386	0.324	-0.001	0.000	0.001	-0.001	-0.002	0.001	0.525
INTERM_CAP_RATIO	0.494	0.484	0.457	0.427	0.361	0.291	0.022	0.030	0.034	0.046	0.052	0.042	0.719*
DISSTR	0.489	0.495	0.462	0.420	0.348	0.288	0.035	0.086	0.105	0.112	0.108	0.099	0.475
BAB	0.475	0.465	0.450	0.414	0.363	0.303	-0.011	-0.023	-0.024	-0.021	-0.013	-0.009	0.921
IA	0.498	0.476	0.440	0.395	0.309	0.263	-0.011	-0.014	-0.012	-0.011	-0.008	-0.008	0.409
PERF	0.525	0.463	0.411	0.381	0.312	0.257	-0.038	-0.055	-0.053	-0.053	-0.047	-0.047	0.651
MGMT	0.497	0.434	0.405	0.360	0.282	0.234	0.017	0.035	0.041	0.041	0.037	0.034	0.631
O_SCORE	0.475	0.443	0.397	0.377	0.297	0.238	-0.010	-0.015	-0.015	-0.015	-0.012	-0.004	0.02
ROE	0.467	0.432	0.408	0.360	0.301	0.232	0.005	0.015	0.020	0.023	0.022	0.017	0.555
BEH_FIN	0.483	0.416	0.367	0.322	0.239	0.191	-0.005	0.003	0.006	0.010	0.008	0.010	0.76
GR_PROF	0.469	0.405	0.363	0.313	0.248	0.188	0.012	0.011	0.011	0.011	0.011	0.008	0.199
QMJ	0.488	0.412	0.350	0.294	0.226	0.186	0.020	0.021	0.019	0.015	0.013	0.013	0.405
RMW	0.470	0.404	0.367	0.317	0.234	0.197	0.012	0.022	0.025	0.026	0.022	0.019	0.292
SKEW	0.454	0.386	0.337	0.290	0.241	0.173	0.000	0.000	0.000	0.000	0.000	0.000	0.438
HML_DEVIL	0.429	0.346	0.300	0.252	0.188	0.147	0.009	0.013	0.010	0.007	0.004	0.003	0.356

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, and posterior mean of factor risk premia, $\mathbb{E}[\lambda_j|\text{data}]$ are computed using the continuous spike-and-slab approach of Section II.2.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1,1)$, yielding a prior expectation for γ_j equal to 50%. The last column reports sample average returns for the tradable factors. The data is monthly, 1973:10 to 2016:12. Test assets: cross-section of 25 Fama-French size-and-book-to-market and 30 industry portfolios. The 51 factors considered are described in Table A1 of the Appendix. Numbers denoted with the asterisk in the last column correspond to the return on the factor-mimicking portfolio of the non-tradable factor, constructed by a linear projection of its values on the set of 51 test assets, and scaled to have the same volatility as the original non-tradable factor.

the weighted average of the posterior means in each possible factor model specification, with weights equal to the posterior probability of each specification being the true data generating process (see, e.g., Roberts (1965), Geweke (1999), Madigan and Raftery (1994)). The results are not very sensitive to the choice of ψ, except when considering very small values of the shrinkage parameter, since in this case posterior means are shrunk toward zero. Interestingly, the estimated price of risk for the market factor (MKT*) is positive, despite it being very often estimated as a negative quantity when considering multifactor models and not dissimilar from the market excess return over the same period. More generally, there is a clear pattern in cross-sectionally estimated (i.e., ex post) factor risk premia and their simple time series average estimates (reported in the last column of Table 5): for the robust factors (HML, MKT*, and SMB*) ex post risk premia are very similar to the time series estimates multiplied by their posterior probabilities, while the opposite holds true for the other factors. In other words, robust factors seems to price themselves well (since, theoretically, their own beta is one), while other factors do not.

IV.3 Estimating 2.6 million sparse factor models

Instead of drawing unrestricted factor models specifications, we now constraint models to have a maximum of five factors, that is, we impose sparsity on the linear factor models, as in most of the previous empirical literature that has tried to identify low-dimensional factor models to explain the cross-section of asset returns. Given our set of 51 factors, this approach yields about 2.6 million candidate models (i.e., the equivalent of about 147 million time series and cross-sectional regressions). Since we now do not sample the possible models, posterior probabilities are computed using the marginal likelihoods of all these models, that is, the posterior probability of model $\boldsymbol{\gamma}_j$ is computed as

$$\Pr(\boldsymbol{\gamma}_j|data) = \frac{p(data|\boldsymbol{\gamma}_j)}{\sum_i p(data|\boldsymbol{\gamma}_i)},$$

where we have assigned equal prior probability to all possible specifications, and $p(data|\boldsymbol{\gamma}_j)$ denotes the marginal likelihood of the j-th model. To both simplify the numerical computation, and to illustrate the qualities of the approach, we use the Dirac spike-and-slab prior of Section II.2.2 since we can leverage its closed form solution for the marginal likelihoods.[26]

Posterior factor probabilities are reported in Figure 5 for a range of values of ψ. Table A2 of the Appendix reports these probabilities jointly with the BMA of risk premia across the sparse candidate models. Note that in this case each factor has an ex ante probability of

[26] Alternatively, one could use the continuous spike-and-slab approach employed in the previous subsection, and drop the draws of specifications with more than five factors, but this significantly reduces computational efficiency.

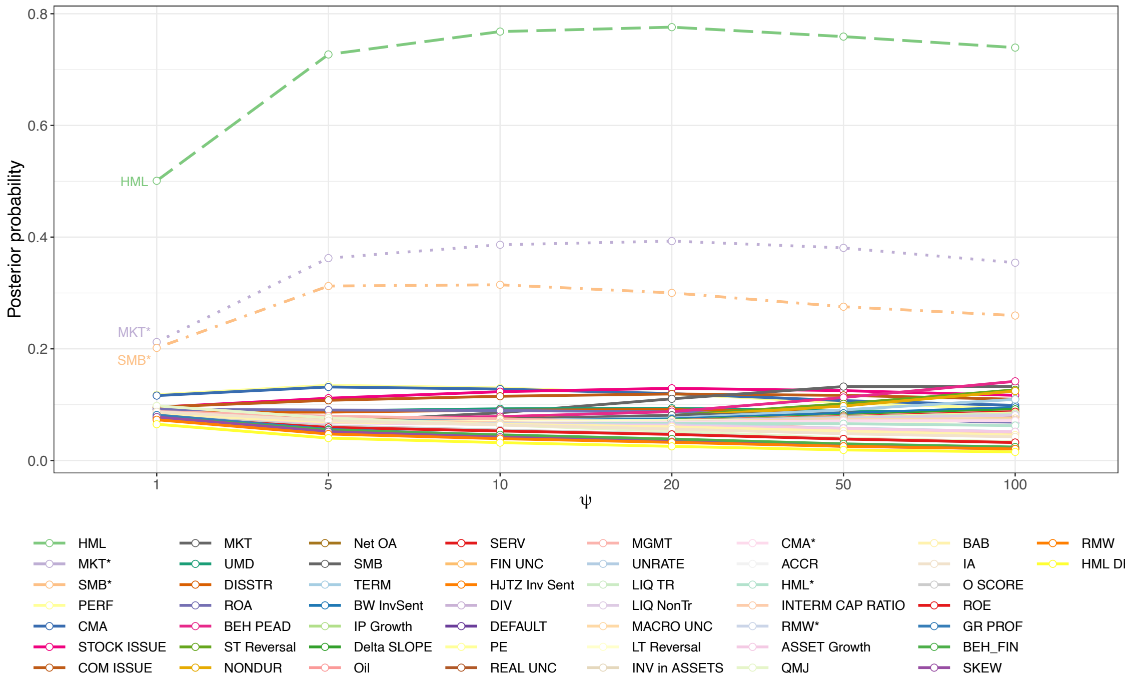


Figure 5: Posterior factor probabilities

Posterior probabilities of factors, $\Pr[\gamma_j = 1|\text{data}]$, computed using the the Dirac spike-and-slab approach of section II.2.2 and 51 factors described in Table A1 of Appendix. Sample: 1973:10-2016:12. Test assets: 25 Fama-French size and book-to-market and 30 Industry portfolios. Prior probability of a factor being included is about 10.38%. Posterior probabilities plotted for $\psi \in [1, \ 100]$.

being included equal to 10.38%. The results are strikingly similar to the ones in Table 5 and Figure 4: as before, only for three factors (HML, MKT*, and SMB*) we observe a marked increase in the posterior probability of inclusion after observing the data. Furthermore, these factors seems to price themselves well, that is, the BMA of their risk premia are very similar to the sample average of their excess returns scaled by the factors' posterior probabilities, while other factors do not.

IV.4 A robust factors model

The previous subsections suggest that only a small number of factors – HML, MKT*, and, to a lesser extent, SMB* – are robust explanators of the cross-section of asset returns. Furthermore, Table 6 reports 10 factor model specifications with the highest posterior probabilities under the continuous spike-and-slab approach (with $\psi = 20$).[27] It shows that the three robust factors tend to be almost always included in the most likely models: HML is featured in all 10 specifications, and both MKT* and SMB* are included, respectively, in seven and eight of the most likely models. Note that the model posterior probabilities in Table 6 might

[27] A $\psi = 10$, as shown in Table OA19 of the Online Appendix, yields very similar results.

Table 6: Factor models with highest posterior probability (continuous spike-and-slab)

	model:									
factor:	1	2	3	4	5	6	7	8	9	10
HML	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MKT*	✓	✓		✓	✓	✓	✓	✓		✓
SMB*	✓		✓	✓	✓	✓		✓	✓	✓
STRev		✓		✓	✓	✓		✓	✓	
IPGrowth	✓	✓			✓	✓				
BEH_PEAD		✓	✓			✓	✓		✓	✓
PE	✓		✓		✓	✓			✓	
CMA*	✓		✓				✓	✓	✓	
TERM				✓	✓		✓	✓	✓	
UMD			✓				✓		✓	✓
DIV	✓	✓	✓	✓	✓				✓	✓
BW_ISENT	✓			✓	✓			✓		
NONDUR			✓	✓	✓					
DeltaSLOPE	✓		✓			✓		✓		
SERV			✓	✓	✓	✓		✓	✓	✓
REAL_UNC	✓	✓	✓	✓	✓				✓	
Oil	✓		✓	✓			✓			✓
LIQ_TR	✓	✓		✓	✓					✓
STOCK_ISS	✓	✓	✓	✓		✓			✓	
LIQ_NT	✓	✓				✓		✓	✓	
FIN_UNC	✓		✓	✓	✓					
UNRATE			✓		✓			✓		
DEFAULT	✓	✓	✓	✓	✓				✓	✓
HJTZ_ISENT			✓		✓		✓	✓		
NetOA	✓			✓	✓	✓			✓	✓
INV_IN_ASSETS		✓			✓		✓	✓		✓
CMA	✓	✓				✓	✓		✓	✓
COMP_ISSUE			✓		✓	✓	✓	✓		
RMW*	✓			✓	✓	✓	✓	✓	✓	✓
HML*			✓	✓		✓				✓
MKT		✓			✓			✓		
ACCR	✓	✓	✓	✓		✓	✓			✓
MACRO_UNC		✓		✓					✓	✓
ROA					✓	✓		✓	✓	
LTRev				✓	✓			✓	✓	✓
SMB		✓			✓					✓
ASSET_Growth	✓		✓				✓			
INTERM_CAP_RATIO		✓		✓	✓	✓			✓	✓
DISSTR										
BAB			✓	✓	✓		✓	✓		✓
IA						✓				
PERF	✓			✓	✓			✓		✓
MGMT	✓			✓					✓	✓
O_SCORE		✓	✓		✓		✓		✓	✓
ROE	✓			✓						
BEH_FIN		✓	✓							
GR_PROF					✓	✓			✓	✓
QMJ									✓	✓
RMW				✓	✓				✓	
SKEW	✓	✓								
HML_DEVIL										
Probability (%)	0.0133	0.0111	0.0111	0.0111	0.0100	0.0100	0.0100	0.0100	0.0100	0.0089

Factors and posterior model probabilities of 10 most likely specifications, computed using the continuous spike-and-slab approach of Section II.2.3, $\psi = 20$, 51 factors, and 2.25 quadrillion models and a model prior probability of the order of 10^{-16}. Specifications organized by columns with the symbol ✓ indicating that the factor in the corresponding row is included. The data is monthly, 1973:10 to 2016:12. Test assets: cross-section of 25 Fama-French size-and-book-to-market and 30 industry portfolios. The 51 factors considered are described in Table A1 of the Appendix.

appear small in absolute terms but are actually of magnitude much larger than the prior model probabilities (equal to one over the number of models considered – 2.25 quadrillion).

Table 7: Posterior probabilities of notable models versus robust factors

	ψ:					
Model:	1	5	10	20	50	100
CAPM	0.02	0.01	0.01	0.01	0.03	0.07
Fama and French (1992)	0.05	0.01	0.01	0.01	0.01	0.00
Fama and French (2016)	0.09	0.06	0.05	0.04	0.03	0.02
Carhart (1997)	0.05	0.01	0.01	0.01	0.01	0.01
Hou, Xue, Zhang (2015)	0.02	0.01	0.00	0.00	0.00	0.00
Pastor and Stambaugh (2000)	0.02	0.01	0.01	0.01	0.02	0.02
Asness, Frazzini and Pedersen (2014)	0.10	0.06	0.04	0.04	0.03	0.02
Robust Factors Model	0.64	0.85	0.87	0.88	0.88	0.86

Posterior model probabilities for the specifications in the first column, for different values of ψ, computed using the Dirac spike-and-slab prior. Models and their factors are described in Table A1 of the Appendix. The model in the last row uses the HML, MKT* and SMB* factors described in Table A1. Sample: 1973:10 to 2016:12. Test assets: 25 Fama-French size-and-value and 30 Industry portfolios.

Therefore, a natural question is whether the three factors identified as robust do indeed deliver a significantly better cross-sectional asset pricing model. We answer this question by comparing the performance of a three-factor model with HML, MKT*, and SMB* as factors, to the one of several notable factor models. In particular, Table 7 reports the model posterior probabilities for the specifications considered, that is, the probability of any of these models being the true data-generating process. Strikingly, for almost any value of ψ, the model posterior probabilities are in the single-digits range for all models but the robust factors one: the probability of this specification is always higher than 85% except when using a very strong shrinkage (in which case it is reduced to 64%). Furthermore, for ψ in the most salient range (10–20), the posterior probability of the robust factors model is about 90%.

IV.5 Model uncertainty and sparsity

The results above on the robust factor model might give the illusion that the data clearly favors one specification. However, this result emerges only when the robust factor model is compared to the set of alternative specifications proposed in the literature. When confronted with the entire universe of possible specifications, the results are quite different. Figure 6 presents the model posterior probabilities of the most likely 1,000 specifications. The first thing to notice is that even the most likely specification is not a clear winner within the set of all possible models – its posterior probability is only 0.0133%. This is a remarkable improvement relative to the prior model probability that is of the order of 10^{-16}, but it clearly does not represent a substantial resolution of model uncertainty. Furthermore, the posterior model probability decays very slowly as we move down the list of most likely models: moving from the best model, it takes about 100–200 models to reach the relative odds of 2:1 (i.e., to reduce the posterior probability by 50%),[28] and at least 750 models to reach the odds of 3:1. Finally, and remarkably, note that the robust three-factor model, which in Table 7 is one order of magnitude more likely to be the true model than the models proposed in the previous literature, is *not* among the top 1,000 models depicted in Figure 6.

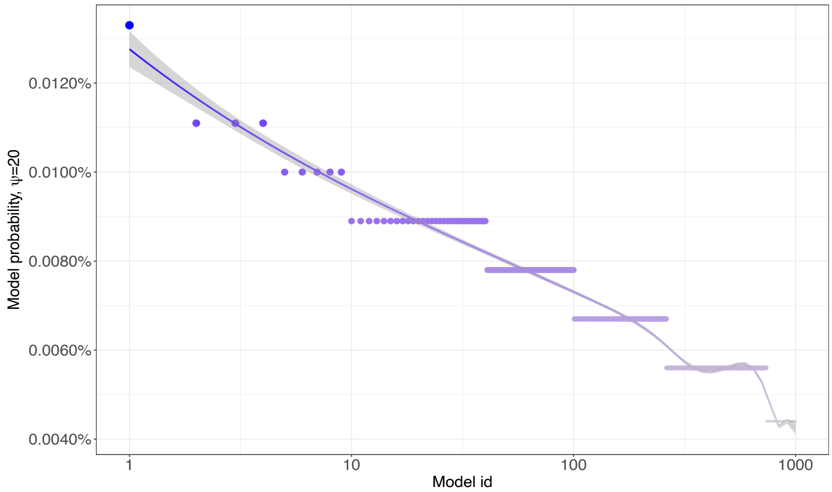


Figure 6: Posterior model probabilities

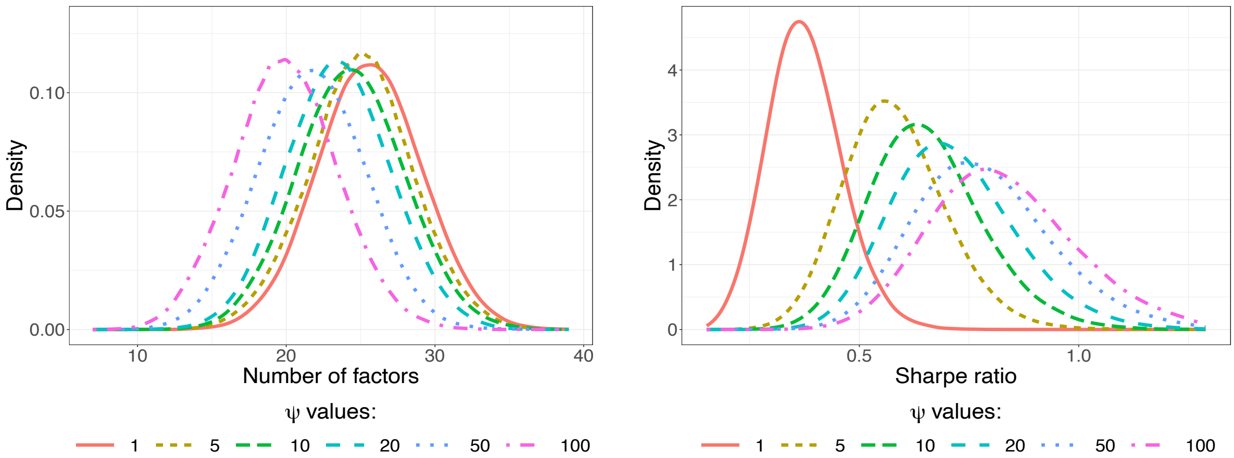
Posterior model probabilities of the 1,000 most likely models computed using the continuous spike-and-slab of Section II.2.3 and 51 factors. The horizontal axis uses a log scale. Sample: 1973:10–2016:12. Test assets: 25 Fama-French size-and-book-to-market and 30 industry portfolios.

But how many of the factors proposed in the literature does it really take to price the cross-section? Thanks to our Bayesian method, even this question can be easily answered. In particular, by using our estimations of about 2.25 quadrillion models and their posterior probabilities, we can compute the posterior distribution of the dimensionality of the "true" model. That is, for any integer number between one and 51, we can compute the posterior probability of the linear factor model being a function of that number of factors.

Figure 7a reports the posterior distributions of the model dimensionality for various values of ψ (these distributions are also summarized in Table 8a). For the most salient values of ψ (10 and 20), the posterior mean of the number of factors in the true model is in the 24–25 range, and the 95% posterior credible intervals are contained in the 17 to 31 factors range. That is, there is substantial evidence that the linear factor model is *dense* in the space of factors considered: given the factors at hand, a relative large number of them is needed to provide an accurate representation of the "true" model. Since most of the literature has focused on very low-dimensional linear factor models, this finding suggests that most empirical results therein have been affected by a very large degree of misspecification.

[28]That is, to a first-order approximation, the frequentist likelihood ratio test of the best performing model versus the 100^{th} one would yield a p-value of 25% at best.

It is worth noticing that, as Figure 7a shows, for very large ψ, that is, with basically a flat prior for factor risk premia, the posterior dimensionality is reduced. This is due to two phenomena we have already outlined. First, if some of the factors are useless (and our analysis points in this direction), under a flat prior they tend to have a higher posterior probability and drive out the true sources of priced risk. Second, a flat prior for the risk premia can generate a "Bartlett Paradox" (see the discussion in Section II.2.1).



(a) Model dimensionality **(b)** Model-implied Sharpe ratio

Figure 7: Posterior densities of model dimensionality and its implied Sharpe ratio.

The left graph presents the posterior density of the true model having the number of factors listed on the horizontal axis. The right graph demonstrates posterior density of the (annualized) Sharpe ratio implied by the linear factor model for various values of $\psi \in [1,\ 100]$. All the parameters are estimated over the 1973:10-2016:12 sample using a cross-section of 25 Fama-French size-and-book-to-market and 30 industry test asset portfolios, computed using the continuous spike-and-slab approach of Section II.2.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1,1)$, yielding a prior expectation for γ_j equal to 50%. The 51 factors considered are described in Table Table A1 of the Appendix.

Table 8: Posterior model dimensionality and its implied Sharpe ratio.

	(a) Number of factors						**(b)** Model-implied Sharpe ratio					
	ψ:						ψ:					
	1	5	10	20	50	100	1	5	10	20	50	100
mean	25.6	25.0	24.4	23.5	21.7	19.9	0.38	0.58	0.65	0.72	0.79	0.83
median	26	25	24	23	22	20	0.37	0.57	0.65	0.71	0.77	0.82
2.5%	19	18	17	17	15	13	0.22	0.38	0.43	0.47	0.51	0.54
5%	20	19	19	18	16	14	0.24	0.41	0.46	0.51	0.55	0.58
95%	31	31	30	29	28	26	0.52	0.78	0.88	0.97	1.07	1.13
97.5%	32	32	31	30	29	27	0.55	0.83	0.94	1.03	1.13	1.20

Summary statistics for posterior number of the factors included in the model and the model-implied Sharpe ratio. Both are summarized for values of $\psi \in [1,\ 100]$. All the parameters are estimated over the 1973:10-2016:12 sample using a cross-section of 25 Fama-French size-and-value and 30 industry portfolios, computed using the continuous spike-and-slab approach of Section II.2.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1,1)$, yielding a prior expectation for γ_j equal to 50%. The 51 factors considered are described in Table A1 of the Appendix.

Note that if the factors proposed in the literature were to capture different and uncorrelated sources of risk, one might worry that a dense model in the space of factors could imply unrealistically high Sharpe ratios (see, e.g., the discussion in Kozak, Nagel, and Santosh (2019)). Since, given a factor model, the SDF-implied maximum Sharpe ratio is just a function of the factors' risk premia and covariance matrix, our Bayesian method allows to construct the posterior distribution of the maximum Sharpe ratio for each of the 2.25 quadrillion models considered. Therefore, using the posterior probabilities of each possible model specification, we can actually construct the (BMA) posterior distribution of the SDF-implied maximum Sharpe ratio (conditional on the data only).

Figure 7b (and Table 8b) reports, respectively, the posterior distribution of the SDF-implied maximum Sharpe ratio (annualized) and its summary statistics for several values of the parameter ψ. Except when a very strong shrinkage (small ψ) is imposed (and, hence, risk premia, and consequently Sharpe ratios, are shrunk toward zero) the posterior distributions of the Sharpe ratio are quite similar for all values of ψ. Furthermore, despite the model being dense in the space of factors, the posterior maximum Sharpe ratio does not appear to be unrealistically high: for example, for $\psi \in [10, 20]$ its posterior mean is about 0.65–0.72, and the 95% posterior credible intervals are in the 0.43–1.03 range. Interestingly, Ghosh, Julliard, and Taylor (2016, 2018) provide a nonparametric estimate of the pricing kernel, extracted using an information-theoretic approach and wide cross-sections of equity portfolios, and find SDF-implied maximum Sharpe ratios of very similar magnitude.

IV.6 Cross-sectional uncertainty

All our results so far relied on using a single cross-section: 25 Fama-French portfolios sorted by size and value and 30 industry portfolios. Since all the estimates of the risk premia and factor posterior probabilities have been constructed *conditional on the cross-section*, it is natural to wonder whether the findings are cross-section-specific or reflective of the general features of the data. Since the issue of building an optimal cross-section for a given set of factors is still largely an open question, our approach is based on using *revealed preferences* instead.

First, we analyze a large number of papers proposing and testing a new linear factor model for cross-sectional asset pricing.[29] Specifically, we focus on all the papers belonging to the following pools: articles published in (12) major economics and finance journal (2007–2018); papers presented at (14) major economics and finance conferences (2006–2018); SSRN working papers (2005–2018); articles and working papers mentioned in Harvey, Liu, and Zhu

[29]We focus only on papers that use linear factor models for pricing cross-sections. We do not consider papers that only construct a tradable strategy and demonstrated that it delivers a substantial alpha.

(2016).[30] We then identify the main specification used in each of the papers and record the *base* cross-sections used. For example, whenever authors choose to estimate the model on the set of 25 portfolios sorted by size and value and 17 industry, we increase the base count for each of these cross-sections. To ensure saliency, we considered only those base test assets that have been used for the main specification in at least three different papers. Table 9 summarizes the empirical counts and frequencies for the nine base cross-sections identified in the data. As expected, 25 Fama-French portfolios sorted by size and value compose the most widespread base cross-section used for empirical analysis, followed by momentum and different selection of the industry portfolios.

Table 9: Base cross-sectional assets

Portfolio set	Number of papers	Frequency	Portfolio set	Number of papers	Frequency
25 size-and-book-to-market	64	54.24%	49 industry	4	3.39%
10 momentum	13	11.02%	25 siz-and-momentum	3	2.54%
10 book-to-market	8	6.78%	5 book-to-market	3	2.54%
10 size	6	5.08%	5 industry	3	2.54%
30 industry	4	3.39%			

The table presents the list of base test assets used in the main specifications from the collection of papers on cross-sectional asset pricing, satisfying criteria outlined in Section IV.6. For each of the base cross-section the table reports the number of papers using the corresponding assets and their empirical frequency.

Second, we combine these base cross-sections into composite sets of test assets large enough for us to evaluate the contribution of different factors. In doing so, we try to match the original frequencies of the base test assets, while also keeping the number of portfolios relatively similar. This leads to the creation of 25 composite cross-sections of test assets listed in Table OA24. To each of these composite cross-section, we assign a "revealed preference" weight equal to the product of its base assets' empirical frequencies.

Finally, using the continuous spike-and-slab approach of Section II.2.3 and 2.25 quadrillion models, we estimate posterior probabilities of factors and posterior mean of their risk premia on each of the composite cross-sections (that is, we jointly evaluate the equivalent of 3.78 quintillion regressions) and then average them out using the revealed preferences weights of the cross-sections.

Table 10 presents our results. When aggregating across the cross-sections of most widely used test assets, only HML and the hedged version of the market factor, MKT* of Daniel, Mota, Rottke, and Santos (2020), seem to be robust explanators of the cross-section of asset returns. Indeed, their posterior factor probabilities are largely above the prior, set at 50%, and the monthly risk premia are roughly equal to their sample returns scaled by the factors' posterior probabilities – indicating that these two factors price themselves accurately. Short-term reversal, the sentiment measure of Baker and Wurgler (2006), and a short-term

[30]See Online Appendix OA.C.4 for a detailed description of the selection procedure.

Table 10: Posterior factor probabilities, $\mathbb{E}[\gamma_j|\text{data}]$, and risk premia in 2.25 quadrillion models averaged across 25 cross-sections.

	$\mathbb{E}[\gamma_j\|\text{data}]$						$\mathbb{E}[\lambda_j\|\text{data}]$						
	ψ:						ψ:						
Factors:	1	5	10	20	50	100	1	5	10	20	50	100	$\bar{F}$
HML	76.44%	80.52%	79.93%	78.33%	75.92%	72.51%	0.130	0.204	0.218	0.224	0.226	0.221	0.377
MKT*	67.50%	69.97%	67.92%	65.50%	58.05%	51.27%	0.092	0.187	0.222	0.247	0.245	0.226	0.514
STRev	51.82%	52.61%	53.69%	55.36%	55.04%	52.57%	0.008	0.031	0.052	0.083	0.133	0.164	0.438
UMD	52.66%	52.73%	51.53%	49.04%	43.26%	37.21%	0.026	0.066	0.086	0.103	0.117	0.116	0.646
BEH_PEAD	50.93%	50.95%	51.42%	52.53%	52.87%	51.21%	0.004	0.012	0.020	0.033	0.058	0.078	0.619
ACCR	49.42%	50.18%	51.15%	49.97%	46.10%	41.42%	-0.010	-0.025	-0.036	-0.046	-0.060	-0.068	0.343
SMB*	57.64%	54.77%	51.10%	46.09%	40.80%	36.37%	0.040	0.068	0.073	0.071	0.073	0.074	0.215
DIV	49.41%	49.11%	50.78%	50.11%	49.00%	49.10%	0.000	0.000	0.000	0.000	0.000	0.000	0.933*
BW_ISENT	49.19%	50.32%	50.35%	50.83%	49.77%	49.19%	0.000	0.001	0.002	0.004	0.007	0.011	0.055*
CMA*	49.95%	50.26%	50.27%	48.73%	46.94%	43.51%	0.000	0.000	-0.001	-0.003	-0.005	-0.006	0.242
LIQ_TR	49.39%	49.92%	50.16%	50.92%	50.64%	49.78%	0.000	0.004	0.010	0.022	0.051	0.081	0.438
REAL_UNC	49.56%	49.06%	50.05%	48.51%	48.26%	45.37%	0.000	0.000	0.000	0.000	0.000	0.000	0.046*
PE	48.82%	50.22%	49.97%	50.56%	50.05%	50.71%	-0.001	-0.004	-0.007	-0.013	-0.023	-0.037	6.826*
NONDUR	49.06%	49.12%	49.75%	50.04%	51.62%	51.30%	0.001	0.002	0.004	0.006	0.013	0.022	0.149*
LIQ_NT	48.78%	48.98%	49.63%	48.71%	47.99%	45.78%	0.002	0.009	0.017	0.029	0.046	0.049	0.471*
COMP_ISSUE	51.28%	52.22%	49.59%	46.81%	39.52%	33.33%	0.032	0.073	0.086	0.094	0.092	0.086	0.497
UNRATE	48.54%	49.35%	49.40%	49.43%	48.80%	46.78%	0.000	-0.001	-0.001	-0.002	-0.004	-0.004	1.171*
DeltaSLOPE	49.26%	49.21%	49.29%	49.55%	50.22%	50.56%	0.000	-0.001	-0.001	-0.002	-0.004	-0.007	0.032*
IPGrowth	50.22%	50.22%	49.21%	50.15%	49.51%	48.99%	0.000	-0.001	-0.001	-0.001	-0.002	-0.004	0.109*
SERV	48.93%	49.98%	49.18%	48.97%	49.69%	50.37%	0.000	0.000	0.000	0.000	-0.001	-0.001	0.047*
RMW*	49.73%	49.70%	49.14%	47.19%	41.27%	36.21%	0.003	0.011	0.016	0.021	0.025	0.027	0.219
TERM	48.73%	48.44%	49.08%	49.78%	50.80%	49.75%	0.001	0.002	0.005	0.008	0.017	0.026	0.929*
MKT	46.43%	49.07%	49.04%	47.41%	42.18%	36.22%	0.038	0.105	0.136	0.158	0.160	0.144	0.563
Oil	48.81%	49.08%	48.92%	49.38%	49.02%	48.82%	0.001	0.008	0.014	0.026	0.052	0.082	0.613*
FIN_UNC	47.89%	48.44%	48.78%	48.89%	49.45%	46.61%	0.000	0.000	0.000	0.000	-0.001	-0.001	0.104*
HJTZ_ISENT	48.31%	49.13%	48.65%	48.31%	49.73%	47.28%	0.000	0.000	-0.001	-0.001	-0.001	-0.001	0.187*
DEFAULT	49.21%	48.85%	48.53%	49.09%	48.31%	47.73%	0.000	0.000	-0.001	-0.001	-0.001	-0.002	0.337*
NetOA	49.51%	49.29%	48.51%	46.53%	41.98%	37.93%	0.007	0.019	0.027	0.035	0.040	0.040	0.544
HML*	49.15%	48.97%	48.47%	46.43%	41.87%	37.71%	0.000	0.003	0.006	0.010	0.012	0.014	0.251
INV_IN_ASS	49.22%	48.00%	48.30%	49.02%	46.14%	42.26%	-0.001	-0.003	-0.005	-0.006	-0.007	-0.008	0.549
MACRO_UNC	48.14%	47.67%	47.72%	47.02%	44.89%	41.10%	0.000	0.000	0.000	0.000	0.000	0.000	0.079*
ROA	50.85%	49.91%	47.72%	43.98%	34.79%	28.08%	0.019	0.059	0.078	0.086	0.072	0.055	0.551
CMA	54.80%	51.54%	47.60%	41.84%	34.25%	27.49%	0.020	0.040	0.044	0.043	0.040	0.033	0.352
LTRev	48.51%	48.65%	47.56%	45.08%	39.85%	34.22%	-0.007	-0.021	-0.027	-0.033	-0.034	-0.032	0.252
STOCK_ISS	49.04%	48.97%	47.21%	43.10%	37.77%	31.27%	-0.022	-0.049	-0.058	-0.062	-0.065	-0.060	0.515
ASS_Growth	49.32%	48.03%	46.92%	44.79%	38.16%	32.24%	0.006	0.019	0.027	0.032	0.033	0.031	0.525
BAB	48.29%	47.95%	46.69%	43.26%	37.65%	32.47%	-0.018	-0.039	-0.045	-0.045	-0.038	-0.030	0.921
INTERM_CAP_RATIO	50.89%	46.87%	43.52%	38.78%	31.68%	25.68%	-0.022	0.013	0.029	0.042	0.050	0.044	0.756*
IA	51.27%	47.44%	42.99%	39.14%	31.13%	26.00%	-0.009	-0.012	-0.010	-0.010	-0.009	-0.010	0.409
DISSTR	50.23%	46.43%	42.97%	37.00%	30.13%	24.49%	0.064	0.092	0.094	0.088	0.078	0.069	0.475
O_SCORE	47.33%	44.47%	40.64%	36.89%	29.23%	23.52%	-0.012	-0.019	-0.020	-0.020	-0.015	-0.010	0.020
SMB	43.72%	41.54%	39.74%	36.44%	29.82%	23.45%	0.021	0.036	0.041	0.042	0.038	0.031	0.257
ROE	47.18%	43.32%	39.19%	34.14%	27.31%	21.84%	0.004	0.010	0.014	0.016	0.015	0.013	0.555
PERF	51.69%	44.62%	38.96%	34.21%	27.05%	22.11%	-0.041	-0.053	-0.049	-0.045	-0.039	-0.036	0.651
MGMT	49.13%	42.81%	38.78%	33.65%	26.03%	20.70%	0.019	0.036	0.039	0.039	0.034	0.029	0.631
GR_PROF	47.85%	41.41%	37.31%	32.07%	25.35%	19.74%	0.023	0.032	0.035	0.035	0.031	0.026	0.199
BEH_FIN	47.84%	40.62%	36.23%	31.04%	23.24%	18.48%	-0.001	0.010	0.014	0.017	0.015	0.013	0.760
SKEW	46.66%	40.57%	36.12%	30.93%	24.75%	19.08%	0.000	0.000	0.000	0.000	0.000	0.000	0.004
RMW	45.71%	38.48%	34.48%	29.51%	22.36%	18.05%	0.006	0.012	0.015	0.016	0.014	0.013	0.292
QMJ	46.70%	37.54%	32.16%	26.92%	20.00%	16.01%	0.012	0.009	0.007	0.006	0.005	0.006	0.405
HML_DEVIL	44.46%	36.28%	32.04%	27.27%	20.88%	16.48%	0.018	0.029	0.029	0.026	0.021	0.017	0.356

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, and posterior mean of factor risk premia, $\mathbb{E}[\lambda_j|\text{data}]$ are computed using the continuous spike-and-slab approach of section II.2.3 and 51 factors. The prior for each factor inclusion is a $Beta(1,1)$, yielding a prior expectation for γ_j equal to 50%. The last column reports sample average returns for the tradable factors, and the numbers denoted with the asterisk correspond to the return on the projection-based factor-mimicking portfolio of the non-tradable factor. The data is monthly, 1973:10 to 2016:12. Posterior probabilities and risk premia are found as the weighted average of the output from the estimation on 25 cross-sections. Cross-sections and weights are summarized in Table OA24 of the Appendix. The 51 factors considered are described in Table A1 of the Appendix and are sorted by the posterior factor probability for $\psi = 10$.

behavioral factor of Daniel, Hirshleifer, and Sun (2019) would be the next candidates, but their posterior factor probability is almost exactly equal to the prior, and the risk premium is very close to zero, especially for a rather plausible prior on the Sharpe ratio, corresponding to $\psi \in [10, 20]$. The posterior probability of SMB^* is larger than its prior probability only for $\psi \leq 10$. Overall, we find that averaging the findings over a large set of popular cross-sections leads to similar results to those described in Section IV.2.[31]

IV.7 Out-of-sample analysis

In the last several years there has been a growing push for the empirical analysis to shift its focus away from the traditional in-sample estimation to examining the out-of-sample evaluation of both existing and new results.[32] In this subsection we investigate the out-of-sample (OOS) performance of our approach, and, in particular, the role of two tuning parameters we have been using throughout the paper: the level of average shrinkage applied to all the factors at the cross-sectional stage (ψ) and its heterogeneity designed to separate strong and weak factors (r).

We focus on the sample of 55 test assets (25 portfolios, sorted by size and value, and 30 industry portfolios) and 51 factors used in the main analysis. In the spirit of cross-validation, we then split all the data into two equal time series subsamples, repeat the estimation of Section IV.2 in each subsample, and save the posterior mean of factor risk premia obtained therein. We then measure the root mean squared error (RMSE) of cross-sectional pricing obtained applying the parameter estimates to the half of the data not used in the estimation. For example, whenever subsample (1) is used to estimate the posterior mean of the factor risk premia $\left(\hat{\lambda}_c^{(1)}, \hat{\boldsymbol{\lambda}}_{\boldsymbol{f}}^{(1)}\right)^\top$, the cross-sectional out-of-sample RMSE is

$$RMSE_{OOS}^{(1)\rightarrow(2)} = \sqrt{\frac{1}{N}\left(\bar{\boldsymbol{R}}^{\boldsymbol{(2)}} - \boldsymbol{1_N}\hat{\lambda}_c^{(1)} - \boldsymbol{\beta}_{\boldsymbol{f}}^{\boldsymbol{(2)}}\hat{\boldsymbol{\lambda}}_{\boldsymbol{f}}^{\boldsymbol{(1)}}\right)^\top \left(\bar{\boldsymbol{R}}^{\boldsymbol{(2)}} - \boldsymbol{1_N}\hat{\lambda}_c^{(1)} - \boldsymbol{\beta}_{\boldsymbol{f}}^{\boldsymbol{(2)}}\hat{\boldsymbol{\lambda}}_{\boldsymbol{f}}^{(1)}\right)}, \quad (27)$$

where moments with superscript (2) represent the OOS moments estimated from a withheld sample and N is the number of test assets. Note that since we focus on the cross-sectional asset pricing and risk premia uncertainty, we use the factor loadings for the validation data directly from the second subsample. Similarly, we fit the model to the second half of the data and examine its out-of-sample pricing errors on the first half construct-

[31] In Online Appendix OA.C.4 we report similar estimation output for alternative weighting schemes (simple averaging across all of cross-sections, and weighting proportional to the Sharpe ratios achievable using the test assets) and estimation results for each of the individual cross-sections are available upon request.

[32] See, e.g., McLean and Pontiff (2016), Kozak, Nagel, and Santosh (2019), Martin and Nagel (2019), and Hodrick and Tomunen (2020), among others.

ing $RMSE_{OOS}^{(2)\to(1)}$. We then compute the average of these pricing errors:[33] $RMSE_{OOS} = 1/2\left(RMSE_{OOS}^{(1)\to(2)} + RMSE_{OOS}^{(2)\to(1)}\right)$. Note that since we work with standardized factors, the units of $RMSE_{OOS}$ are Sharpe ratios and can be directly compared to the Sharpe ratio of the test assets.

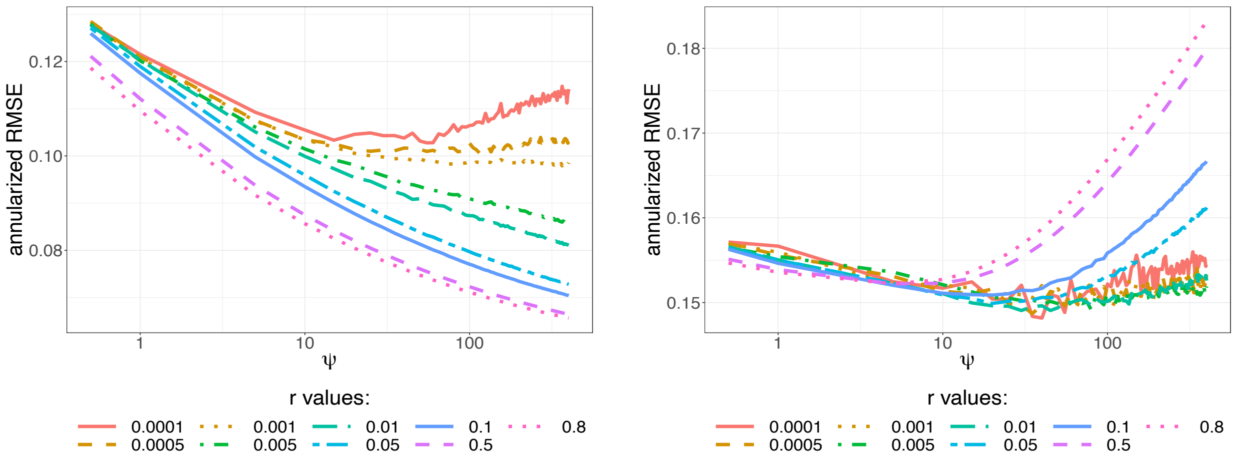
We explore how the out-of-sample performance of our method varies across different values of r and ψ. Intuitively, a larger level of ψ is equivalent to and overall lower level of shrinkage applied to the factor risk premia, which should increase the in-sample model goodness-of-fit, but not necessarily the OOS one. Similarly, r controls the degree of selective (heterogeneous) factor shrinkage: when r is close to 1 (e.g., 0.7-0.8 and higher), it imposes almost identical shrinkage on the risk premia of both useful and useless factors. In contrast, when r is extremely small (e.g., 0.0001), it imposes a much stronger separation between the treatment of the selected and residual factors ($\gamma_k = 1$ and $\gamma_k = 0$, respectively). This heterogeneity could have a non-trivial effect on the model performance. Hence, we examine the impact of both types of shrinkage on both the in-sample and out-of-sample pricing errors for a wide range of tuning parameters.[34] As with in the other empirical applications in the paper, we investigate the cases, in which the prior variance of risk premia is proportional to the demeaned correlation between test assets and factors[35]: $\psi_k = \psi \times \tilde{\boldsymbol{\rho}}_{\boldsymbol{k}}^\top \tilde{\boldsymbol{\rho}}_{\boldsymbol{k}}$ for each factor k.

Figure 8 summarizes our findings by reporting the annualized in- and out-of-sample RMSEs. Not surprisingly, the less shrinkage is applied to all the factors (i.e., when ψ is relatively high), and the less is the selective shrinkage among them (i.e., when r is close to 1), the better is the model fit in-sample. This is due to both an automatic effect of including a larger number of variables in the model without imposing any discipline on their coefficients (i.e., standard overfitting) and as consequence of the identification failure, when including spurious factors seems to improve model fit. However, out of sample the situation is substantially different, as moderate-to-high levels of shrinkage in both dimensions clearly deliver a better performance. That is, Figure 8 makes clear that the increase in in-sample model fit generated by removing the shrinkage of useless factor (by setting both ψ and r

[33]Note that is usually standard in the time series analysis to do a five– or ten–fold cross-validation. However, we are limited by the data availability and high level of estimation uncertainty, both of which are characteristic features of the cross-sectional asset pricing. Hence, our main results are based only on two-fold cross-validation.

[34]$r \in \{0.0001, 0.0005, 0.001, 0.05, 0.01, 0.05, 0.1, 0.5, 0.8\}$ and $\psi \in \{0.1, 0.5, 1, 5, 10, ..., 400\}$.

[35]Empirical results for additional parametrizations, e.g., using non-demeaned correlations, or excluding the common intercept during the cross-section estimation stage, are presented in the Online Appendix OA.C.5. Furthermore, we have done the same out-of-sample exercise using standard off-the-shelf lasso and ridge penalties applied to the second-pass regressions. Compared to the Bayesian estimation with our spike-and-slab approach, we found their performance to be subpar. This is not surprising since these methods were not designed to be robust to identification failures in factor selection. These additional results are available upon request.



(a) In-sample pricing errors **(b)** Out-of-sample pricing errors

Figure 8: In-sample and out-of-sample cross-sectional root mean squared errors.

Cross-sectional model fit RMSE obtained with a two-fold cross-validation for different values of shrinkage parameters, r and ψ. Penalty, given in equation (21), based on demeaned correlations between factor and returns. Cross-sectional equation includes a common intercept. Estimation based on the 55 portfolios and 51 factors BMA used in Section IV.2.

to large values), is illusory in that it comes at the cost of a substantial worsening of the out-of-sample performance.

Note that while a certain degree of shrinkage (both average and heterogenous) seems to be important for the model to reliably match the data, its out-of-sample performance is remarkably stable for a wide range of parameters: For example, the average mean squared error stays stable at around 0.15 for a large set of relatively small values of both r and ψ. Importantly, the whole range of plausible tuning parameter values that correspond to reasonable economic priors of the agent, as discussed in Section II.2.3, lies within this region. Hence, the out-of-sample performance of our model-averaging approach is remarkably stable across all the realistic choices of the shrinkage parameters and is not substantially affected by their particular values. Moreover, for such reasonable values of the prior parameters, both in- and out-of-sample RMSE are quite small relative to both the ex post maximum Sharpe ratio obtainable with the tests assets (about 2.3) and the Sharpe ratios of the individual assets (that have an interquartile range of 0.35–0.51).

V Extensions

In addition to the extensions formalized in remarks 1 (on how to handle generated factors such as principal components and factor mimicking portfolios) and 3 (on how to handle the identification failure generated by "level factors"), our method can be feasibly extended to encompass several salient generalizations, including estimating linear SDF models directly.

First, based on economic considerations, one might possibly want to bound the maximum risk premia (or the maximum Sharpe ratios) associated with the factors. This can be achieved by replacing the Gaussian distributions in our spike-and-slab priors with (rescaled and centered) Beta distributions, since the latter have bounded support. Furthermore, for the sake of expositional simplicity and closed-form solutions, we have focused on regularizing spike-and-slab priors with exponential tails. Nevertheless, our approach, which shrinks useless factors based on their correlation with asset returns, could be also implemented using polynomial tailed (i.e., heavy-tailed) mixing priors (see Polson and Scott (2011) for a general discussion of priors for regularization and shrinkage).[36] The rationale for using heavy-tailed priors is that, when the likelihood has thick tails while the prior has a thin tail, if the likelihood peak moves too far from the prior mean, the posterior eventually reverts toward the prior. Nevertheless, note that this mechanism (first pointed out in Jeffreys (1961)) is actually desirable in our settings in order to shrink the risk premia of useless factors toward zero.[37]

Second, Lewellen, Nagel, and Shanken (2010) point out that the first pass time series regression is often affected by having a strong factor structure in the residuals. Given the hierarchical structure of our Bayesian approach, one can add latent linear components in the time series regression of asset returns on factors, reformulate the time series estimation step as a state-space problem, and filter the latent components (e.g., via the Kalman filter). The posterior sampling of the time series parameters would then be enriched by the drawing of the added terms as in Bryzgalova and Julliard (2018). Furthermore, one could allow the latent time series factors to be potential priced in the cross-section (again as in Bryzgalova and Julliard (2018)). This extension would increase the numerical complexity of the procedure in the time series step, but would nonetheless leave unchanged the method proposed in this paper for the cross-sectional step (with the only difference, that the time series loadings of the latent factors could be included in the cross-sectional step as if these latent factors were observable). This extended approach would lead to valid posterior inference and model selection.

Third, again thanks to the hierarchical structure of our method, time-varying time series betas could be accommodated by adopting the time varying parameters approach of Primiceri (2005) in the time series step. And since in our approach the asset-specific expected risk premia are parameters estimated in the time series step, this extension would also allow

[36]For example, albeit alternative distributions with desirable properties exist, our spike-and-slab could be implemented using a Cauchy prior with location parameter set to zero and scale parameter proportional to ψ_j, as defined in equation (19) or (21).

[37]Risk premia and Sharpe ratios have a natural support, hence, as discussed in Section II.2.3 the prior can be chosen (adjusting the parameter ψ) to attach high probability to it. And since useless factors will tend to generate heavy-tailed likelihoods (in the limit, the likelihood is an improper "uniform" on $\mathbb{R}$), with peaks for risk premia that deviate toward infinity, the posterior risk premia of such factors will be shrunk toward the prior mean if the prior has thin tails.

for time variation in the risk premia of the test assets. Furthermore, albeit this would significantly increase the numerical complexity of the cross-sectional inference step, the time varying parameters formulation could also be used for the modeling of the factor risk premia in the cross-sectional step.

VI Conclusions

We have developed a novel (Bayesian) method for the analysis of linear factor models in asset pricing. The approach can handle the quadrillions of models generated by the zoo of traded and non-traded factors, and it delivers inference that is robust to the common identification failures, and spurious inference problems, caused by useless and level factors.

We have applied our approach to the study of more than two quadrillion factor model specifications and have found that: 1) only a handful of factors (the Fama and French (1992) "high-minus-low" proxy for the value premium, and the adjusted versions of both market and size factors of Daniel, Mota, Rottke, and Santos (2020)) seem to be robust explanators of the cross-sections of asset returns; 2) jointly, the three robust factors provide a model that is, compared to the previous empirical literature, one order of magnitude more likely to have generated the observed asset returns (its posterior probability is about 85–88%); 3) nevertheless, with very high probability the "true" latent SDF is dense in the space of factors proposed in the previous literature, that is, capturing its characteristics requires the use of 24–25 factors (at the posterior mean of the SDF sparsity); and 4) however, despite being dense in the space of factors, the SDF-implied maximum Sharpe ratio is not excessive, suggesting a high degree of commonality, in terms of captured risks, among the factors in the zoo.

As a by-product of our novel framework for empirical asset pricing, we provide a very simple Bayesian version of the Fama and MacBeth (1973) regression method (BFM). We show that this simple procedure (which requires neither optimization nor tuning parameters and is not harder to implement than, e.g., the Shanken (1992) correction for standard errors), makes useless factors easily detectable in finite sample. In extensive simulations, the BFM and its GLS analogue (BFM-GLS) perform well even with relatively small time and large cross-sectional dimensions. We apply BFM and BFM-GLS to several notable factor models, and document that a range of trade and non-traded factors are only weakly identified at best and are characterized by a substantial degree of model misspecification and uncertainty.

Finally, thanks to its hierarchical structure, our framework is extremely flexible and can be extended to accommodate, and deliver robust inference in the presence of, 1) pre-estimated factors (e.g., mimicking portfolios and principal components), 2) latent, priced and unpriced, factors, and 3) time varying betas and risk premia.

References

Allena, R. (2019): "Comparing Asset Pricing Models with Traded and Non-Traded Factors," working paper.

Anderson, E., and A.-R. Cheng (2016): "Robust Bayesian Portfolio Choices," *Review of Financial Studies*, 29, 1330–1375.

Avramov, D., and G. Zhou (2010): "Bayesian Portfolio Analysis," *Annual Review of Financial Economics*, 2, 25–47.

Baker, M., and J. Wurgler (2006): "Investor Sentiment and the Cross-Section of Stock Returns," *Journal of Finance*, 61, 1645–80.

Baks, K. P., A. Metrick, and J. Wachter (2001): "Should Investors Avoid All Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation," *Journal of Finance*, 56, 45–85.

Barillas, F., and J. Shanken (2018): "Comparing Asset Pricing Models," *Journal of Finance*, 73(2), 715–754.

Bartlett, M. S. (1957): "Comment on a Statistical Paradox by D. V. Lindley," *Biometrika*, 44(1-2), 533–534.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on Treatment Effects after Selection among High-Dimensional Controls," *Review of Economic Studies*, 81, 608–650.

Bryzgalova, S. (2015): "Spurious Factors in Linear Asset Pricing Models," working paper.

Bryzgalova, S., and C. Julliard (2018): "Consumption in Asset Returns," London School of Economics Manuscript.

Chernov, M., L. Lochstoer, and S. Lundeby (2019): "Conditional Dynamics and the Multi-Horizon Risk-Return Trade-Off," working paper.

Chib, S., X. Zeng, and L. Zhao (2020): "On Comparing Asset Pricing Models," *Journal of Finance*.

Cochrane, J. H. (2005): *Asset Pricing: Revised Edition*. Princeton University Press.

Cremers, K. M. (2002): "Stock Return Predictability: A Bayesian Model Selection Perspective," *Review of Financial Studies*, 15(4), 1223–1249.

Daniel, K., D. Hirshleifer, and L. Sun (2019): "Short- and Long-Horizon Behavioral Factors," *Review of Financial Studies*.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020): "The Cross-Section of Risk and Return," *Review of Financial Studies*, forthcoming.

Fabozzi, F., D. Huang, and G. Zhou (2010): "Robust Portfolios: Contributions from Operations Research and Finance," *Annals of Operations Research*, 172, 191–220.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465.

Fama, E. F., and K. R. French (2016): "Dissecting Anomalies with a Five-Factor Model," *Review of Financial Studies*, 29, 69–103.

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

Garlappi, L., R. Uppal, and T. Wang (2007): "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *Review of Financial Studies*, 20, 41–81.

George, E. I., and R. E. McCulloch (1993): "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.

——— (1997): "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–373.

Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–73.

Ghosh, A., C. Julliard, and A. Taylor (2016): "What is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models," *Review of Financial Studies*, 30, 442–504.

——— (2018): "An Information-Theoretic Asset Pricing Model," London School of Economics Manuscript.

Giannone, D., M. Lenza, and G. E. Primiceri (2018): "Economic Predictions with Big Data: The Illusion of Sparsity," FRB of New York Staff Report No. 847.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989): "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57(5), 1121–1152.

Giglio, S., G. Feng, and D. Xiu (2019): "Taming the Factor Zoo," *Journal of Finance*, forthcoming.

Giglio, S., and D. Xiu (2018): "Asset Pricing with Omitted Factors," working paper.

Gospodinov, N., R. Kan, and C. Robotti (2014): "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors," *Review of Financial Studies*, 27(7), 2139–2170.

——— (2019): "Too Good to Be True? Fallacies in Evaluating Risk Factor Models," *Journal of Financial Economics*, 132(2), 451–471.

Goyal, A., Z. L. He, and S.-W. Huh (2018): "Distance-Based Metrics: A Bayesian Solution to the Power and Extreme-Error Problems in Asset-Pricing Tests," Swiss Finance Institute Research Paper No. 18-78, available at SSRN: https://ssrn.com/abstract=3286327.

Harvey, C., and Y. Liu (2019): "Cross-Sectional Alpha Dispersion and Performance Evaluation," *Journal of Financial Economics*, forthcoming.

Harvey, C., and G. Zhou (1990): "Bayesian Inference in Asset Pricing Tests," *Journal of Financial Economics*, 26, 221–254.

Harvey, C. R. (2017): "Presidential Address: The Scientific Outlook in Financial Economics," *Journal of Finance*, 72(4), 1399–1440.

Harvey, C. R., Y. Liu, and H. Zhu (2016): "... and the Cross-Section of Expected Returns," *Review of Financial Studies*, 29(1), 5–68.

He, A., D. Huang, and G. Zhou (2018): "New Factors Wanted: Evidence from a Simple Specification Test," working paper, available at SSRN: https://ssrn.com/abstract=3143752.

Hodrick, R. J., and T. Tomunen (2020): "Taking the Cochrane-Piazzesi Term Structure Model Out of Sample: More Data, Additional Currencies, and FX Implications," *Critical Finance Review*, forthcoming.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14(4), 382–401.

Huang, D., J. Li, and G. Zhou (2018): "Shrinking Factor Dimension: A Reduced-Rank Approach," working paper, available at SSRN: https://ssrn.com/abstract=3205697.

Ishwaran, H., J. S. Rao, et al. (2005): "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies," *Annals of Statistics*, 33(2), 730–773.

Jagannathan, R., and Z. Wang (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51(1), 3–53.

Jeffreys, H. (1961): *Theory of Probability*. Oxford: Calendron Press.

Jones, C., and J. Shanken (2005): "Mutual Fund Performance with Learning across Funds," *Journal of Financial Economics*, 78, 507–552.

Kan, R., and C. Zhang (1999a): "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54(1), 103–127.

——— (1999b): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54(1), 203–235.

Kass, R. E., and A. E. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90(430), 773–795.

Kelly, B., S. Pruitt, and Y. Su (2019): "Characteristics Are Covariances: A Unified Model of Risk and Return," *Journal of Financial Economics*, 134, 501–524.

Kleibergen, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149(2), 149–173.

Kleibergen, F., and Z. Zhan (2015): "Unexplained Factors and Their Effects on Second Pass R-

Squareds," *Journal of Econometrics*, 189(1), 101–116.

Kozak, S., S. Nagel, and S. Santosh (2019): "Shrinking the Cross-Section," *Journal of Financial Economics*, forthcoming.

Lancaster, T. (2004): *Introduction to Modern Bayesian Econometrics*. Wiley.

Lettau, M., and S. Ludvigson (2001): "Resurrecting the (C) CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109(6), 1238–1287.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A Skeptical Appraisal of Asset Pricing Tests," *Journal of Financial Economics*, 96(2), 175–194.

Ludvigson, S., S. Ma, and S. Ng (2019): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?," *American Economic Journal: Macroeconomics*, forthcoming.

Madigan, D., and A. E. Raftery (1994): "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window," *Journal of the American Statistical Association*, 89(428), 1535–1546.

Martin, I., and S. Nagel (2019): "Market Efficiency in the Age of Big Data," NBER Working Paper 26586.

McLean, R. D., and J. Pontiff (2016): "Does Academic Research Destroy Stock Return Predictability?," *Journal of Finance*, 71(1), 5–32.

Pástor, L. (2000): "Portfolio Selection and Asset Pricing Models," *The Journal of Finance*, 55(1), 179–223.

Pástor, L., and R. F. Stambaugh (2000): "Comparing Asset Pricing Models: an Investment Perspective," *Journal of Financial Economics*, 56(3), 335–381.

Pástor, L., and R. F. Stambaugh (2002): "Investing in Equity Mutual Funds," *Journal of Financial Economics*, 63, 351–380.

Phillips, D. L. (1962): "A Technique for the Numerical Solution of Certain Integral Equations of the First Kind," *J. ACM*, 9(1), 84–97.

Polson, N. G., and J. G. Scott (2011): "Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction," in *Bayesian Statistics 9*, ed. by J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West. Oxford University Press.

Primiceri, G. E. (2005): "Time-Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997): "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92(437), 179–191.

Roberts, H. V. (1965): "Probabilistic Prediction," *Journal of the American Statistical Association*, 60(309), 50–62.

Shanken, J. (1987): "A Bayesian Approach to Testing Portfolio Efficiency," *Journal of Financial Economics*, 19, 195–215.

——— (1992): "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5, 1–33.

Sims, C. A. (2007): "Thinking about Instrumental Variables," mimeo.

Stock, J. H. (1991): "Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series," *Journal of Monetary Economics*, 28(3), 435–459.

Tikhonov, A., A. Goncharsky, V. Stepanov, and A. G. Yagola (1995): *Numerical Methods for the Solution of Ill-Posed Problems*, Mathematics and Its Applications. Springer Netherlands.

Uppal, R., P. Zaffaroni, and I. Zviadadze (2018): "Correcting Misspecified Stochastic Discount Factors," working paper.

Yogo, M. (2006): "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance*, 61(2), 539–580.

A Appendix

A.1 Additional derivations and proofs

A.1.1 Derivation of the posterior distribution in Section II.1

Let's consider first the time series regression. We assume that $\boldsymbol{\epsilon_t} \overset{\text{iid}}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma})$, or $\boldsymbol{\epsilon} \sim \mathcal{MVN}(\boldsymbol{0_{T\times N}}, \boldsymbol{\Sigma} \otimes \boldsymbol{I_T})$. The time series likelihood of the data $(\boldsymbol{R}, \boldsymbol{F})$ is therefore

$$p(data|\boldsymbol{B}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{NT}{2}} |\boldsymbol{\Sigma}|^{-\frac{T}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(\boldsymbol{R}-\boldsymbol{F}\boldsymbol{B})^\top(\boldsymbol{R}-\boldsymbol{F}\boldsymbol{B})]}.$$

After assigning flat prior for $(\boldsymbol{B}, \boldsymbol{\Sigma})$, $\pi(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{N+1}{2}}$, we simplify the likelihood function by exploiting the fact that OLS estimated residuals are orthogonal to the regressors.[38] Therefore, the posterior distribution in the first (time series) step is

$$p(\boldsymbol{B}, \boldsymbol{\Sigma}|data) \propto |\boldsymbol{\Sigma}|^{-\frac{T+N+1}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(T\hat{\boldsymbol{\Sigma}})]} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}-\hat{\boldsymbol{B}}_{\boldsymbol{ols}})^\top \boldsymbol{F}^\top \boldsymbol{F}(\boldsymbol{B}-\hat{\boldsymbol{B}}_{\boldsymbol{ols}})]}.$$

where $\hat{\boldsymbol{B}}_{\boldsymbol{ols}} = (\hat{\boldsymbol{a}}, \hat{\boldsymbol{\beta}})\top = (\boldsymbol{F}^\top\boldsymbol{F})^{-1}\boldsymbol{F}^\top\boldsymbol{R}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{ols}} = \frac{1}{T}(\boldsymbol{R} - \boldsymbol{F}\hat{\boldsymbol{B}}_{\boldsymbol{ols}})^\top(\boldsymbol{R} - \boldsymbol{F}\hat{\boldsymbol{B}}_{\boldsymbol{ols}})$. Hence, the posterior distribution of $\boldsymbol{B}$ conditional on data and $\boldsymbol{\Sigma}$ is

$$p(\boldsymbol{B}|\boldsymbol{\Sigma}, data) \propto e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}-\hat{\boldsymbol{B}}_{\boldsymbol{ols}})^\top \boldsymbol{F}^\top \boldsymbol{F}(\boldsymbol{B}-\hat{\boldsymbol{B}}_{\boldsymbol{ols}})]},$$

and the above is the kernel of the multivariate normal in equation (8).

If we further integrate out $\boldsymbol{B}$, it is easy to show that

$$p(\boldsymbol{\Sigma}|data) \propto |\boldsymbol{\Sigma}|^{-\frac{T+N-K}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(T\hat{\boldsymbol{\Sigma}})]}.$$

Therefore, the posterior distribution of $\boldsymbol{\Sigma}$ is the inverse-Wishart in equation (9).

Recall that $\boldsymbol{\beta} = (\boldsymbol{1_N},\, \boldsymbol{\beta_f})$, $\boldsymbol{\lambda}^\top = (\lambda_c,\, \boldsymbol{\lambda_f}^\top)$. If we assume that the pricing error α_i follows an independent and identical normal distribution $\mathcal{N}(0, \sigma^2)$, the cross-sectional likelihood function in the second step is

$$p(data|\boldsymbol{\lambda}, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{a}-\boldsymbol{\beta}\boldsymbol{\lambda})^\top(\boldsymbol{a}-\boldsymbol{\beta}\boldsymbol{\lambda})}, \tag{28}$$

where data in the second step include $(\boldsymbol{a},\, \boldsymbol{\beta_f})$ drawn from the first step. Assuming the diffuse prior $\pi(\boldsymbol{\lambda},\, \sigma^2) \propto \frac{1}{\sigma^2}$ the posterior distribution of $(\boldsymbol{\lambda},\, \sigma^2)$ is

$$p(\boldsymbol{\lambda}, \sigma^2|data, \boldsymbol{B}, \boldsymbol{\Sigma}) \propto (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{a}-\boldsymbol{\beta}\boldsymbol{\lambda})^\top(\boldsymbol{a}-\boldsymbol{\beta}\boldsymbol{\lambda})} = (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{a}-\boldsymbol{\beta}\hat{\boldsymbol{\lambda}}+\boldsymbol{\beta}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}))^\top(\boldsymbol{a}-\boldsymbol{\beta}\hat{\boldsymbol{\lambda}}+\boldsymbol{\beta}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}))}$$

[38] Note that $(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{B})^\top(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{B}) = T\hat{\boldsymbol{\Sigma}} + (\boldsymbol{B} - \hat{\boldsymbol{B}}_{\boldsymbol{ols}})^\top\boldsymbol{F}^\top\boldsymbol{F}(\boldsymbol{B} - \hat{\boldsymbol{B}}_{\boldsymbol{ols}})$.

$$\therefore \;\; p(\boldsymbol{\lambda}|\sigma^2, data, \boldsymbol{B}, \boldsymbol{\Sigma}) \propto e^{-\frac{(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^\top \boldsymbol{\beta}^\top \boldsymbol{\beta}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})}{2\sigma^2}},$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}$ and $\hat{\sigma}^2 = \frac{(\boldsymbol{a}-\boldsymbol{\beta}\hat{\boldsymbol{\lambda}})^\top(\boldsymbol{a}-\boldsymbol{\beta}\hat{\boldsymbol{\lambda}})}{N}$. Therefore, the posterior conditional distribution of $\boldsymbol{\lambda}$ is the one in equation (12). Finally, we can derive the posterior distribution of σ^2 by integrating out $\boldsymbol{\lambda}$ as follows:

$$p(\sigma^2|data, \boldsymbol{B}, \boldsymbol{\Sigma}) = \int p(\boldsymbol{\lambda}, \sigma^2|data, \boldsymbol{B}, \boldsymbol{\Sigma}) d\sigma^2 \propto (\sigma^2)^{-\frac{N-K+1}{2}} e^{-\frac{N\hat{\sigma}^2}{2\sigma^2}},$$

hence obtaining the posterior distribution in equation (13).

A.1.2 Formal derivation of the flat prior pitfall for risk premia

Following the derivation in section A.1.1, the likelihood function in the second step is given in equation (28). Assigning flat prior to the parameters[39] $(\boldsymbol{\lambda}, \sigma^2)$, the marginal cross-sectional likelihood function conditional on model index $\boldsymbol{\gamma}$ is

$$\begin{aligned} p(data|\boldsymbol{\gamma}) &= \iint p(data|\boldsymbol{\gamma}, \boldsymbol{\lambda}, \sigma^2)\pi(\boldsymbol{\lambda}, \sigma^2|\boldsymbol{\gamma}) d\lambda d\sigma^2 \propto \iint (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{a}-\boldsymbol{\beta_\gamma}\boldsymbol{\lambda_\gamma})^\top(\boldsymbol{a}-\boldsymbol{\beta_\gamma}\boldsymbol{\lambda_\gamma})} d\boldsymbol{\lambda} d\sigma^2 \\ &= \iint (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{N\hat{\sigma}^2_\gamma}{2\sigma^2}} e^{-\frac{(\boldsymbol{\lambda_\gamma}-\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}})^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top \boldsymbol{\beta_\gamma}(\boldsymbol{\lambda_\gamma}-\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}})}{2\sigma^2}} d\boldsymbol{\lambda} d\sigma^2 \\ &= (2\pi)^{\frac{p_\gamma+1}{2}} |\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma}|^{-\frac{1}{2}} \int (\sigma^2)^{-\frac{N-p_\gamma+1}{2}} e^{-\frac{N\hat{\sigma}^2_\gamma}{2\sigma^2}} d\sigma^2 = (2\pi)^{\frac{p_\gamma+1}{2}} |\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma}|^{-\frac{1}{2}} \frac{\Gamma(\frac{N-p_\gamma+1}{2})}{(\frac{N\hat{\sigma}^2_\gamma}{2})^{\frac{N-p_\gamma+1}{2}}}, \end{aligned}$$

where $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma})^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{a}$, $\hat{\sigma}^2_\gamma = \frac{(\boldsymbol{a}-\boldsymbol{\beta_\gamma}\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}})^\top(\boldsymbol{a}-\boldsymbol{\beta_\gamma}\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}})}{N}$ and Γ denotes the Gamma function.

A.1.3 Proof of Proposition 2

Proof.
Sampling $\boldsymbol{\lambda}_\gamma$. From Bayes' theorem we have that

$$\begin{aligned} p(\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}) &\propto p(data|\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma}) \\ &\propto (2\pi)^{-\frac{p_\gamma}{2}} |\boldsymbol{D_\gamma}|^{\frac{1}{2}} (\sigma^2)^{-\frac{N+p_\gamma}{2}} e^{-\frac{1}{2\sigma^2}[(\boldsymbol{a}-\boldsymbol{\beta_\gamma}\boldsymbol{\lambda_\gamma})^\top(\boldsymbol{a}-\boldsymbol{\beta_\gamma}\boldsymbol{\lambda_\gamma})+\boldsymbol{\lambda}_{\boldsymbol{\gamma}}^\top\boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}]} \\ &= (2\pi)^{-\frac{p_\gamma}{2}} |\boldsymbol{D_\gamma}|^{\frac{1}{2}} (\sigma^2)^{-\frac{N+p_\gamma}{2}} e^{-\frac{(\boldsymbol{\lambda_\gamma}-\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}})^\top(\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma}+\boldsymbol{D_\gamma})(\boldsymbol{\lambda_\gamma}-\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}})}{2\sigma^2}} e^{\left\{-\frac{SSR_\gamma}{2\sigma^2}\right\}}, \end{aligned}$$

where $SSR_\gamma = \boldsymbol{a}^\top\boldsymbol{a} - \boldsymbol{a}^\top\boldsymbol{\beta_\gamma}(\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma})^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{a} = \min_{\boldsymbol{\lambda_\gamma}}\{(\boldsymbol{a} - \boldsymbol{\beta_\gamma}\boldsymbol{\lambda_\gamma})^\top(\boldsymbol{a} - \boldsymbol{\beta_\gamma}\boldsymbol{\lambda_\gamma}) + \boldsymbol{\lambda}_{\boldsymbol{\gamma}}^\top\boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}\}$. Hence, defining $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma})^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{a}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma) = \sigma^2(\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma})^{-1}$, we obtain the posterior distribution in (16).

[39] More precisely, the priors for $(\boldsymbol{\lambda}, \sigma^2)$ are $\pi(\boldsymbol{\lambda_\gamma}, \sigma^2) \propto \frac{1}{\sigma^2}$ and $\boldsymbol{\lambda}_{-\boldsymbol{\gamma}} = 0$.

Using our priors and integrating out $\boldsymbol{\lambda}$ yields

$$p(data|\sigma^2,\boldsymbol{\gamma}) = \int p(data|\boldsymbol{\lambda},\sigma^2,\boldsymbol{\gamma})\pi(\boldsymbol{\lambda}|\sigma^2,\boldsymbol{\gamma})d\boldsymbol{\lambda} \propto (\sigma^2)^{-\frac{N}{2}} \frac{|\boldsymbol{D_\gamma}|^{\frac{1}{2}}}{|\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D_\gamma}|^{\frac{1}{2}}} e^{-\frac{SSR_\gamma}{2\sigma^2}}.$$

Sampling σ^2. From the Bayes' theorem we have that the posterior of σ^2 given by

$$p(\sigma^2|data,\boldsymbol{\gamma}) \propto p(data|\sigma^2,\boldsymbol{\gamma})\pi(\sigma^2) \propto (\sigma^2)^{-\frac{N}{2}-1} e^{-\frac{SSR_\gamma}{2\sigma^2}}.$$

Hence, the posterior distribution of σ^2 is the inverse-Gamma in (17).

Finally, we obtain the marginal likelihood of the data in (18) by integrating out σ^2as follows:

$$p(data|\boldsymbol{\gamma}) = \int p(data|\sigma^2,\boldsymbol{\gamma})\pi(\sigma^2)d\sigma^2 \propto \frac{|\boldsymbol{D_\gamma}|^{\frac{1}{2}}}{|\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D_\gamma}|^{\frac{1}{2}}} \frac{1}{(SSR_\gamma/2)^{\frac{N}{2}}},$$

where $SSR_\gamma = \boldsymbol{a}^\top\boldsymbol{a} - \boldsymbol{a}^\top\boldsymbol{\beta}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{D_\gamma})^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{a}$. ■

A.1.4 Proof of Corollary 1

Proof. To begin with, we introduce the following matrix notations:

$$\boldsymbol{\beta_\gamma} = (\boldsymbol{\beta_{\gamma'}},\boldsymbol{\beta_p}),\ \ \boldsymbol{D_\gamma} = \begin{pmatrix} \boldsymbol{D_{\gamma'}} & \\ & \frac{1}{\psi_p} \end{pmatrix},\ \ \boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma} = \begin{pmatrix} \boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}} & \boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_p} \\ \boldsymbol{\beta}_{\boldsymbol{p}}^\top\boldsymbol{\beta_{\gamma'}} & \boldsymbol{\beta}_{\boldsymbol{p}}^\top\boldsymbol{\beta_p} + \frac{1}{\psi_p} \end{pmatrix},$$

$$|\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma}| = |\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}| \times |\boldsymbol{\beta}_{\boldsymbol{p}}^\top\boldsymbol{\beta_p} + \frac{1}{\psi_p} - \boldsymbol{\beta}_{\boldsymbol{p}}^\top\boldsymbol{\beta_{\gamma'}}(\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}})^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_p}|, \text{ and}$$

$$|\boldsymbol{D_\gamma}| = |\boldsymbol{D_{\gamma'}}| \times \frac{1}{\psi_p}.$$

Equipped with the above, we have by direct calculation

$$\begin{aligned}
\frac{p(data|\gamma_j = 1, \boldsymbol{\gamma_{-j}})}{p(data|\gamma_j = 0, \boldsymbol{\gamma_{-j}})} &= \frac{\frac{|\boldsymbol{D_\gamma}|^{\frac{1}{2}}}{|\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma}|^{\frac{1}{2}}} \frac{1}{\left(\frac{SSR_\gamma}{2}\right)^{\frac{N}{2}}}}{\frac{|\boldsymbol{D_{\gamma'}}|^{\frac{1}{2}}}{|\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}|^{\frac{1}{2}}} \frac{1}{\left(\frac{SSR_{\gamma'}}{2}\right)^{\frac{N}{2}}}} = \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \left(\frac{|\boldsymbol{D_\gamma}|}{|\boldsymbol{D_{\gamma'}}|}\right)^{\frac{1}{2}} \left(\frac{|\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}|}{|\boldsymbol{\beta}_{\boldsymbol{\gamma}}^\top\boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma}|}\right)^{\frac{1}{2}} \\
&= \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \psi_p^{-\frac{1}{2}} \left|\boldsymbol{\beta}_{\boldsymbol{p}}^\top\boldsymbol{\beta_p} + \frac{1}{\psi_p} - \boldsymbol{\beta}_{\boldsymbol{p}}^\top\boldsymbol{\beta_{\gamma'}}\left(\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}\right)^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_p}\right|^{-\frac{1}{2}} \\
&= \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \left(1 + \psi_p\boldsymbol{\beta}_{\boldsymbol{p}}^\top\left[\boldsymbol{I_N} - \boldsymbol{\beta_{\gamma'}}\left(\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}\right)^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma'}}^\top\right]\boldsymbol{\beta_p}\right)^{-\frac{1}{2}},
\end{aligned}$$

where $\boldsymbol{\beta_p}^\top \left[\boldsymbol{I_N} - \boldsymbol{\beta_{\gamma'}}(\boldsymbol{\beta_{\gamma'}}^\top \boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}})^{-1}\boldsymbol{\beta_{\gamma'}}^\top\right]\boldsymbol{\beta_p} = \min_{\boldsymbol{b}}\{(\boldsymbol{\beta_p} - \boldsymbol{\beta_{\gamma'}}\boldsymbol{b})^\top(\boldsymbol{\beta_p} - \boldsymbol{\beta_{\gamma'}}\boldsymbol{b}) + \boldsymbol{b}^\top \boldsymbol{D_{\gamma'}}\boldsymbol{b}\}$, which is the minimal value of the penalised sum of squared errors when we use $\boldsymbol{\beta_{\gamma'}}$ to predict $\boldsymbol{\beta_p}$. ■

A.1.5 Proof of Proposition 4

Proof.
Sampling $\boldsymbol{\lambda_\gamma}$. Combining the likelihood and the prior for $\boldsymbol{\lambda}$ we have the following:

$$p(\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}) \propto p(data|\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma})p(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma}) \propto e^{-\frac{1}{2\sigma^2}\left[\boldsymbol{\lambda}^\top(\boldsymbol{\beta}^\top\boldsymbol{\beta}+\boldsymbol{D})\boldsymbol{\lambda}-2\boldsymbol{\lambda}^\top\boldsymbol{\beta}^\top\boldsymbol{a}\right]}.$$

Therefore, defining $\hat{\boldsymbol{\lambda}} = (\boldsymbol{\beta}^\top\boldsymbol{\beta} + \boldsymbol{D})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}) = \sigma^2(\boldsymbol{\beta}^\top\boldsymbol{\beta} + \boldsymbol{D})^{-1}$, we have the posterior in equation (23).
Sampling $\{\gamma_j\}_{j=1}^K$. Given a ω_j, the conditional Bayes factor for the j-th risk factor is[40]

$$\frac{p(\gamma_j = 1|data, \boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2, \boldsymbol{\gamma_{-j}})}{p(\gamma_j = 0|data, \boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2, \boldsymbol{\gamma_{-j}})} = \frac{\omega_j}{1-\omega_j}\frac{p(\lambda_j|\gamma_j = 1, \sigma^2)}{p(\lambda_j|\gamma_j = 0, \sigma^2)}.$$

Sampling $\boldsymbol{\omega}$. From Bayes' theorem we have

$$\begin{aligned} p(\omega_j|data, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^2) &\propto \pi(\omega_j)\pi(\gamma_j|\omega_j) \propto \omega_j^{\gamma_j}(1-\omega_j)^{1-\gamma_j}\omega_j^{a_\omega-1}(1-\omega_j)^{b_\omega-1} \\ &\propto \omega_j^{\gamma_j+a_\omega-1}(1-\omega_j)^{1-\gamma_j+b_\omega-1}. \end{aligned}$$

Therefore, the posterior distribution of ω_j is the Beta in equation (25).
Sampling σ^2. Finally,

$$p(\sigma^2|data, \boldsymbol{\omega}, \boldsymbol{\lambda}, \boldsymbol{\gamma}) \propto (\sigma^2)^{-\frac{N+K+1}{2}-1}e^{-\frac{1}{2\sigma^2}[(\boldsymbol{a}-\boldsymbol{\beta\lambda})^\top(\boldsymbol{a}-\boldsymbol{\beta\lambda})+\boldsymbol{\lambda}^\top\boldsymbol{D}\boldsymbol{\lambda}]}.$$

Hence, the posterior distribution of σ^2 is the inverse-Gamma in equation (26). ■

[40] If we had instead imposed $\omega_j = 0.5$, as in section II.2.2, the Bayes factor would be fully determined by $\frac{p(\lambda_j|\gamma_j=1,\sigma^2)}{p(\lambda_j|\gamma_j=0,\sigma^2)}$.

A.2 Additional tables

Table A1: List of factors for cross-sectional asset pricing models

Factor ID	reference	Factor ID	Reference
MKT	Sharpe (1964, Journal of Finance), Lintner (1965, Journal of Finance)	HML_DEVIL	Asness and Frazzini (2013, Journal of Portfolio Management)
SMB	Fama and French (1992, Journal of Finance)	QMJ	Asness, Frazzini, and Pedersen (2019, Review of Accounting Studies)
HML	Fama and French (1992, Journal of Finance)	FIN_UNC	Jurado, Ludvigson, and Ng (2015, American Economy Review), Ludvigson, Ma, and Ng (2019, AEJ: Macroeconomics)
RMW	Fama and French (2015, Journal of Financial Economics)	REAL_UNC	Jurado, Ludvigson, and Ng (2015, American Economy Review), Ludvigson, Ma, and Ng (2019, AEJ: Macroeconomics)
CMA	Fama and French (2015, Journal of Financial Economics)	MACRO_UNC	Jurado, Ludvigson, and Ng (2015, American Economy Review), Ludvigson, Ma, and Ng (2019)
UMD	Carhart (1997, Journal of Finance), Jegadeesh and Titman (1993, Journal of Finance)	TERM	Chen, Ross and Roll (1986, Journal of Business), Fama and French (1993, Journal of Financial Economics)
STREV	Jegadeesh and Titman (1993, Journal of Finance)	DELTA_SLOPE	Ferson and Harvey (1991, Journal of Political Economy)
LTREV	Jegadeesh and Titman (2001, Journal of Finance)	CREDIT	Chen, Ross and Roll (1986, Journal of Business), Fama and French (1993, Journal of Financial Economics)
q_IA	Hou, Xue, Zhang (2015, Review of Financial Studies)	DIV	Campbell (1996, Journal of Political Economy)
q_ROE	Hou, Xue, Zhang (2015, Review of Financial Studies)	PE	Basu (1977, Journal of Finance), Ball (1978, Journal of Financial Economics)
LIQ_NT	Pastor and Stambaugh (2003, Journal of Political Economy)	BW_INV_SENT	Baker and Wurgler (2006, Journal of Finance)
LIQ_TR	Pastor and Stambaugh (2003, Journal of Political Economy)	HJTZ_INV_SENT	Huang, Jiang, Tu, and Zhou (2015, Review of Financial Studies)
MGMT	Stambaugh and Yuan (2016, Review of Financial Studies)	BEH_PEAD	Daniel, Hirshleifer, and Sun (2019, Review of Financial Studies)
PERF	Stambaugh and Yuan (2016, Review of Financial Studies)	BEH_FIN	Daniel, Hirshleifer, and Sun (2019, Review of Financial Studies)
ACCR	Sloan (1996, Accounting Review)	MKT*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
DISSTR	Campbell, Hilscher, and Szilagyi (2008, Journal of Finance)	SMB*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
ASSET_Growth	Cooper, Gulen, and Schill (2008, Journal of Finance)	HML*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
COMP_ISSUE	Daniel and Titman (2006, Journal of Finance)	RMW*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
GR_PROF	Novy-Marx (2013, Journal of Financial Economics)	CMA*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
INV_IN_ASSETS	Titman, Wei, and Xie (2004, Journal of Financial and Quantitative Analysis)	SKEW	Langlois (2019, Journal of Financial Economics)
NetOA	Hirshleifer, Kewei, Teoh, and Zhang (2004, Journal of Accounting and Economics)	NONDUR	Chen, Ross and Roll (1986, Journal of Business), Breeden, Gibbons, and Litzenberger (1989, Journal of Finance)
OSCORE	Ohlson (1980, Journal of Accounting Research)	SERV	Breeden, Gibbons, and Litzenberger (1989, Journal of Finance), Hall (1978, Journal of Political Economy)
ROA	Chen, Novy-Marx, and Zhang (2010, working paper)	UNRATE	Gertler and Grinols (1982, Journal of Money, Credit, and Banking)
STOCK_ISS	Ritter (1991, Journal of Finance), Fama and French (2008, Journal of Finance)	IND_PROD	Chan, Chen, and Hsieh (1985, Journal of Financial Economics), Chen, Ross and Roll (1986, Journal of Business)
INTERM_CR	He, Kelly, and Manela (2017, Journal of Financial Economics)	OIL	Chen, Ross and Roll (1986, Journal of Business)
BAB	Frazzini and Pedersen (2014, Journal of Financial Economics)		

The table presents the list of factors used in Section IV.2. For each of the variables we present their identification index, the nature of the factor, and the source of data for downloading and/or constructing the time series. Full description of the factors, sources, and references can be found in Table OA13 of the Online Appendix.

Table A2: Posterior factor probabilities and risk premia of 2.6 million sparse models

	$\mathbb{E}[\gamma_j \| \text{data}]$						$\mathbb{E}[\lambda_j \| \text{data}]$						
	ψ:						ψ:						
Factors:	1	5	10	20	50	100	1	5	10	20	50	100	$\bar{F}$
HML	0.501	0.727	0.768	0.776	0.759	0.739	0.104	0.213	0.240	0.252	0.253	0.249	0.377
MKT*	0.212	0.362	0.386	0.393	0.381	0.354	0.030	0.114	0.149	0.178	0.199	0.197	0.514
SMB*	0.202	0.312	0.315	0.300	0.275	0.260	0.025	0.083	0.099	0.105	0.105	0.103	0.215
PERF	0.118	0.135	0.130	0.119	0.103	0.092	-0.013	-0.033	-0.038	-0.039	-0.037	-0.035	0.651
CMA	0.116	0.132	0.128	0.120	0.107	0.099	0.008	0.019	0.022	0.023	0.023	0.022	0.351
STOCK_ISS	0.095	0.112	0.123	0.129	0.125	0.117	-0.006	-0.022	-0.034	-0.045	-0.053	-0.055	0.515
COMP_ISSUE	0.096	0.108	0.115	0.119	0.117	0.110	0.007	0.026	0.040	0.054	0.065	0.068	0.497
MKT	0.074	0.069	0.085	0.110	0.132	0.133	0.003	0.012	0.025	0.045	0.068	0.075	0.563
UMD	0.087	0.089	0.093	0.093	0.089	0.085	0.004	0.013	0.019	0.025	0.029	0.032	0.646
DISSTR	0.086	0.087	0.090	0.091	0.086	0.078	0.009	0.029	0.043	0.055	0.063	0.062	0.475
ROA	0.092	0.090	0.090	0.087	0.078	0.067	0.009	0.024	0.032	0.038	0.041	0.037	0.551
BEH_PEAD	0.082	0.075	0.079	0.087	0.113	0.142	0.001	0.002	0.004	0.008	0.020	0.039	0.619
STRev	0.081	0.072	0.073	0.080	0.102	0.127	0.000	0.002	0.004	0.009	0.024	0.048	0.438
NONDUR	0.081	0.072	0.073	0.079	0.098	0.124	0.000	0.000	0.001	0.001	0.003	0.008	0.151*
NetOA	0.082	0.074	0.075	0.079	0.084	0.083	0.001	0.004	0.007	0.011	0.019	0.024	0.544
SMB	0.079	0.070	0.074	0.081	0.086	0.082	0.008	0.010	0.012	0.015	0.017	0.017	0.257
TERM	0.081	0.071	0.071	0.076	0.091	0.110	0.000	0.001	0.001	0.002	0.006	0.013	0.962*
BW_ISENT	0.081	0.071	0.071	0.074	0.085	0.095	0.000	0.000	0.001	0.001	0.003	0.005	0.101*
IPGrowth	0.081	0.070	0.070	0.072	0.081	0.091	0.000	0.000	0.000	0.000	-0.001	-0.002	0.097*
DeltaSLOPE	0.081	0.070	0.070	0.072	0.081	0.092	0.000	0.000	0.000	0.000	-0.001	-0.002	0.059*
Oil	0.080	0.070	0.069	0.072	0.079	0.088	0.000	0.001	0.002	0.005	0.013	0.026	0.740*
SERV	0.080	0.070	0.069	0.071	0.078	0.087	0.000	0.000	0.000	0.000	0.000	0.000	0.045*
FIN_UNC	0.080	0.070	0.069	0.071	0.078	0.085	0.000	0.000	0.000	0.000	0.000	0.000	0.103*
HJTZ_ISENT	0.080	0.070	0.069	0.071	0.077	0.084	0.000	0.000	0.000	0.000	-0.001	-0.001	0.242*
DIV	0.080	0.070	0.069	0.071	0.078	0.085	0.000	0.000	0.000	0.000	-0.001	-0.002	0.926*
DEFAULT	0.080	0.070	0.069	0.071	0.077	0.084	0.000	0.000	0.000	0.000	0.000	0.000	0.333*
PE	0.080	0.070	0.069	0.071	0.077	0.085	0.000	-0.001	-0.002	-0.003	-0.007	-0.014	6.770*
REAL_UNC	0.080	0.070	0.069	0.071	0.077	0.082	0.000	0.000	0.000	0.000	0.000	0.000	0.046*
MGMT	0.088	0.079	0.075	0.068	0.057	0.050	0.005	0.012	0.014	0.015	0.015	0.014	0.631
UNRATE	0.080	0.070	0.069	0.071	0.077	0.082	0.000	0.000	0.000	0.000	-0.001	-0.002	1.157*
LIQ_TR	0.080	0.069	0.069	0.070	0.077	0.084	0.000	0.000	0.001	0.002	0.005	0.010	0.438
LIQ_NT	0.081	0.070	0.069	0.070	0.076	0.081	-0.001	-0.001	-0.001	-0.002	-0.003	-0.005	0.428*
MACRO_UNC	0.081	0.070	0.069	0.070	0.073	0.075	0.000	0.000	0.000	0.000	0.000	0.000	0.078*
LTRev	0.080	0.070	0.070	0.072	0.069	0.063	-0.001	-0.004	-0.007	-0.012	-0.017	-0.017	0.252
INV_IN_ASSETS	0.080	0.069	0.068	0.069	0.072	0.073	0.000	0.001	0.001	0.001	0.003	0.005	0.549
CMA*	0.081	0.069	0.068	0.068	0.071	0.071	0.000	0.000	0.000	0.000	-0.001	-0.002	0.242
ACCR	0.082	0.070	0.068	0.067	0.066	0.064	-0.001	-0.003	-0.003	-0.004	-0.005	-0.005	0.343
HML*	0.080	0.068	0.066	0.067	0.066	0.063	0.000	0.000	0.001	0.002	0.003	0.004	0.251
INTERM_CAP_RATIO	0.086	0.073	0.068	0.063	0.055	0.048	0.007	0.013	0.015	0.016	0.018	0.017	0.719*
RMW*	0.079	0.068	0.066	0.064	0.058	0.052	0.000	0.002	0.003	0.004	0.005	0.006	0.219
ASSET_Growth	0.081	0.068	0.065	0.062	0.057	0.051	-0.001	-0.002	-0.002	-0.003	-0.004	-0.005	0.525
QMJ	0.098	0.076	0.064	0.052	0.040	0.034	0.007	0.010	0.009	0.008	0.007	0.006	0.405
BAB	0.082	0.069	0.065	0.060	0.053	0.046	-0.003	-0.007	-0.008	-0.010	-0.011	-0.010	0.921
IA	0.087	0.071	0.064	0.056	0.048	0.042	-0.003	-0.005	-0.006	-0.006	-0.006	-0.005	0.409
O_SCORE	0.080	0.062	0.055	0.047	0.039	0.032	-0.002	-0.004	-0.005	-0.005	-0.004	-0.003	0.02
ROE	0.078	0.059	0.053	0.047	0.038	0.032	0.001	0.003	0.004	0.005	0.005	0.005	0.555
GR_PROF	0.081	0.056	0.045	0.037	0.028	0.022	0.004	0.005	0.004	0.003	0.002	0.002	0.199
BEH_FIN	0.077	0.055	0.046	0.038	0.030	0.024	-0.002	-0.001	0.000	0.001	0.001	0.001	0.76
SKEW	0.078	0.052	0.042	0.034	0.026	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.438
RMW	0.073	0.047	0.039	0.033	0.025	0.021	0.002	0.002	0.002	0.003	0.002	0.002	0.292
HML_DEVIL	0.065	0.040	0.032	0.025	0.019	0.015	0.001	0.002	0.002	0.001	0.001	0.001	0.356

Posterior probabilities of factors, $\Pr[\gamma_j = 1|\text{data}]$, and posterior mean of factor risk premia, $\mathbb{E}[\lambda_j|\text{data}]$, computed using the Dirac spike-and-slab approach of section II.2.2, 51 factors, and all possible models with up to 5 factors, yielding about 2.6 million candidate models. The prior probability of a factor being included is about 10.38%. The data is monthly, 1973:10 to 2016:12. Test assets: cross-section of 25 Fama-French size-and-book-to-market and 30 industry portfolios. The 51 factors considered are described in Table A1 of the Appendix.