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

Svetlana Bryzgalova† Jiantao Huang‡ Christian Julliard§

May 23, 2022

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 price of risk estimates for 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 Bayesian model averaging (BMA-SDF), if there is no clear winner. We analyze 2.25 quadrillion models generated by a large set of factors, and find that the BMA-SDF outperforms existing models in- and out-of-sample.

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, Doron Avramov, Mikhail Chernov, Pierre Collin-Dufresne, Aureo de Paula, Marcelo Fernandes, Stefano Giglio, Rodrigo Guimaraes, Raymond Kan, Bryan Kelly, Lars Lochstoer, Albert Marcet, Marcelo Medeiros, Alexander Michaelides, Olivier Scaillet, Chris Sims, George Tauchen, Fabio Trojani, Dacheng Xiu, Motohiro Yogo, Irina Zviadadze, and seminar and conference participants at HBS, Princeton University, Carnegie Mellon, Cambridge Judd, ICEF Moscow, Goethe University Frankfurt, University College London, University of Lugano, London Business School, London School of Economics, Second David Backus Memorial Conference on Macro-Finance, SITE Summer Workshops, SoFiE seminar, SITE workshop on Asset Pricing, Macro Finance, and Computation, AFA 2021, Fourth International Workshop in Financial Econometrics, SOFIE virtual seminar, Virtual Finance Workshop, and CEPR Advanced Forum in Financial Economics, NBER Asset Pricing 2021, Brazilian Finance Society. We have read the *Journal of Finance*'s disclosure policy and have no conflict of interest to disclose.

†Department of Finance, London Business School, and CEPR; 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, thanks to the factor zoo phenomenon, in the near future we might have as many empirically "priced" sources of risk as stock returns. Second, the so-called weak factors (i.e., factors whose true covariance with asset returns is asymptotically zero) are likely to both appear empirically relevant and invalidate inference on the true sources of risk (see, e.g., Gospodinov, Kan, and Robotti (2019), and Kleibergen and Zhan (2020)). 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 entire factor zoo, *iii*) remains valid under misspecification, *iv*) is robust to the weak inference problem, and, importantly, *v*) delivers an empirical pricing kernel that outperforms (in- and out-of-sample) popular models (with either observable or latent factors). And that is exactly what we provide.

We develop a *unified framework* for tackling linear asset pricing models. In the case of stand-alone model estimation, our method provides reliable price of risk estimates, hypothesis testing, and confidence intervals for these parameters, as well as all other objects of interest – alphas, R^2's, Sharpe ratios, etc. Furthermore, even when all the pricing kernels are misspecified and non-nested, our approach delivers factor selection – *if* a dominant model exists – or model averaging, if there is no clear winner given the data. The method is numerically simple, fast, easy to use, and can be feasibly applied to literally quadrillions of candidate factor models.

Empirically, we find that the Stochastic Discount Factor (SDF) constructed as the Bayesian Model Averaging (BMA) over the space of 2.25 quadrillion models, prices a wide cross-section of anomalies better than both celebrated (observable) factor models and the latent factor approach of Kozak, Nagel, and Santosh (2020). This outperformance arises not only in sample but also out-of-sample in *both* time series and cross-sectional dimensions.[1] There are three key drivers of this performance. First, our method reliably identifies a small subset of observable factors that should be included in any SDF with high probability. Second, although these factors alone are already sufficient to outperform notable (observable) factor models, they do not fully characterize the SDF. The latter, as we show, is *dense* in the space of observable factors. As a result, the BMA optimally (in the predictive density sense) aggregates multiple imperfect measures of the same sources of risk. Third, our method relies on a novel prior that is fully driven by the researcher's belief about the Sharpe ratio in the economy, and that effectively controls potential overfitting. The BMA-SDF neither requires arbitrary tuning parameters nor separates factor extraction and aggregation. Instead, unlike most of the existing literature, it delivers an SDF in one step, driven by transparent and economically motivated priors.

[1]In cross-sectional out-of-sample exercises, we first estimate the BMA-SDF in a baseline cross-section, and then use it to price several other cross-sections *without* any further parameter estimation.

As stressed by Harvey (2017) in his AFA presidential address, the factor zoo naturally calls for a Bayesian solution – and we develop one. Furthermore, we show that factor proliferation and spurious inference are tightly connected problems, and a naïve Bayesian model selection fails in the presence of weak factors. We develop a reliable solution focused on the SDF representation, since the key question posed by the factor zoo lies in whether candidate risk factors have non-zero price of risk. Our Bayesian SDF formulation (B-SDF) is intuitively similar to the standard frequentist OLS/GLS estimation that imposes the self-pricing of tradable factors when they are part of the test assets. However, it is robust to identification failure, allows us to easily compare and aggregate non-nested models, and provides robust inference for all the quantities of interest within stand-alone models and across the whole model space. Remarkably, unlike the frequentist alternatives, the B-SDF estimator performs well in both small and large samples, even with fairly large cross-sections.

Our empirical results are based on what is arguably a representative cross-section of test assets: 60 portfolios based on a large number of firm-specific characteristics. We examine 51 factors proposed in the previous literature, yielding a total of 2.25 quadrillion possible models to analyze. We find that only a handful of factors proposed in the literature are robust explanators of the cross-section of returns, and a three (at most six) *most likely* factor model easily outperforms canonical reduced-form benchmarks. Nevertheless, there is no clear "winner" across the whole space of potential models: Hundreds of possible specifications that combine tradable and non-tradable factors, 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 variables is needed to fully capture its pricing implications.[2] 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. BMA, therefore, emerges naturally as an optimal way of aggregating models that load on the same set of underlying risks: It aggregates all the possible factors and models based on their likelihood to have generated the data. Crucially, this approach allows for *both* selection and aggregation based on the posterior probabilities of the factors being part of the pricing kernel, and allows the data to decide on the optimal structure of the SDF. Empirically, we find that the BMA-SDF performs well both in- and out-of-sample (OOS). Its OOS performance is stable across subsamples (going both into the future and into the past), and, most importantly, it prices well cross-sections not used for its construction, including the notoriously challenging 49 industry portfolios.

Our contribution is fourfold. First, we develop a very simple Bayesian estimator for linear

[2]Interestingly, the SDF remains dense even when we include either the five principal components or the five RP-PCs of Lettau and Pelger (2020).

SDFs with both traded and non-traded factors. This approach makes weak factors easily detectable in finite sample, while providing valid inference on the strong factors' price of risk, measures of cross-sectional fit, and other objects of interest. Our robust approach is very simple to implement and use, and it does not require pre-testing or pre-estimation.

Second, we provide a method for inference on the *entire* factor zoo with 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 a flat prior for the price of risk: Asymptotically, weakly identified factors are selected with probability one even if they have zero price of risk.[3] 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. Building upon the literature on predictor selection (see, e.g., Ishwaran, Rao, et al. (2005) and Giannone, Lenza, and Primiceri (2021)), 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. It is uninformative (the "slab") for strong factors but shrinks away (the "spike") the weak ones. This prior also: *i*) makes it computationally feasible to analyze quadrillions of alternative factor models, *ii*) allows the researcher to encode prior beliefs (or lack thereof) about the sparsity of the true SDF without imposing hard thresholds, *iii*) restores the validity of hypothesis testing, and *iv*) performs well in numerous simulation settings. The prior is entirely pinned down by economic quantities: It maps into beliefs about the Sharpe ratio of the risk factors. We regard this approach as a solution for the high-dimensional inference problem generated by the factor zoo.[4]

Third, we provide a new way of selecting robust observable factors. Indeed, we find a new 3–6 observable factor model, combining variables from different papers, that dominates all the popular reduced-form benchmarks. However, even that model would be strongly rejected by the data: No sparse factor model is among the most likely 2000 data-generating processes that we consider. Furthermore, a *unique* best performing combination of the factors (sparse or dense in observables) does not seem to exist: Hundreds of possible models, never proposed in the previous literature, deliver almost equivalent performance, which indicates fragility of conventional model selection and horse races, popular among reduced-form sparse factor models.

Fourth, our results do not rely on ex ante unverifiable assumptions of existence, uniqueness, and sparsity of the true SDF representation among the candidate models (unlike LASSO and

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

[4] Despite a seemingly prohibitive dimension of the model space, the estimation is numerically simple and computationally feasible. Our Markov Chain, used to evaluate the whole space of 2.25 quadrillions of models and deliver all the baseline results from the paper, takes about four hours on a 3.0GHz 10-core Intel Xeon W processor and 128 GB of RAM. Furthermore, we formally test its convergence and establish that the posterior distributions converge already after *less than one fifth* of the Markov Chain draws, making our method easily applicable for most researchers.

other popular frequentist methods). When a dominant model for the SDF does not arise in the data (as in our analysis), our method does not stop at selection. Instead, it efficiently aggregates pricing information from (potentially) the entire factor zoo. Interestingly, we show that solely extracting leading standard latent factors from a wide range of predictors using PCA or RP-PCA, is not sufficient to characterize the SDF. In fact, we find that observable and (some) leading latent factors are complementary for such a characterization. Therefore, our results indicate that there is scope for both more efficient latent factor extraction and better aggregation informed by economic fundamentals.

The remainder of the paper is organized as follows. In the next subsection we review the most closely related literature and our contribution to it. Section II provides a brief overview of the benchmark frequentist approach, while Section III outlines the Bayesian SDF estimation and its properties for inference, selection, and model aggregation. Section IV provides simulation evidence on both small- and large-sample behavior of our method. Section V presents our empirical results. Finally, Section VI discusses potential extensions of our procedure and concludes.[5]

I.1 Closely Related Literature

There are numerous strands of literature relying on Bayesian tools, especially for asset allocation (for an excellent overview, see Avramov and Zhou (2010)), model selection (e.g., Chib, Zeng, and Zhao (2020)), and performance evaluation (Baks, Metrick, and Wachter (2001), Pástor and Stambaugh (2002), and Harvey and Liu (2019)). Therefore, we aim to provide only an overview of the literature that is most closely related to our paper.

Shanken (1987) and Harvey and Zhou (1990) are probably the first to use the Bayesian framework in portfolio choice and develop GRS-type tests (cf. Gibbons, Ross, and Shanken (1989)) for mean-variance efficiency. While Shanken (1987) is 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 priors on the deep model parameters. Interestingly, we show that there is a tight link between using the most popular, diffuse, priors for the price of risk and the failure of the standard estimation techniques in the presence of weak factors.

Pástor and Stambaugh (2000) and Pástor (2000) assign a prior distribution to the vector of pricing errors $\boldsymbol{\alpha}$, $\boldsymbol{\alpha} \sim \mathcal{N}(0, \kappa \boldsymbol{\Sigma_R})$, where $\boldsymbol{\Sigma_R}$ is the variance-covariance matrix of returns and $\kappa \in \mathbb{R}_+$, and apply it to portfolio choice. This prior imposes a degree of shrinkage on the alphas: When factor models are misspecified, pricing errors cannot be too large a priori. This prior effectively places a bound on the Sharpe ratio achievable in this economy.

[5] Additional results are reported in the Internet Appendix: `https://ssrn.com/abstract=3627010`.

Barillas and Shanken (2018) extend the aforementioned prior and derive a closed-form solution for the Bayes factor when all the 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.

There is a general close connection between the Bayesian approach to model selection and 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. The ridge-based approach to recovering the SDF of Kozak, Nagel, and Santosh (2020) can also be interpreted from a Bayesian perspective with priors on the expected returns distribution.

To the best of our knowledge, our paper is the first attempt to develop a general Bayesian approach for both tradable and non-tradable factors, capable of imposing tradable restriction on the price of risk when needed. Flat priors for the price of risk, we show, lead to erroneous model selection in the presence of weak factors. Hence, we develop a novel one that depends on the degree of parameter identification. This prior is heterogenous among factors, depending on the correlation between test assets and the factor itself. In the spirit of Pástor and Stambaugh (2000), our prior directly maps into beliefs about the Sharpe ratio achievable in the economy, yet without imposing a hard threshold on it. Not only does it restore the validity of model selection, but it also allows for sharp inference in small sample on all the economic quantities of interest.

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 estimators 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 such that when the vector of correlation coefficients between asset returns and a factor is close to zero, the prior variance for the price of risk also goes to zero, effectively shrinking the posterior toward zero.

Our method does not require any pretesting, works well in small and large time-series and cross-sectional dimensions. Furthermore, due to its hierarchical structure, it can be feasibly extended to handle time variation in the factor exposure and asset risk premia, and it accommodates both observable and latent factors. Most importantly, our approach provides a robust unified framework for evaluation of stand-alone models, factor and model selection, as well as aggregation, even when all the potential models are misspecified.

Naturally, our paper also contributes to the very active (and growing) body of work that critically re-evaluates existing findings in the empirical asset pricing literature and develop robust inference methods. There is ample empirical evidence that most linear factor models are misspecified (e.g., Chernov, Lochstoer, and Lundeby (2019), and He, Huang, and Zhou (2018)). 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. Gospodinov, Kan, and Robotti (2014) develop a general approach for misspecification-robust inference, while Giglio and Xiu (2021) exploit the invariance principle of the PCA and recover the price of risk of a given factor from the projection on the span of latent factors driving a cross-section of returns. Similarly, Uppal, Zaffaroni, and Zviadadze (2018) recover latent factors from the residuals of an asset pricing model, effectively completing the span of the SDF. Giglio, Feng, and Xiu (2020) 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 when the true SDF is sparse. Huang, Li, and Zhou (2018) use a reduced rank approach to select from not only the observable factors but their total span, effectively allowing for sparsity in both factors and their combinations.

We do not take a stand on the origin of the factors, the "unique" true model being among the candidate specifications, and a priori SDF sparsity. Instead, we consider the whole universe of potential models that can be created from a wide set of factors proposed in the empirical literature (observable and latent) and let the data speak. We find that the cross-sectional likelihood across many best-performing (dense) models is flat. Hence, the data seem to call for aggregation, rather than selection.

Avramov (2002, 2004) brought model uncertainty to the forefront of asset pricing. Building on these seminal papers, Anderson and Cheng (2016) develop a BMA approach to portfolio choice that, with formal recognition of model uncertainty, delivers robust asset allocation and superior out-of-sample performance. Similarly, we find that there is a large degree of model uncertainty in cross-sectional asset pricing, suggesting a large degree of model misspecification and rendering canonical selection unreliable. We therefore develop a BMA method that explicitly targets cross-sectional pricing of asset returns. The resulting averaging over the space of SDFs delivers superior pricing in- and out-of-sample.

In reality, the BMA-SDF has – endogenously – elements of both selection and aggregation: While a small subset of factors delivers large individual contributions to the SDF, other factors are efficiently bundled together to deliver the best predictive density of the cross-sectional pricing kernel. In the recent literature, model *selection* (see, e.g., Giglio, Feng, and Xiu (2020)) or *aggregation* (see, e.g., Kozak, Nagel, and Santosh (2020)) of pricing factors, have been largely mutually exclusive alternatives. Our framework, instead, successfully combines both.

II Frequentist Estimation of Linear SDFs

This section introduces the notation and reviews the basics of linear SDF models as well as related (frequentist) Generalized Method of Moments (GMM) estimation. 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. The returns of N test assets, which are long-short portfolios, are denoted by $\boldsymbol{R}_t = (R_{1t} \dots R_{Nt})^\top$. Throughout the paper, $\mathbb{E}[X]$ or μ_X denotes the unconditional expectation of arbitrary random variable X and $\bar{X}$ denotes the sample mean operator.

Consider linear stochastic discount factors (M), that is models of the form $M_t = 1 - (\boldsymbol{f}_t - \mathbb{E}[\boldsymbol{f}_t])^\top \boldsymbol{\lambda_f}$. In the absence of arbitrage opportunities $\mathbb{E}[M_t \boldsymbol{R_t}] = \boldsymbol{0_N}$, which implies that expected returns are given by $\boldsymbol{\mu_R} = \mathbb{E}[\boldsymbol{R}_t] = \boldsymbol{C_f}\boldsymbol{\lambda_f}$, where $\boldsymbol{C_f}$ is the covariance matrix between $\boldsymbol{R}_t$ and $\boldsymbol{f}_t$ and $\boldsymbol{\lambda_f} \in \mathbb{R}^K$ denotes the vector of prices of risk associated with the factors. The latter can therefore be estimated via the cross-sectional regression:

$$\boldsymbol{\mu_R} = \lambda_c \boldsymbol{1_N} + \boldsymbol{C_f}\boldsymbol{\lambda_f} + \boldsymbol{\alpha} = \boldsymbol{C}\boldsymbol{\lambda} + \boldsymbol{\alpha}, \tag{1}$$

where $\boldsymbol{C} = (\boldsymbol{1_N}, \boldsymbol{C_f})$, $\boldsymbol{\lambda}^\top = (\lambda_c, \boldsymbol{\lambda_f}^\top)$, λ_c is a scalar average mispricing (equal to zero under the null of the model being correctly specified), $\boldsymbol{1_N}$ 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 (also equal to zero under the null of the model).

Such a model is usually estimated via GMM (see Hansen (1982)) with the following moment conditions:

$$\mathbb{E}[\boldsymbol{g_t}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f})] = \mathbb{E}\begin{pmatrix} \boldsymbol{R_t} - \lambda_c \boldsymbol{1_N} - \boldsymbol{R_t}(\boldsymbol{f_t} - \boldsymbol{\mu_f})^\top \boldsymbol{\lambda_f} \\ \boldsymbol{f_t} - \boldsymbol{\mu_f} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0_N} \\ \boldsymbol{0_K} \end{pmatrix}, \tag{2}$$

with corresponding sample analogue function $\boldsymbol{g_T}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f}) \equiv \frac{1}{T}\sum_{t=1}^T \boldsymbol{g_t}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f})$. Combining the latter with a weighting matrix $\boldsymbol{W}$ yields the GMM estimates as the minimizer of the following objective function:

$$\{\widehat{\lambda}_c, \widehat{\boldsymbol{\lambda}}_{\boldsymbol{f}}, \widehat{\boldsymbol{\mu}}_{\boldsymbol{f}}\} \equiv \underset{\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f}}{\arg\min}\ \boldsymbol{g_T}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f})^\top \boldsymbol{W} \boldsymbol{g_T}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f}).$$

Different weighting matrices deliver different point estimates. Following Cochrane (2005, pp.256-

258), two popular choices are

$$\boldsymbol{W_{ols}} = \begin{pmatrix} \boldsymbol{I_N} & \boldsymbol{0_{N\times K}} \\ \boldsymbol{0_{K\times N}} & \kappa \boldsymbol{I_K}, \end{pmatrix}, \text{ and } \boldsymbol{W_{gls}} = \begin{pmatrix} \boldsymbol{\Sigma_R^{-1}} & \boldsymbol{0_{N\times K}} \\ \boldsymbol{0_{K\times N}} & \kappa \boldsymbol{I_K} \end{pmatrix},$$

where $\boldsymbol{\Sigma_R}$ is the covarince matrix of returns, and $\kappa > 0$ is a large constant so that $\widehat{\boldsymbol{\mu}}_{\boldsymbol{f}} \equiv \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{f_t}$. These weighting matrices yield, respectively, the following prices of risk estimates:

$$\widehat{\boldsymbol{\lambda}}_{\boldsymbol{ols}} = (\widehat{\boldsymbol{C}}^\top \widehat{\boldsymbol{C}})^{-1} \widehat{\boldsymbol{C}}^\top \bar{\boldsymbol{R}}, \text{ and} \tag{3}$$

$$\widehat{\boldsymbol{\lambda}}_{\boldsymbol{gls}} = (\widehat{\boldsymbol{C}}^\top \boldsymbol{\Sigma_R^{-1}} \widehat{\boldsymbol{C}})^{-1} \widehat{\boldsymbol{C}}^\top \boldsymbol{\Sigma_R^{-1}} \bar{\boldsymbol{R}}, \tag{4}$$

where $\widehat{\boldsymbol{C}} = (\boldsymbol{1_N}, \widehat{\boldsymbol{C}}_{\boldsymbol{f}})$ and $\widehat{\boldsymbol{C}}_{\boldsymbol{f}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{R_t}(\boldsymbol{f_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{f}})^\top$.

GMM provides valid inference on the price of risk under a set of well-known assumptions (Newey and McFadden (1994)). In particular, equations (3) and (4) make it clear that OLS and GLS (but also GMM more generally) require the matrix of factor exposures $\boldsymbol{C}$ to have full rank – that is, the price of risk to be identified. However, there is a growing body of literature that finds this assumption to be often empirically violated.[6] Most famously, this problem arises in the case of a *weak* factor f_j that does not have enough comovement with any of the assets but is nonetheless considered to be a part of the SDF, that is $C_{i,j} \sim O(T^{-1/2}),\ \ i \in 1 \ldots N$. In such a model, risk prices are no longer identified and their estimates diverge with the sample size, leading to wrong inference for *both* strong and weak factors (Kan and Zhang (1999a)). Another widespread example of weak identification arises with the inclusion of a *level* factor, f_j, characterized by a lack of cross-sectional spread in factor exposures, that is, $\sum_{i=1}^{N}(C_{i,j} - \bar{C}_j)^2 \sim O(T^{-1})$, where $\bar{C}_j \equiv \frac{1}{N}\sum_{i=1}^{N} C_{i,j}$.

Identification problems arise not only when using the GMM in estimating linear SDF models but equally so in Fama-MacBeth regressions (Kan and Zhang (1999b), Kleibergen (2009)) and Maximum Likelihood Estimation (Gospodinov, Kan, and Robotti (2019)). In addition to creating inference problems for model parameters, weak identification also tends to inflate the standard measures of cross-sectional fit (Kleibergen and Zhan (2015)). Consequently, several papers have attempted to develop alternative statistical procedures that are robust to the presence of weak factors and general cases of rank deficiency of the matrix $\boldsymbol{C}$. In particular, 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 GMM estimates of factor risk prices in the linear stochastic discount factor framework and prove that t-statistics calculated using their standard errors are robust even when the model is misspecified and a weak factor is included. Bryzgalova

[6]For recent applications, see Kleibergen and Zhan (2020) and Gospodinov and Robotti (2021a,b).

(2015) introduces a LASSO-like penalty term that identifies weak factors and eliminates their impact on the model. Finally, since factor strength depends on the choice of returns used in the estimation, Giglio, Xiu, and Zhang (2021) recently developed an iterative procedure for constructing a cross-section of model-specific test assets that specifically addresses the problem of weak factors.

In this paper, we provide a Bayesian inference and model selection framework that *i*) can be easily used for robust inference in the presence of weak and level factors (section III) 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 weak) risk factors – that is, the entire factor zoo.

Although we focus on the estimation of linear SDF representations, our approach can be adapted (with minimal adjustments) to deliver a robust Bayesian version of the canonical Fama-MacBeth estimation approach (see Fama and MacBeth (1973) and Fama and French (1993)). We consider this extension in a companion paper, Bryzgalova, Huang, and Julliard (2022).

III Bayesian Analysis of Linear SDFs

This section introduces our hierarchical Bayesian estimation of linear SDF models, B-SDF. A more detailed derivation is presented in Appendix A.1.1.

Consider first the time-series dimension of the estimation problem. Let $\boldsymbol{f}_t \equiv (f_{1t} \dots f_{Kt})^\top$, $t = 1, \dots T$ denote a vector of factors. Without loss of generality, we order the K_1 tradable factors first ($\boldsymbol{f}_t^{(1)}$), followed by K_2 non-tradable factors ($\boldsymbol{f}_t^{(2)}$), hence, $\boldsymbol{f} \equiv (\boldsymbol{f}_t^{(1),\top}, \boldsymbol{f}_t^{(2),\top})^\top$ and $K_1 + K_2 = K$.

Let $\boldsymbol{Y_t}$ denote the union of factors and returns, that is, $\boldsymbol{Y_t} \equiv \boldsymbol{f_t} \cup \boldsymbol{R_t}$, where $\boldsymbol{Y_t}$ is a p-dimensional vector. If one requires the tradable factors to price themselves (as we do in our empirical applications), then $\boldsymbol{Y_t}^\top \equiv (\boldsymbol{R_t}^\top, \boldsymbol{f}_t^{(2),\top})^\top$ and $p = N + K_2$.

We assume that $\{\boldsymbol{Y_t}\}_{t=1}^T$ follows an iid multivariate Gaussian distribution, that is, $\boldsymbol{Y_t} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$, where $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ denote, respectively, the unconditional means vector and the unconditional covariance matrix. This modeling choice can easily be modified to accommodate different distributional assumptions, predictability, and time-varying volatility, albeit at the cost of losing analytical solutions in most cases. In particular, as discussed in Section VI, we could accommodate time-varying means and variances, as well as autocorrelations. The resulting likelihood function for the time-series layer of our hierarchical modeling is

$$p(\mathbf{Y}|\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2} tr \left[\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \sum_{t=1}^{T} (\boldsymbol{Y_t} - \boldsymbol{\mu_Y})(\boldsymbol{Y_t} - \boldsymbol{\mu_Y})^\top\right]\right\}, \tag{5}$$

where $\mathbf{Y} \equiv \{\boldsymbol{Y_t}\}_{t=1}^{T}$. For simplicity, we use the diffuse prior: $\pi(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{p+1}{2}}$. This implies the following posterior distribution of $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$:

$$\boldsymbol{\mu_Y}|\boldsymbol{\Sigma_Y}, \mathbf{Y} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}},\ \boldsymbol{\Sigma_Y}/T\right), \tag{6}$$

$$\boldsymbol{\Sigma_Y}|\mathbf{Y} \sim \mathcal{W}^{-1}\left(T-1,\ \sum_{t=1}^{T}\left(\boldsymbol{Y_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}\right)\left(\boldsymbol{Y_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}\right)^{\top}\right), \tag{7}$$

where $\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}} \equiv \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Y_t}$ and $\mathcal{W}^{-1}$ is the inverse-Wishart distribution (a multivariate generalization of the inverse-gamma distribution). Note that the above posterior distribution is well defined even in the presence of weak factors, since the time-series layer does not depend on the strength of the factors or their tradability. Furthermore, the above posterior is analogous to the canonical t-distribution result for the parameters of a linear regression model.

The Normal-inverse-Wishart posterior in equations (6)–(7) implies that we can sample the distribution of the parameters $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$ by first drawing the covariance matrix $\boldsymbol{\Sigma_Y}$ from the inverse-Wishart distribution conditional on the data, and then by drawing $\boldsymbol{\mu_Y}$ from a multivariate normal distribution conditional on the data and the draw of $\boldsymbol{\Sigma_Y}$.

If the SDF is correctly specified, in the sense that all true factors are included, expected asset returns should be fully explained by their risk exposure, $\boldsymbol{C}$, and the prices of risk $\boldsymbol{\lambda}$, that is, $\boldsymbol{\mu_R} = \boldsymbol{C\lambda}$, where $\boldsymbol{\mu_R}$ is the sub-vector of $\boldsymbol{\mu_Y}$ corresponding to asset returns and $\boldsymbol{C}$ is the corresponding covariance sub-matrix of $\boldsymbol{\Sigma_Y}$. Therefore, we can define our first estimator.[7] In Appendix A.1.1 we show formally that it arises, under the assumption of correct specification, as a particular case of our general posterior presented in equations (11)-(12) below.

Definition 1 (Bayesian SDF (B-SDF) Estimates) *Conditional on $\boldsymbol{\mu_Y}$, $\boldsymbol{\Sigma_Y}$ and the data $\mathbf{Y} = \{\boldsymbol{Y_t}\}_{t=1}^{T}$, under the null of unique correct SDF specification*[8] *and any diffuse prior, the posterior distribution of $\boldsymbol{\lambda}$ is a Dirac distribution (that is, a constant) at $(\boldsymbol{C}^{\top}\boldsymbol{C})^{-1}\boldsymbol{C}^{\top}\boldsymbol{\mu_R}$. Therefore, conditional on only the data $\mathbf{Y} = \{\boldsymbol{Y_t}\}_{t=1}^{T}$ and the null, the posterior distribution of $\boldsymbol{\lambda}$ can be sampled by drawing $\boldsymbol{\mu}_{Y,(j)}$ and $\boldsymbol{\Sigma}_{Y,(j)}$ from the Normal-inverse-Wishart (6)–(7) and computing the draw $\boldsymbol{\lambda}_{(j)} \equiv \left(\boldsymbol{C}_{(j)}^{\top}\boldsymbol{C}_{(j)}\right)^{-1}\boldsymbol{C}_{(j)}^{\top}\boldsymbol{\mu}_{R,(j)}$.*

The posterior distribution of $\boldsymbol{\lambda}$, defined above, accounts for both the uncertainty about expected returns – via the sampling of $\boldsymbol{\mu_R}$ – and the uncertainty about the factor loadings – via the sampling of $\boldsymbol{C_f}$. Note that for completeness in the above we have allowed for a common

[7]The B-SDF estimator, and its GLS version, as shown in Appendix A.1.1, are particular cases of the more general posterior characterizations in equations (11)–(12) and (13)–(14). For expositional purposes we focus on the particular OLS- and GLS-like Bayesian estimators. Nevertheless, for any conformable matrix $\boldsymbol{A}$ such that $\boldsymbol{AC}$ is invertible, we have that under the null of unique correct specification, $\boldsymbol{\lambda}$ has (under any non-dogmatic prior) a degenerated posterior at $(\boldsymbol{AC})^{-1}\boldsymbol{A\mu_R}$ conditional on $\boldsymbol{A}$, $\boldsymbol{C}$, and $\boldsymbol{\mu_R}$.

[8]That is, $\boldsymbol{\mu_R} = \boldsymbol{C\lambda}$ holds for a unique value of $\boldsymbol{\lambda}$ as assumed in standard frequentist estimation.

cross-sectional intercept, λ_c. However, this can be readily constrained to be equal to zero, and we consider this case in our empirical analysis.

From the B-SDF definition, it is intuitive why we expect posterior inference to detect weak factors in finite sample. For such factors, the near singularity of $(\boldsymbol{C}_{(j)}^{\top}\boldsymbol{C}_{(j)})^{-1}$ will cause the draws for $\boldsymbol{\lambda}_{(j)}$ to diverge from zero (as in the frequentist point estimate). Nevertheless, the posterior uncertainty about factor loadings and asset risk premia will cause $\boldsymbol{C}_{(j)}^{\top}\boldsymbol{\mu}_{R,(j)}$ 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 risk prices $\boldsymbol{\lambda}$, 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

$$R^2_{ols} = 1 - \frac{(\boldsymbol{\mu_R} - \boldsymbol{C\lambda})^{\top}(\boldsymbol{\mu_R} - \boldsymbol{C\lambda})}{(\boldsymbol{\mu_R} - \bar{\mu}_R\boldsymbol{1_N})^{\top}(\boldsymbol{\mu_R} - \bar{\mu}_R\boldsymbol{1_N})}, \tag{8}$$

where $\bar{\mu}_R = \frac{1}{N}\sum_i^N \mu_{R,i}$. That is, for each posterior draw of $(\boldsymbol{\mu_R},\ \boldsymbol{C},\ \boldsymbol{\lambda})$, we can construct the corresponding draw for the R^2 from equation (8), hence, tracing out its posterior distribution. Equation (8) can be thought of as the population R^2, where $\boldsymbol{\mu_R}$, $\boldsymbol{C}$, and $\boldsymbol{\lambda}$ are unknown. After observing the data, we infer the posterior distribution of $\boldsymbol{\mu_R}$, $\boldsymbol{C}$, and $\boldsymbol{\lambda}$, and from these we can recover the distribution of the R^2.

Often the cross-sectional step of the frequentist estimation is performed via GLS rather than least squares. In our setting, under the null of the model, this leads to the following GLS estimator (see Appendix A.1.1 for a formal derivation).

Definition 2 (Bayesian SDF GLS (B-SDF-GLS)) *Conditional on $\boldsymbol{\mu_Y}$, $\boldsymbol{\Sigma_Y}$ and the data $\mathbf{Y} = \{\boldsymbol{Y_t}\}_{t=1}^T$, under the null of unique correct SDF specification and any diffuse prior, the posterior distribution of $\boldsymbol{\lambda}$ is a Dirac distribution (that is, a constant) at $(\boldsymbol{C}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{\mu_R}$. Therefore, conditional on <u>only</u> the data $\mathbf{Y} = \{\boldsymbol{Y_t}\}_{t=1}^T$ and the null, the posterior distribution of $\boldsymbol{\lambda}$ can be sampled by drawing $\boldsymbol{\mu}_{Y,(j)}$ and $\boldsymbol{\Sigma}_{Y,(j)}$ from the Normal-inverse-Wishart (6)–(7) and computing $\boldsymbol{\lambda}_{(j)} \equiv (\boldsymbol{C}_{(j)}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R},(j)}^{-1}\boldsymbol{C}_{(j)})^{-1}\boldsymbol{C}_{(j)}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R},(j)}^{-1}\boldsymbol{\mu}_{R,(j)}$.*

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

$$R^2_{gls} = 1 - \frac{(\boldsymbol{\mu_R} - \boldsymbol{C\lambda})^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{C\lambda})}{(\boldsymbol{\mu_R} - \bar{\mu}_R\boldsymbol{1_N})^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{\mu_R} - \bar{\mu}_R\boldsymbol{1_N})}. \tag{9}$$

Once again, we can think of equation (9) as the unknown population GLS R^2, which is a function of the unknown quantities $\boldsymbol{\mu_R}$, $\boldsymbol{C}$, and $\boldsymbol{\lambda}$. Since after observing the data we infer the posterior distribution of the parameters, we obtain the posterior distribution of the R^2_{gls} as well.

Realistically, models are rarely true. Therefore, we now allow for the presence of model-implied *average* pricing errors, $\boldsymbol{\alpha}$.[9] This can be easily accommodated within our Bayesian framework since in this case the data-generating process in the cross-section becomes $\boldsymbol{\mu_R} = \boldsymbol{C\lambda} + \boldsymbol{\alpha}$. Adding an assumption on the cross-sectional distribution of the pricing errors yields a Bayesian hierarchical structure to the estimation that naturally separates the time series and cross-sectional dimensions of the inference problem. To continue the analogy with OLS and GLS estimators, we consider two distributional assumptions for the average pricing errors $\boldsymbol{\alpha}$.

First, we consider the case of spherical cross-sectional errors, that is, $\alpha_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, in the spirit of OLS. Under this assumption, the cross-sectional likelihood function (i.e., conditional on the time-series parameters $\boldsymbol{\mu_R}$ and $\boldsymbol{C}$) is

$$p(data|\boldsymbol{\lambda}, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R} - \boldsymbol{C\lambda})^\top(\boldsymbol{\mu_R} - \boldsymbol{C\lambda})\right\}. \tag{10}$$

In the cross-sectional regression, the "data" are the expected risk premia, $\boldsymbol{\mu_R}$, and the factor loadings, $\boldsymbol{C}$. These quantities are not directly observable to the researcher but can be sampled from the Normal-inverse-Wishart posterior distribution in equations (6)–(7). Conceptually, this is not very different from the Bayesian modeling of latent variables. In the benchmark case, we assume a diffuse prior[10] 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

$$\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\mu_R}, \boldsymbol{C} \sim \mathcal{N}\left(\underbrace{(\boldsymbol{C}^\top\boldsymbol{C})^{-1}\boldsymbol{C}^\top\boldsymbol{\mu_R}}_{\widehat{\boldsymbol{\lambda}}}, \underbrace{\sigma^2(\boldsymbol{C}^\top\boldsymbol{C})^{-1}}_{\boldsymbol{\Sigma_\lambda}}\right) \text{ and} \tag{11}$$

$$\sigma^2|\boldsymbol{\mu_R}, \boldsymbol{C} \sim \mathcal{IG}\left(\frac{N-K-1}{2}, \frac{(\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})^\top(\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})}{2}\right), \tag{12}$$

where $\mathcal{IG}$ denotes the inverse-Gamma distribution. The conditional distribution in equation (11) makes it clear that the posterior takes into account both the uncertainty about prices of risk stemming from the time series parameters $\boldsymbol{C}$ and $\boldsymbol{\mu_R}$ (that are drawn from the Normal-inverse-Wishart posterior in equations (6)–(7)) and the random pricing errors $\boldsymbol{\alpha}$ that have the conditional posterior variance distribution given in equation (12). If test assets' expected excess returns are fully explained by $\boldsymbol{C}$, there are no pricing errors and $\sigma^2(\boldsymbol{C}^\top\boldsymbol{C})^{-1}$ converges to zero; otherwise, this layer of uncertainty always exists. Similarly, if one assumes that the cross-sectional model is correctly specified, that is, $\sigma^2 \to 0$, we are back to the B-SDF estimator

[9] As we show in the next section, this natural assumption is essential for model selection.

[10] As shown in the next subsection, in the presence of weak factors, such a 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.

in Definition 1.[11]

The OLS assumption ignores the fact that average pricing errors could be cross-sectionally correlated, which motivates our second, non-spherical, cross-sectional distributional assumption for $\boldsymbol{\alpha}$. Suppose that the model is correctly specified, that is, $\boldsymbol{R_t} = \lambda_c \mathbf{1_N} + \boldsymbol{C_f \lambda_f} + \boldsymbol{\epsilon_t}$, where $\boldsymbol{\epsilon_t} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0_N}, \boldsymbol{\Sigma_R})$. Since $\mathbb{E}_T[\boldsymbol{R_t}] = \lambda_c \mathbf{1_N} + \boldsymbol{C_f \lambda_f} + \mathbb{E}_T[\boldsymbol{\epsilon_t}]$, the pricing error $\boldsymbol{\alpha}$ should be equal to $\mathbb{E}_T[\boldsymbol{\epsilon_t}]$.[12] Hence, in the spirit of the central limit theorem, a natural distributional assumption for the pricing errors is $\boldsymbol{\alpha} \mid \boldsymbol{\Sigma_R} \sim \mathcal{N}(\mathbf{0_N}, \frac{1}{T}\boldsymbol{\Sigma_R})$. However, since we allow for mispricing, and its degree is endogeously determined by the observed data, a scaling of the covariance matrix is desirable. Therefore, we assign the following distributional assumption for $\boldsymbol{\alpha}$: $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0_N}, \sigma^2 \boldsymbol{\Sigma_R})$. We call this the GLS assumption. Recall that $\boldsymbol{\Sigma_R}$ is the covariance matrix of returns $\boldsymbol{R_t}$. Hence, the difference between the OLS and GLS assumption is that non-diagonal elements are non-zeros in the latter case. Since all models are misspecified to a certain degree, we would expect the estimated σ^2 to be larger than $1/T$.

The posterior distribution of $(\boldsymbol{\lambda}, \sigma^2)$ under the GLS distributional assumption, and conditional on $\boldsymbol{\mu_R}$, $\boldsymbol{\Sigma_R}$ and $\boldsymbol{C}$, is then (see derivation in Appendix A.1.1)

$$\boldsymbol{\lambda} | \sigma^2, \boldsymbol{\mu_R}, \boldsymbol{C}, \boldsymbol{\Sigma_R} \sim \mathcal{N}\Bigg(\underbrace{(\boldsymbol{C}^\top \boldsymbol{\Sigma_R^{-1}} \boldsymbol{C})^{-1} \boldsymbol{C}^\top \boldsymbol{\Sigma_R^{-1}} \boldsymbol{\mu_R}}_{\widehat{\boldsymbol{\lambda}}}, \underbrace{\sigma^2 (\boldsymbol{C}^\top \boldsymbol{\Sigma_R^{-1}} \boldsymbol{C})^{-1}}_{\boldsymbol{\Sigma_\lambda}}\Bigg) \text{ and} \tag{13}$$

$$\sigma^2 | \boldsymbol{\mu_R}, \boldsymbol{C}, \boldsymbol{\Sigma_R} \sim \mathcal{IG}\left(\frac{N-K-1}{2}, \frac{(\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})^\top \boldsymbol{\Sigma_R^{-1}} (\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})}{2}\right). \tag{14}$$

And once again $\boldsymbol{\mu_R}$, $\boldsymbol{\Sigma_R}$, and $\boldsymbol{C}$ can be sampled from the the Normal-inverse-Wishart posterior in equations (6)–(7). Furthermore, as before, by setting $\sigma^2 \to 0$ we recover the B-SDF-GLS in Definition 2.

Remark 1 (Generated factors) *Often factors are estimated, as, for example, 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, thanks to their hierarchical structure, it is relatively simple to adjust the above-defined Bayesian estimators 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*

[11] When pricing errors $\boldsymbol{\alpha}$ are assumed to be exactly zero under the null, the posterior distribution of $\boldsymbol{\lambda}$ in equation (11) collapses to a degenerate distribution, where $\boldsymbol{\lambda}$ equals $(\boldsymbol{C}^\top \boldsymbol{C})^{-1} \boldsymbol{C}^\top \boldsymbol{\mu_R}$ with probability one.

[12] Where $\mathbb{E}_T$ is the sample analog of the unconditional expectation operator.

which the PCs are constructed follows an inverse-Wishart distribution.[13] *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 above.*

Note that while we focus on the case of linear SDF models, our method can be easily extended to the estimation of beta representations of the fundamental pricing equation used in the two-pass procedure, such as Fama-MacBeth regressions (see Bryzgalova, Huang, and Julliard (2022)).

III.1 Model Selection and Aggregation

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

However, given the plethora of risk factors that have been proposed in the literature, a robust approach for model selection, across not necessarily nested models, that can handle a very large universe 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 hypotheses within the Bayesian framework is through Bayes factors and posterior probabilities, which 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 to the presence of 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 on the time series), that is, on the basis of the risk prices that the factors command.

Our approach hinges upon the introduction of suitable and economically driven priors that deliver valid marginal likelihoods and posterior model probabilities. With valid posterior probabilities, our framework allows to also *aggregate* multiple candidate factors and specifications into the most likely, given the data, representation of the true unknown SDF (via BMA).[14] Hence, our method endogenously selects a dominant subset of factors – if such a set exists uniquely – and instead aggregates factors optimally, if no dominant low-dimensional representation arises. But, unlike the canonical dichotomy of observable factors selection versus pure

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

[14]See, e.g., Raftery, Madigan, and Hoeting (1997), and Hoeting, Madigan, Raftery, and Volinsky (1999).

aggregation (e.g., principal component and entropy methods), our approach combines both. In a sense, it jointly delivers model selection and "smart" latent factor extraction.

In this subsection, we show first that flat priors for risk prices are not suitable for model selection in the presence of weak 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 weak factors. These priors allow us to test hypotheses using *valid* Bayes factors and model probabilities. Furthermore, they are particularly powerful in high-dimensional model selection, that is, when one wants, as in our empirical application, to consider all the factors in the zoo. Finally, we show how, as a by-product of the estimation and selection method, factors and models can be optimally aggregated.

III.1.1 Pitfalls of Flat Priors for Risk Prices

We start this section by discussing why flat priors for prices of risk are not suitable for 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 flat (in the sense of Jeffreys) priors for the time-series parameters $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$.

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 representation $\boldsymbol{\mu_R} = \boldsymbol{C}\boldsymbol{\lambda}+\boldsymbol{\alpha}$. For illustrative purposes, we focus on spherical cross-sectional errors (i.e., the case analogous to the GMM-OLS). Nevertheless, all the results in this and following subsections are also generalized to the non-spherical error setting (i.e., the case analogous to the GMM-GLS).

To model variable selection, we introduce a vector of binary latent variables $\boldsymbol{\gamma}^\top = (\gamma_0, \gamma_1, \dots, \gamma_K)$, where $\gamma_j \in \{0, 1\}$. When $\gamma_j = 1$, factor j (with associated loadings $\boldsymbol{C_j}$) should be included into the model and vice versa. Therefore, the number of included factors is $p_\gamma \equiv \sum_{j=0}^{K} \gamma_j$. Note that we always include the intercept, that is, $\gamma_0 = 1$ always. The notation $\boldsymbol{C_\gamma} = [\boldsymbol{C}_j]_{\gamma_j=1}$ represents a p_γ-columns sub-matrix of $\boldsymbol{C}$.

When testing whether the risk price 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 weak factors, risk prices are not identified; hence, testing whether they are equal to any given value is problematic per se. Nevertheless, as we show in the next section, with appropriate priors, whether a factor should be included or not is a well-defined question even in the presence of weak 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 $\boldsymbol{\mu_Y}$, $\boldsymbol{\Sigma_Y}$ and σ^2, 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 prices.

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

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

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

Therefore, if model $\boldsymbol{\gamma}$ includes a weak factor (whose $\boldsymbol{C_j}$ asymptotically converges to zero), the matrix $\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{C}_{\boldsymbol{\gamma}}$ 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 weak factor goes to one.[15] Consequently, under a flat prior for risk prices, the model containing a weak factor will always be selected asymptotically. However, the posterior distribution of $\boldsymbol{\lambda}$ for the weak 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 the null H_0, 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 are even more unlikely under H_1. Therefore, a disperse prior for λ_j may push the posterior probabilities to favor H_0 and make it fail to identify true factors.[16]

Note also that flat (hence improper) priors for risk prices are not appropriate, since 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 the data, $p(data|H_0)$ and $p(data|H_1)$, are both undetermined, we cannot define the Bayes'

[15]Note that a similar problem also arises when using mimicking portfolios of weak factors. In this case the singularity in the determinant in equation (15) would be generated by the projection of the non-tradable factors on the space of returns.

[16]This phenomenon is known as the Bartlett Paradox (see Bartlett (1957)).

factor $\frac{p(data|H_1)}{p(data|H_0)}$ (as stressed in, 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.[17] Similarly, we can still assign improper priors for $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ in the first time-series step.

The final reason why it might be undesirable to use a flat prior for risk prices 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 prices that is both robust to weak factors and can be used for model selection, even when dealing with a very large number of potential models.

III.1.2 Spike-and-Slab Prior for Risk Prices

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

This prior belongs to the so-called *spike-and-slab* family. For illustrative purposes, in this section we consider a Dirac spike and show analytically its implications for model selection. In the next subsection we generalize the method 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, $\pi(\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma}) \propto \pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2)\pi(\boldsymbol{\gamma})$. When $\gamma_j = 1$, and, hence, the factor should be included in the model, the prior (the "slab") 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 (crucial) quantity that we define below. When instead $\gamma_j = 0$, and the corresponding risk factor should not be included in the model, the prior (the "spike") is a Dirac distribution at zero – since, if the factor is not a part of the SDF, its price of risk should be zero.[18] For the cross-sectional variance of the pricing errors we keep the canonical diffuse prior:[19] $\pi(\sigma^2) \propto \sigma^{-2}$.

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 $\boldsymbol{\gamma}$, where c is a small positive number corresponding to the common

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

[18]Obviously, this does not imply that the risk premium on the factor should be zero, since the factor might correlate with the true sources of risk.

[19]Note that since the parameter σ is common across models and has the same support in each model, the marginal likelihoods obtained under this improper prior are valid and comparable (see Proposition 1 of Chib, Zeng, and Zhao (2020)).

cross-sectional intercept (λ_c). The prior for the prices of risk ($\boldsymbol{\lambda_\gamma}$) of model $\boldsymbol{\gamma}$ is then

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

Given this prior, we sample the posterior distribution by sequentially drawing from the conditional distributions of the parameters (i.e., we use a Gibbs sampling approach)[20] presented in the following proposition.

Proposition 2 (B-SDF OLS Posterior with Dirac Spike-and-Slab) *The posterior distribution of $(\lambda_\gamma, \sigma^2, \boldsymbol{\gamma})$ under the assumption of Dirac spike-and-slab prior and spherical $\boldsymbol{\alpha}$ (OLS), conditional on the draws of $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ from equations (6)–(7), is given by the following conditional distributions:*

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

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

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

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

Proposition 3 (B-SDF GLS Posterior with Dirac Spike-and-Slab) *The posterior distribution of $(\lambda_\gamma, \sigma^2, \boldsymbol{\gamma})$ under the assumption of Dirac spike-and-slab prior and non-spherical $\boldsymbol{\alpha}$ (GLS), conditional on the draws of $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ from equations (6)–(7), is given by the following conditional distributions:*

$$\boldsymbol{\lambda_\gamma}|data, \sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}_\gamma, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma)\right), \tag{19}$$

$$\sigma^2|data, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N}{2}, \frac{SSR_\gamma}{2}\right), \text{ and} \tag{20}$$

$$p(\boldsymbol{\gamma} \mid data) \propto \frac{|\boldsymbol{D}_\gamma|^{\frac{1}{2}}}{|\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma|^{\frac{1}{2}}} \frac{1}{(SSR_\gamma/2)^{\frac{N}{2}}}, \tag{21}$$

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

[20] We do not standardize $\boldsymbol{Y_t}$ in the time-series regression. In the empirical implementation, after obtaining posterior draws for $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$, we calculate $\boldsymbol{\mu_R}$ and $\boldsymbol{C_f}$ as the standardized expected returns of test assets and correlation between test assets and factors. Then $\boldsymbol{C}$ is a matrix containing a vector of ones and $\boldsymbol{C_f}$.

The above propositions are proved, respectively, in Appendices A.1.3 and A.1.4.

Note that SSR_γ is the minimized sum of squared errors under the spherical pricing errors assumption, and is instead the minimized squared Sharpe ratio of pricing errors in the non-spherical case, where the term $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}^\top \boldsymbol{D}_{\boldsymbol{\gamma}} \boldsymbol{\lambda}_{\boldsymbol{\gamma}}$ is akin to a generalized ridge regression penalty.

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 setting, arises in the presence of weak factors).

The key element and novelty of our method is that the "shrinkage" applied to the factors is endogenously heterogeneous and designed to target weak factors: It leverages the correlation between factors and returns by setting ψ_j as

$$\psi_j = \psi \times \boldsymbol{\rho}_{\boldsymbol{j}}^\top \boldsymbol{\rho}_{\boldsymbol{j}}, \tag{22}$$

where $\boldsymbol{\rho}_{\boldsymbol{j}}$ 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 degree of shrinkage over all factors.[21] But, unlike tuning parameters in frequentist inference, as we show below, ψ is uniquely pinned down by the researcher's beliefs about Sharpe ratios being achievable in the economy.

When the correlation between f_{jt} and $\boldsymbol{R_t}$ is very low, as in the case of a weak factor, the penalty for λ_j, which is the reciprocal of $\psi \boldsymbol{\rho}_{\boldsymbol{j}}^\top \boldsymbol{\rho}_{\boldsymbol{j}} \equiv (\{\boldsymbol{D}_{\boldsymbol{\gamma}}\}_{jj})^{-1}$, is very large and dominates the sum of squared errors.

Equation (16) (and, similarly, equation (19)) makes clear why this Bayesian formulation is robust to weak factors. When $\boldsymbol{C}$ converges to zero, $(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{C}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})$ is dominated by $\boldsymbol{D}_{\boldsymbol{\gamma}}$, so the identification condition for the prices of risk no longer fails. When a factor is weak, 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 $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}} = (\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{C}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\mu}_{\boldsymbol{R}}$, 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 weak 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 $\boldsymbol{C}$ dominates the prior information, since in this case the absolute size of $\boldsymbol{D}_{\boldsymbol{\gamma}}$ is small relative to $\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{C}_{\boldsymbol{\gamma}}$.

Remark 4 (Level Factors) *Identification failure of factors' risk prices can arise in the presence of "level factors," that is factors to which asset returns have non-zero exposure but lack cross-sectional spread. These factors help explain the average level of returns but not their cross-sectional dispersion, and, hence, are collinear with the common cross-sectional intercept. Our*

[21] Alternatively, we could have set $\psi_j = \psi \times \boldsymbol{C}_{\boldsymbol{j}}^\top \boldsymbol{C}_{\boldsymbol{j}}$, where $\boldsymbol{C}_{\boldsymbol{j}}$ is a $N \times 1$ vector of covariances of the test assets with factor j. However, $\boldsymbol{\rho}_{\boldsymbol{j}}$ has the advantage of being invariant to the units in which factors are measured. Furthermore, in the empirical analysis the cross-sectional step is implemented using returns and factors scaled by their standard deviations, making the distinction immaterial.

approach can handle this case by using variance standardized variables in the cross-sectional part of the estimation and replacing the penalty in (22) with

$$\psi_j = \psi \times \widetilde{\boldsymbol{\rho}}_j^\top \widetilde{\boldsymbol{\rho}}_j, \tag{23}$$

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

When comparing two models, using posterior model probabilities for specification selection 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 shows that, unlike in the flat prior case discussed earlier, under the Dirac spike, the Bayes factors (and posterior probabilities) are well-defined even in the presence of weak factors.[22] Therefore, they can be used for model selection and hypotheses testing.

Corollary 1 (Model 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* $K \times 1$ *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.*

Under the spherical assumption for $\boldsymbol{\alpha}$ *(OLS), the Bayes factor is*

$$BF_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} = \left(\frac{SSR_{\gamma'}}{SSR_{\gamma}}\right)^{\frac{N}{2}} \left|1 + \psi_p \boldsymbol{C}_{\boldsymbol{p}}^\top \left[\boldsymbol{I}_{\boldsymbol{N}} - \boldsymbol{C}_{\gamma'}(\boldsymbol{C}_{\gamma'}^\top \boldsymbol{C}_{\gamma'} + \boldsymbol{D}_{\gamma'})^{-1}\boldsymbol{C}_{\gamma'}^\top\right] \boldsymbol{C}_{\boldsymbol{p}}\right|^{-\frac{1}{2}}, \tag{24}$$

where $SSR_\gamma = \boldsymbol{\mu}_{\boldsymbol{R}}^\top \boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{\mu}_{\boldsymbol{R}}^\top \boldsymbol{C}_\gamma(\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1}\boldsymbol{C}_\gamma^\top \boldsymbol{\mu}_{\boldsymbol{R}} = \min_{\boldsymbol{\lambda}_\gamma}\{(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma)^\top(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma) + \boldsymbol{\lambda}_\gamma^\top \boldsymbol{D}_\gamma \boldsymbol{\lambda}_\gamma\}$*. Under the non-spherical assumption for* $\boldsymbol{\alpha}$ *(GLS), the Bayes factor is*

$$BF_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} = \left(\frac{SSR_{\gamma'}}{SSR_{\gamma}}\right)^{\frac{N}{2}} \left|1 + \psi_p\left[\boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}} - \boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'}\left(\boldsymbol{C}_{\gamma'}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'} + \boldsymbol{D}_{\gamma'}\right)^{-1}\boldsymbol{C}_{\gamma'}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}}\right]\right|^{-\frac{1}{2}}. \tag{25}$$

where $SSR_\gamma = \boldsymbol{\mu}_{\boldsymbol{R}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{\mu}_{\boldsymbol{R}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_\gamma(\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1}\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{\mu}_{\boldsymbol{R}} = \min_{\boldsymbol{\lambda}_\gamma}\{(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma)^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma) + \boldsymbol{\lambda}_\gamma^\top \boldsymbol{D}_\gamma \boldsymbol{\lambda}_\gamma\}$.

The proof can be found in Appendix A.1.5.

[22]The corollary can be trivially extended to the case of different prior probabilities for the two models, since in this case the Bayes factor is simply the ratio of marginal likelihoods multiplied by the prior odds.

Since $\boldsymbol{C}_{\boldsymbol{p}}^{\top}[\boldsymbol{I}_{\boldsymbol{N}} - \boldsymbol{C}_{\boldsymbol{\gamma'}}(\boldsymbol{C}_{\boldsymbol{\gamma'}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma'}} + \boldsymbol{D}_{\boldsymbol{\gamma'}})^{-1}\boldsymbol{C}_{\boldsymbol{\gamma'}}^{\top}]\boldsymbol{C}_{\boldsymbol{p}}$ 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 (24) dominates, and the Bayes factor will be much greater than 1, hence, providing evidence in favor of model $\boldsymbol{\gamma}$.

Recall that $SSR_\gamma = \min_{\boldsymbol{\lambda}_\gamma}\{(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma\boldsymbol{\lambda}_\gamma)^\top(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma\boldsymbol{\lambda}_\gamma) + \boldsymbol{\lambda}_\gamma^\top\boldsymbol{D}_\gamma\boldsymbol{\lambda}_\gamma\}$, 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, 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 (24), lowering the Bayes factor, and acting as a penalty for dimensionality.

A particularly interesting case is when the factor added by model $\boldsymbol{\gamma}$ is weak: $\boldsymbol{C}_{\boldsymbol{p}}$ converges to zero, but the penalty term $1/\psi_p \propto 1/\boldsymbol{\rho}_{\boldsymbol{p}}^\top\boldsymbol{\rho}_{\boldsymbol{p}}$ goes to infinity. On the one hand, the first term in equation (24) will converge to 1; on the other hand, since $\psi_p \approx 0$ in large sample, the second term in equation (24) will also be around 1. Therefore, the Bayes factor for a weak factor will go to 1 asymptotically.[23] 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 prices of risk by centering the Dirac spike at the null value rather than at zero.

Corollary 2 (Hypothesis Testing for Risk Prices (Bayesian p-values)) *Suppose that we want to test the point hypothesis $\boldsymbol{\lambda}_{-\gamma} = \tilde{\boldsymbol{\lambda}}_{-\gamma}$ and as before we have the prior $\boldsymbol{\lambda}_\gamma|\sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2\boldsymbol{D}_\gamma^{-1})$ in model $\boldsymbol{\gamma}$. In this case, the posterior distributions in Propositions 2 and 3 still hold with $SSR_{\boldsymbol{\gamma}}$ therein replaced by the $\widetilde{SSR}_{\boldsymbol{\gamma}}$ defined below.*

Under the spherical assumption for $\boldsymbol{\alpha}$ (OLS),

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

where $\tilde{\boldsymbol{\mu}}_{\boldsymbol{R}} \equiv \boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma}$ denotes the vector of cross-sectional residual expected returns that are unexplained by factors $f_{-\gamma}$ with prices of risk $\tilde{\boldsymbol{\lambda}}_{-\gamma}$.

[23] 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 $H_0 : \gamma_p = 0$.

Under the non-spherical assumption for $\boldsymbol{\alpha}$ (GLS),

$$\begin{aligned}\widetilde{SSR}_\gamma &= (\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma})^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma}) - \\ &\quad (\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma})^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_\gamma(\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1}\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma}) \\ &= \min_{\boldsymbol{\lambda}_\gamma}\{(\tilde{\boldsymbol{\mu}}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma\boldsymbol{\lambda}_\gamma)^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\tilde{\boldsymbol{\mu}}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma\boldsymbol{\lambda}_\gamma) + \boldsymbol{\lambda}_\gamma^\top \boldsymbol{D}_\gamma\boldsymbol{\lambda}_\gamma\},\end{aligned}$$

A Bayesian p-value for the null hypothesis is then constructed by integrating $1 - p(\boldsymbol{\gamma} \mid data)$ in equation (18) (equation (21) in the case of spherical (non-spherical) pricing errors), with respect to the Normal-inverse-Wishart in equations (6)–(7).

The proof of the corollary follows the same steps as the proofs of Propositions 2 and 3 in Appendices A.1.3 and A.1.4.

Corollary 2 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.

III.1.3 Continuous Spike

We extend the Dirac spike-and-slab prior by encoding a continuous spike for λ_j, when γ_j equals 0. While the closed-form solutions obtained with a Dirac spike allow to feasibly evaluate *millions* of models, this extension allows to efficiently sample *quadrillions* of alternative specifications.

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 $\pi(\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega}) = \pi(\boldsymbol{\lambda} \mid \sigma^2, \boldsymbol{\gamma})\pi(\sigma^2)\pi(\boldsymbol{\gamma} \mid \boldsymbol{\omega})\pi(\boldsymbol{\omega})$, where

$$\lambda_j \mid \gamma_j, \sigma^2 \sim \mathcal{N}\left(0, r(\gamma_j)\psi_j\sigma^2\right). \tag{26}$$

Note the introduction of two new elements, $r(\gamma_j)$ and $\pi(\boldsymbol{\omega})$, in the prior. When the factor should be included, $r(\gamma_j = 1) = 1$, hence we have the same "slab" as before. When the factor should not be in the model $r(\gamma_j = 0) = r \ll 1$. Hence the Dirac "spike" is replaced by a Gaussian spike, which is extremely concentrated at zero (we set $r = 0.001$ in our empirical analysis). Note that in this case ψ_j has an effect on the spike, but given a small value for r this effect is virtually immaterial. As we explain below, the additional prior $\pi(\boldsymbol{\omega})$ encodes our ex ante beliefs about the sparsity of the true model in terms of observable factors.

We now 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 (22). In matrix notation, the prior for $\boldsymbol{\lambda}$ is therefore: $\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2\boldsymbol{D}^{-1})$. 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. Hence, the spike component concentrates the posterior mass of $\boldsymbol{\lambda}$ around 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 III.1.2.

Furthermore, this prior encodes beliefs about the fraction of the total Sharpe ratio of the test assets ascribable to the factors and to the pricing errors. To see this, consider the case in which (as in our empirical applications) both factors and returns are standardized. It then follows that

$$\frac{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \boldsymbol{\gamma}, \sigma^2]}{\mathbb{E}_\pi[SR^2_{\boldsymbol{\alpha}} \mid \sigma^2]} = \frac{\sum_{k=1}^K r(\gamma_k)\psi_k}{N} = \frac{\psi \sum_{k=1}^K r(\gamma_k) \tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k}{N}, \tag{27}$$

where $SR_{\boldsymbol{f}}$ and $SR_{\boldsymbol{\alpha}}$ denote, respectively, the Sharpe ratios of all factors[24] ($\boldsymbol{f_t}$) 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, $\sum_{k=1}^K \tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k / N \simeq 3.22$.Hence, for ψ in the 1–5 range, if, say, 50% of the factors are selected, our prior expectation is that the factors should explain about 62%–89% of the squared Sharpe ratio of test assets.

The prior $\pi(\boldsymbol{\omega})$ not only gives us a way of sampling from the space of potential models, but also encodes belief about the sparsity of the true model using the prior distribution $\pi(\gamma_j = 1|\omega_j) = \omega_j$. Following the literature on predictors selection, we set:

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

Different hyper-parameters a_ω and b_ω determine whether one a priori favors more parsimonious models or not.[25] 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, and iii) the sparsity of the "true" model, *uniquely* determine the parameters ψ, a_ω, b_ω.[26]

Potentially, this prior specification could be improved along two dimensions. First, we do not formally rule out all near-arbitrage opportunities – hence, we potentially leave on the table some performance improvement that could have been achieved by exploiting such an economic constraint. Second, the prior does not make ex ante use of the covariance structure between factors (but our posterior does): that is, equally strong variables are treated identically by the

[24]The squared Sharpe ratio implied by the SDF is $\boldsymbol{\lambda}_{\boldsymbol{f}}^\top \boldsymbol{\Sigma}_{\boldsymbol{f}} \boldsymbol{\lambda}_{\boldsymbol{f}}$. Since $\boldsymbol{\lambda}_{\boldsymbol{f}}$ are assumed to be independently distributed in the prior level, $\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \boldsymbol{\gamma}, \sigma^2]$ is equal to $\sum_{k=1}^K \mathbb{E}_\pi[\lambda_k^2 \mid \gamma_k, \sigma^2]$.

[25]The prior expected probability of selecting a factor is simply $\frac{a_\omega}{a_\omega + b_\omega}$. We set $a_\omega = b_\omega = 1$ in the benchmark case; that is, each factor has an ex ante expected probability of being selected equal to 50%. However, we could for instance, set $a_\omega = 1$ and $b_\omega >> 1$ in order to favor a sparser model.

[26]For a discussion on the importance of using priors on observables and economic quantities, rather than deep model parameters, see Jarociński and Marcet (2019).

prior, irrespectively of their covariance structure. In principle, one could modify the prior to be over the space of groups of factors rather than individual factors themselves.

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 Dirac spike in Section III.1.2. Instead, treating ω_j (hence, γ_j), as a parameter to be sampled is particularly useful in high-dimensional cases. For instance, suppose that there are 30 candidate factors. With the Dirac spike-and-slab prior we have to calculate the posterior model probabilities for 2^{30} different models. Given that we update $(\boldsymbol{\mu_R}, \boldsymbol{C_f})$ at each sampling round, posterior probabilities for all models are re-computed for every new draw of these quantities, rendering the computational cost very large. In contrast, with the continuous spike-and-slab approach one can simply use the posterior mean of γ_j to estimate the posterior marginal probability of the j-th factor, since they are the same quantity.

Similar to the Dirac spike-and-slab case, we use sequential sampling from the conditional distributions of the parameters $(\boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2)$ and, most importantly, $\boldsymbol{\gamma}$, as presented in the following propositions.

Proposition 5 (B-SDF OLS Posterior 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 and spherical $\boldsymbol{\alpha}$ (OLS), conditional on the draws of $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ from equations (6)–(7), is given by the following conditional distributions:*

$$\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})\right), \tag{28}$$

$$\frac{p(\gamma_j = 1|data, \boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2, \boldsymbol{\gamma_{-j}})}{p(\gamma_j = 0|data, \boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2, \boldsymbol{\gamma_{-j}})} = \frac{\omega_j}{1-\omega_j}\frac{p(\lambda_j|\gamma_j = 1, \sigma^2)}{p(\lambda_j|\gamma_j = 0, \sigma^2)}, \tag{29}$$

$$\omega_j|data, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^2 \sim Beta\left(\gamma_j + a_\omega, 1 - \gamma_j + b_\omega\right), \ and \tag{30}$$

$$\sigma^2|data, \boldsymbol{\omega}, \boldsymbol{\lambda}, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N+K+1}{2}, \frac{(\boldsymbol{\mu_R} - \boldsymbol{C}\boldsymbol{\lambda})^\top(\boldsymbol{\mu_R} - \boldsymbol{C}\boldsymbol{\lambda}) + \boldsymbol{\lambda}^\top\boldsymbol{D}\boldsymbol{\lambda}}{2}\right), \tag{31}$$

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

Proposition 6 (B-SDF GLS Posterior 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 and non-spherical $\boldsymbol{\alpha}$ (GLS), conditional on the draws of $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ from equations (6)–(7), differs from the one in Proposition 5 only for the posterior distributions of $(\boldsymbol{\lambda}, \sigma^2)$:*

$$\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})\right), \text{ and} \tag{32}$$

$$\sigma^2|data, \boldsymbol{\omega}, \boldsymbol{\lambda}, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N+K+1}{2}, \frac{(\boldsymbol{\mu_R} - \boldsymbol{C}\boldsymbol{\lambda})^\top\boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R} - \boldsymbol{C}\boldsymbol{\lambda}) + \boldsymbol{\lambda}^\top\boldsymbol{D}\boldsymbol{\lambda}}{2}\right), \tag{33}$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{C}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{C} + \boldsymbol{D})^{-1} \boldsymbol{C}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{\mu}_{\boldsymbol{R}}$ *and* $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}) = \sigma^2 (\boldsymbol{C}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{C} + \boldsymbol{D})^{-1}$.

The proofs of the above propositions are reported in Appendix A.1.6.

III.1.4 Selection vs. Aggregation

The posterior probabilities of models and factors obtained above with spike-and-slab priors, can be used not only for model selection but also efficient aggregation using *all* possible specification.

If we are interested in some quantity Δ that is well-defined for every model $m = 1, ..., \bar{m}$ (e.g., price of risk, risk premia, and maximum Sharpe ratio), from the Bayes theorem we have

$$\mathbb{E}\left[\Delta|\text{data}\right] = \sum_{m=0}^{\bar{m}} \mathbb{E}\left[\Delta|\text{data}, \text{model} = m\right] \Pr\left(\text{model} = m|\text{data}\right), \tag{34}$$

where $\mathbb{E}\left[\Delta|\text{data}, \text{model} = m\right] = \lim_{L\to\infty} \frac{1}{L}\sum_{l=1}^{L} \Delta(\theta_l^{(m)})$ 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), and Hoeting, Madigan, Raftery, and Volinsky (1999)).

The BMA efficiently aggregates information about Δ over the space of all models, rather than conditioning on a particular model. At the same time, *if* a dominant model exists – hence it has posterior probability approaching one – the BMA will use that model alone.

For each model $\boldsymbol{\gamma}$ that one could construct with the universe of factors, we have the corresponding SDF: $M_{\boldsymbol{\gamma},t} = 1 - (\boldsymbol{f}_{\boldsymbol{\gamma},t} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma},t}])^\top \boldsymbol{\lambda}_{\boldsymbol{\gamma}}$. Therefore, one can construct a BMA of the SDF using the model posterior probabilities derived in the previous sections. Note that these probabilities are based upon the ability of the factors and models to explain the cross-section of asset return; that is, they explicitly target the key property of a valid SDF. Aggregation is particularly appealing when multiple candidate factors load on the same underlying sources of risk (plus factor-specific noise). Crucially, BMA creates a weighted average that endogenously maximizes the SDF signal-to-noise ratio for cross-sectional pricing.

The BMA *is* the optimal aggregation procedure for a very wide spectrum of optimality criteria and, in particular, it is optimal under the quadratic loss function and is "optimal on average", that is, no alternative estimator can beat the BMA for all values of the true unknown parameters (see, e.g., Raftery and Zheng (2003), and Schervish (1995)). Furthermore, the BMA predictive distribution minimizes the Kullback-Leibler information divergence relative to the true unknown data generating process. Hence, it delivers the *most likely* SDF given the data, and the estimated density is as close as possible to the true unknown one, even if all the models considered are misspecified.

A powerful feature of the BMA method is that equation (34) can be evaluated by generating a Markov Chain over the space of possible models. This is exactly what the continuous spike-and-slab method allows us to do: We sample models in the unrestricted space of 2.25 quadrillion specifications, computing all the desired quantities of interest for each specification sampled, and then aggregate the results. The Markov Chain endogenously over-samples the more likely specifications and under-samples the ones that are less likely to have generated the observed data. The Markov Chain can then be stopped when the posterior means of interest have converged according to the standard tests. We use as a convergence criterion the Separate Partial Mean test (see, e.g., Geweke (2005)) for *each* factor specific parameter (i.e., posterior probability and price of risk).

Recent literature has usually pursued either *selection* (see, e.g., Giglio, Feng, and Xiu (2020)) or *aggregation* (see, e.g., Kozak, Nagel, and Santosh (2020)) of pricing factors. Our approach, instead, combines *both*. The BMA-SDF includes both factors that are clear drivers of asset returns, that is, factors with posterior probability of inclusion ($\Pr[\gamma_j = 1|\text{data}]$) approaching 1, and also an optimal combination of factors that are, given the data, individually less salient.

IV Simulation

We build a simple setting for a linear factor model that includes both strong and weak 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 GMM-OLS estimation, which makes the vector of simulated expected returns equal to their sample mean estimates of 25 Fama-French portfolios. Finally, a useless factor is simulated from an independent normal distribution with mean zero and standard deviation 1%. In summary,

$$f_{t,useless} \overset{\text{iid}}{\sim} \mathcal{N}(0, (1\%)^2), \qquad \begin{pmatrix} \boldsymbol{R_t} \\ f_{t,hml} \end{pmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(\begin{bmatrix} \bar{\boldsymbol{\mu}}_{\boldsymbol{R}} \\ \bar{f}_{hml} \end{bmatrix}, \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{R}} & \widehat{\boldsymbol{C}}_{\boldsymbol{hml}} \\ \widehat{\boldsymbol{C}}_{\boldsymbol{hml}}^{\top} & \widehat{\sigma}^2_{hml} \end{bmatrix} \right), \text{ and}$$

$$\boldsymbol{\mu_R} = \begin{cases} \widehat{\lambda}_c \mathbf{1_N} + \widehat{\boldsymbol{C}}_{\boldsymbol{f}} \widehat{\lambda}_{HML}, & \text{if the model is correct, and} \\ \bar{\boldsymbol{R}}, & \text{if the model is misspecified,} \end{cases}$$

where factor loadings, risk prices, and variance-covariance matrix of returns and factors are equal to their sample estimates from the time series and cross-sectional regressions of the GMM-OLS 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 three estimation setups: (a) the model includes only a strong factor (HML), (b) the model includes only a useless factor as a stylized example for a weak 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, 1,000, and 20,000. We compare the performance of the OLS/GLS standard frequentist and Bayesian SDF estimators (GMM and B-SDF, respectively) with the focus on risk prices recovery, testing, and identification of strong and useless factors for model comparison.

IV.1 B-SDF Estimation of Risk Prices

In this section we focus on the most realistic (and challenging) model setup, which includes both useless and strong factors and allows for model misspecification. We found similar performance of the B-SDF approach in a wide range of alternative simulation settings (e.g., considering correctly specified models and cross-sections of different dimensions).[27]

Table 1 compares the performance of frequentist and Bayesian estimators of the price of risk and reports their empirical test size and confidence intervals for cross-sectional R^2. In the case of the Bayesian estimation we report results for both the flat and normal priors for the price of risk (the latter, in a single stand-alone model case, corresponds to the spike-and-slab approach). Since the model is misspecified, true cross-sectional R^2 has the population value of 43.87% (6.69%) for OLS (GLS)). In the case of the standard GMM approach, tests are constructed using standard t-statistics, and in the case of the B-SDF we rely on the quantiles of the posterior distribution to form the credible confidence intervals. The last two columns also report the quantiles of the posterior distribution of the R^2 mode across the simulations, corresponding to the peak of the cross-sectional likelihood.

As expected, in the conventional frequentist 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% (87%) of the simulations in the asymptotic case of $T = 20,000$. On the contrary, the Bayesian confidence intervals detect the useless factor and reject the null of zero price of risk attached to the useless factor with frequency asymptotically approaching the size of the tests independently from the prior.

The presence of useless factors can also bias parameter estimates for the strong ones and often leads to their *crowding out* from the model. Panel A in Table 1 serves as a good illustration of this possibility, with the GMM price of risk estimates for the strong factor clearly biased due to the weak identification problem. In this case B-SDF provides reliable, albeit conservative in the flat prior case, confidence bounds for model parameters, and effectively restores statistical

[27] These additional results are reported in Internet Appendix IA.A.1.

inference. Note that the empirical size of the B-SDF (normal prior) credible confidence intervals is very close to the nominal one even for relatively small sample sizes.

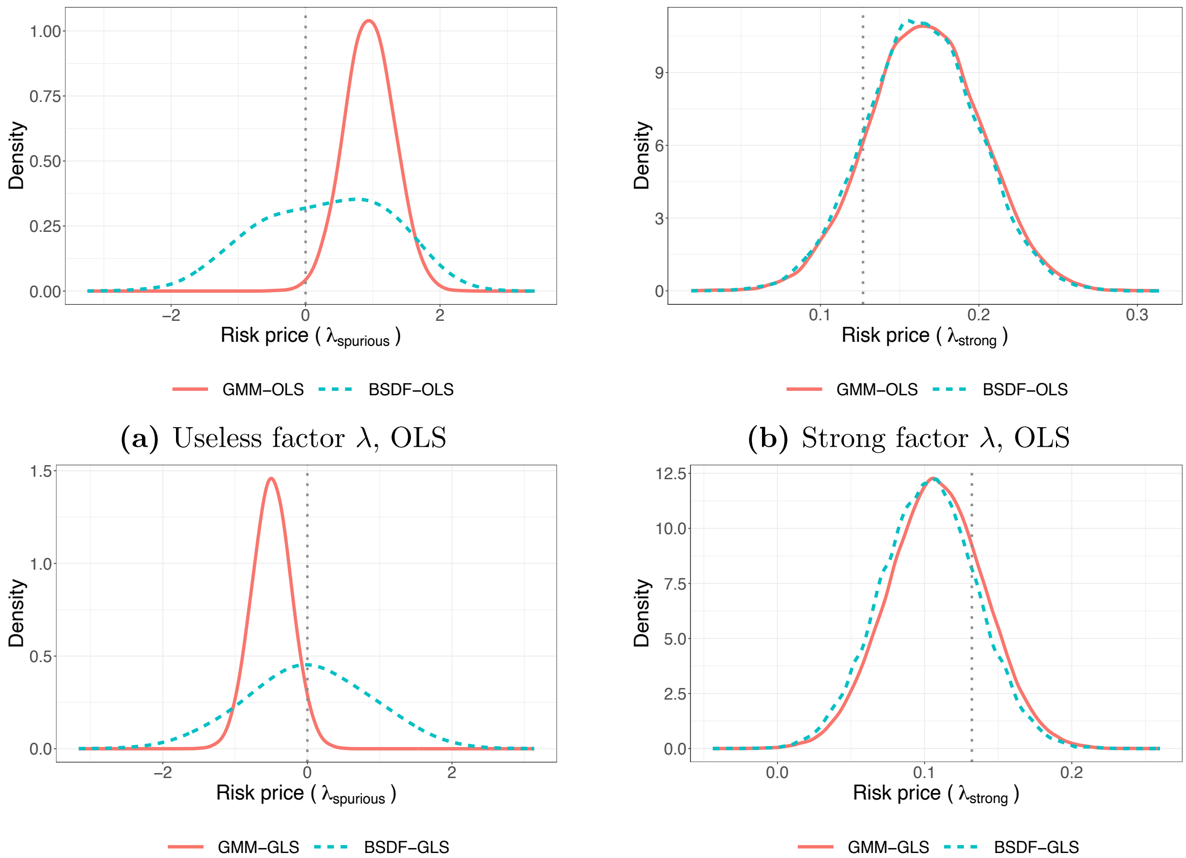
Table 1: Price of risk tests in a misspecified model with useless and strong factors

			λ_c			λ_{strong}			$\lambda_{useless}$		R^2_{adj}	
	T	10%	5%	1%	10%	5%	1%	10%	5%	1%	5th	95th
						Panel A: OLS						
GMM–$\boldsymbol{W_{ols}}$	100	0.083	0.033	0.007	0.065	0.03	0.005	0.082	0.029	0.004	-4.35%	70.21%
	200	0.084	0.039	0.006	0.058	0.025	0.003	0.119	0.047	0.006	-2.38%	69.17%
	600	0.075	0.034	0.009	0.074	0.032	0.005	0.255	0.140	0.024	8.42%	67.27%
	1000	0.078	0.03	0.004	0.070	0.031	0.001	0.311	0.181	0.048	16.85%	65.40%
	20000	0.066	0.019	0.001	0.052	0.022	0.001	0.752	0.585	0.288	36.92%	58.64%
B-SDF, flat prior	100	0.037	0.015	0.001	0.032	0.007	0.001	0.003	0.001	0.000	16.62%	49.24%
	200	0.054	0.021	0.002	0.036	0.013	0.001	0.006	0.001	0.000	13.54%	54.05%
	600	0.053	0.027	0.005	0.047	0.015	0.002	0.019	0.006	0.001	14.72%	58.72%
	1000	0.059	0.027	0.004	0.050	0.018	0.000	0.040	0.013	0.002	19.57%	58.85%
	20000	0.015	0.005	0.000	0.010	0.003	0.000	0.089	0.043	0.009	39.19%	52.86%
B-SDF, normal prior	100	0.062	0.029	0.005	0.047	0.019	0.002	0.003	0.001	0.000	7.47%	43.43%
	200	0.084	0.04	0.008	0.067	0.031	0.005	0.006	0.002	0.000	3.66%	48.19%
	600	0.087	0.048	0.018	0.093	0.044	0.010	0.019	0.006	0.001	4.87%	54.33%
	1000	0.094	0.052	0.011	0.106	0.051	0.010	0.040	0.013	0.002	9.64%	54.13%
	20000	0.100	0.050	0.011	0.102	0.052	0.009	0.088	0.043	0.009	34.47%	46.84%
						Panel B: GLS						
GMM–$\boldsymbol{W_{gls}}$	100	0.095	0.048	0.007	0.076	0.035	0.004	0.146	0.070	0.012	-7.66%	20.08%
	200	0.104	0.051	0.008	0.086	0.045	0.007	0.235	0.142	0.031	-6.97%	19.19%
	600	0.090	0.045	0.009	0.105	0.047	0.008	0.433	0.326	0.163	-4.81%	20.93%
	1000	0.096	0.044	0.010	0.106	0.054	0.008	0.535	0.444	0.273	-3.38%	19.52%
	20000	0.084	0.034	0.006	0.091	0.037	0.009	0.889	0.865	0.807	1.42%	19.32%
B-SDF, flat prior	100	0.114	0.061	0.011	0.046	0.020	0.001	0.029	0.009	0.000	-1.99%	9.64%
	200	0.094	0.050	0.012	0.056	0.023	0.003	0.034	0.012	0.001	-3.04%	10.27%
	600	0.090	0.045	0.008	0.066	0.028	0.004	0.068	0.029	0.004	-2.31%	12.68%
	1000	0.080	0.036	0.007	0.071	0.026	0.002	0.075	0.035	0.007	-1.10%	12.98%
	20000	0.017	0.002	0.000	0.013	0.004	0.002	0.105	0.050	0.011	3.43%	12.65%
B-SDF, normal prior	100	0.133	0.070	0.014	0.054	0.023	0.002	0.029	0.008	0.000	-3.50%	7.72%
	200	0.111	0.057	0.018	0.075	0.033	0.006	0.034	0.012	0.001	-5.08%	7.24%
	600	0.105	0.061	0.013	0.093	0.047	0.008	0.068	0.029	0.004	-5.30%	7.85%
	1000	0.108	0.055	0.014	0.099	0.049	0.010	0.075	0.035	0.007	-4.42%	7.86%
	20000	0.090	0.046	0.010	0.113	0.057	0.009	0.105	0.050	0.011	0.62%	4.10%

The table shows the 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 an intercept, a strong and a useless factor. The true value of the cross-sectional R^2_{adj} is 43.87% (6.69%) for the OLS (GLS) estimation. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $\lambda_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a (annualized) prior SR of the factor model equal to 1.239, 1.305, 1.386, 1.413, and 1.497 for $T \in \{100, 200, 600, 1,000, \text{and } 20,000\}$.

Why does the Bayesian approach work while the frequentist one fails? The argument is probably best illustrated by Figure 1, which plots posterior distributions of B-SDF $\hat{\lambda}$ for both strong and useless factors from one of the simulations, along with their pseudo-true values of the price of risk (defined as 0 for the useless factor).

In this particular simulation, GMM estimates of $\lambda_{useless}$ imply significant price of risk for both OLS and GLS versions of the weight matrix, with traditional hypothesis testing rejecting the null of $\lambda_{useless} = 0$, even at 1% significance level. Instead, the B-SDF posteriors (blue lines in Figure 1) of the useless factor price of risk are diffuse and centered around 0. Intuitively,



(a) Useless factor λ, OLS

(b) Strong factor λ, OLS

(c) Useless factor λ, GLS

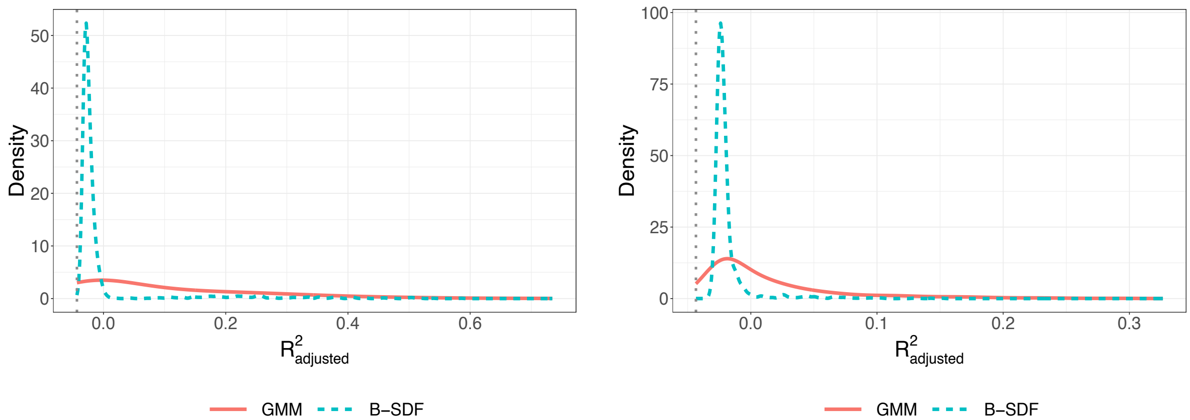
(d) Strong factor λ, GLS

Figure 1: Distribution of the price of risk estimates

Posterior distribution of the price of risk (blue dashed line) from B-SDF estimation of a misspecified one-factor model based on a single simulation with $T = 1000$ and asymptotic distribution of the frequentist GMM estimate (red solid line). The dotted line corresponds to the pseudo-true value of the parameter (defined to be 0 for a useless factor). Panels (a) and (c) correspond to the estimation of a model including a single useless factor. Panels (b) and (d) correspond to the case of including a single strong, well-identified factor.

the main driving force behind it is the fact that in B-SDF, $\boldsymbol{C}$ (the covariance of factors with returns) is updated continuously: When $\hat{\boldsymbol{C}}$ is close to zero, the posterior draws of $\boldsymbol{C}$ will be randomly positive or negative, which implies that the conditional expectation of $\boldsymbol{\lambda}$ in equation (11) will also switch sign from draw to draw. As a result, the posterior distribution of $\lambda_{useless}$ is centered around 0, and so is its confidence interval. The same logic applies to both OLS and GLS B-SDF formulations. Note that the Bayesian prior does not have any significant impact on the price of risk estimation of strong factors: In the case of well-identified sources of risk (Figure 1, panels (b) and (d)), the Bayesian and frequentist approach give very similar results.

Our setting also allows us to perform formal hypothesis testing via posterior probabilities and Bayes factors, following Corollary 2, even as $T \to \infty$, using the spike-and-slab prior of Section III.1.2. We report corresponding simulation results for the Bayesian p-value in Internet Appendix IA.A.1. Figure IA1 shows that useless factors are easily detected (their $p-$values, as



(a) OLS-type estimation

(b) GLS-type estimation

Figure 2: Cross-sectional distribution of OLS R^2_{adj} in a model with a useless factor

Empirical distribution of cross-sectional R^2 achieved by a misspecified model with a useless factor across 2,000 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 GMM estimates. The grey dotted line stands for the true value of R^2_{adj}.

expected, are sharply concentrated around the prior inclusion probability of 50% for any sample size), while true sources of risk are successfully selected with probability fast approaching 1.

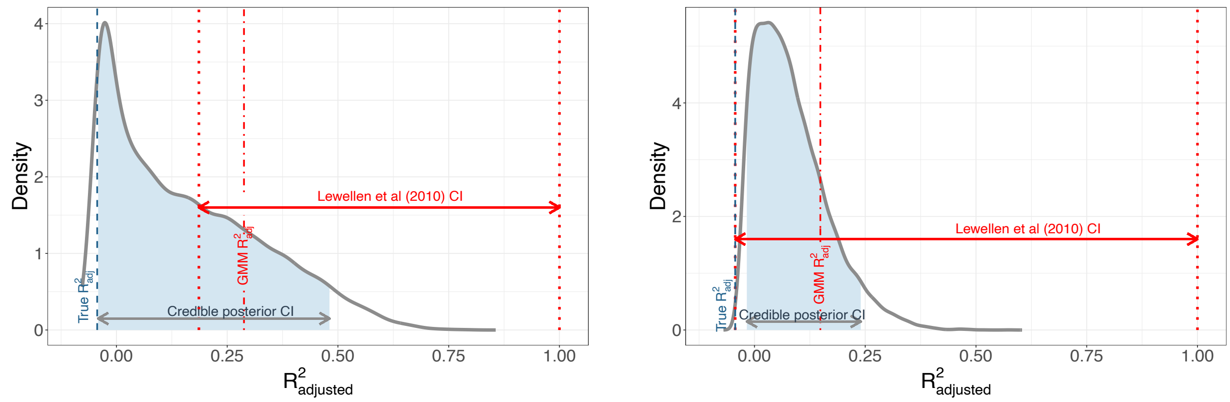
IV.1.1 Evaluating Cross-Sectional Fit

Weak identification notoriously affects not only parameter estimates but also conventional measures of fit, such as cross-sectional R^2 (Kleibergen and Zhan (2015)). We now show that the B-SDF approach restores not only inference on the price of risk but also the validity of the measures of cross-sectional fit.

Figure 2 shows the distribution of cross-sectional R^2 across a large number of simulations for the asymptotic case of $T = 20,000$ and a misspecified process for returns. For brevity, we focus on the most illustrative case of a single useless factor in the model. In this case frequentist estimation yields an extremely spreadout distribution of R^2 across simulations, which makes the researcher likely to conclude that the useless factor actually has significant explanatory power in the cross-section of returns.[28] This unfortunate property of the frequentist approach is not shared by our hierarchical Bayesian approach: The mode of the posterior distribution is tightly concentrated (across simulations) in the proximity of the true R^2 value.

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 B-SDF incorporates this uncertainty directly

[28]Gospodinov, Kan, and Robotti (2019) show examples of perfect fit obtainable with artificially generated useless factors and a family of one-step estimators.



(a) OLS-type estimation **(b)** GLS-type estimation

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). The blue dashed line denotes the true R^2_{adj}. The red dashed-dotted line depicts the GMM R^2_{adj} estimate with 90% Lewellen, Nagel, and Shanken (2010) confidence intervals (red dotted lines).

into the shape of its posterior distribution, one needs to rely on bootstrap-like algorithms to build a similar analogue in the frequentist case. As frequentist benchmark, we use the approach of Lewellen, Nagel, and Shanken (2010) to construct the confidence interval.

Figure 3 presents the posterior distribution of cross-sectional R^2 for a model that contains a useless factor and contrasts it with the frequentist values and their confidence intervals. The true adjusted R^2 is marginally negative, yet not only are its frequentist estimates economically large (29% and 19% for the OLS and GLS estimation types, respectively), but also the standard approach of Lewellen, Nagel, and Shanken (2010) yields extremely wide confidence intervals. Interestingly, they include a level of fit up to 100%, but not the true value. In contrast, while there is still considerable estimation uncertainty, the posterior distribution of the adjusted R^2 peaks in the proximity of 0 and is concentrated on much lower values. As shown in the last two columns of Table 1, this is a general property of the B-SDF estimation across simulation designs, sample sizes, and types of prior.

The B-SDF estimator performed well in a wide range of additional simulations that we have conducted. In particular, in Section IA.A.2 of the Internet Appendix we show that the B-SDF-based inference stays reliable even in the presence of what is typically considered a large cross-section (100 portfolios). This is reassuring, as it implies that our estimator does not require any specific adjustments for applications with either small time-series dimension or a large cross-sectional one (unlike popular frequentist alternatives).

IV.2 Selection via Bayes Factors

How well do flat and spike-and-slab priors work empirically in selecting relevant and detecting useless factors in the cross-section of asset returns? We revisit the theoretical results from Section III using the same simulation design therein.

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

T		55%	57%	59%	61%	63%	65%		55%	57%	59%	61%	63%	65%
							Panel A: Flat prior							
200	f_{strong}:	0.636	0.602	0.570	0.538	0.509	0.470	$f_{useless}$:	0.980	0.950	0.856	0.724	0.581	0.437
600		0.821	0.802	0.784	0.764	0.733	0.710		0.996	0.983	0.970	0.932	0.878	0.791
1,000		0.880	0.850	0.840	0.840	0.800	0.800		1.000	1.000	0.990	0.980	0.940	0.910
							Panel B: Spike-and-Slab, prior of $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]} = 0.295$							
200	f_{strong}:	0.815	0.761	0.721	0.675	0.630	0.581	$f_{useless}$:	0.004	0.000	0.000	0.000	0.000	0.000
600		0.974	0.961	0.954	0.943	0.926	0.899		0.000	0.000	0.000	0.000	0.000	0.000
1,000		0.980	0.970	0.970	0.960	0.960	0.940		0.000	0.000	0.000	0.000	0.000	0.000
							Panel C: Spike-and-Slab, prior of $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]} = 0.807$							
200	f_{strong}:	0.527	0.489	0.449	0.412	0.381	0.349	$f_{useless}$:	0.041	0.007	0.004	0.000	0.000	0.000
600		0.859	0.832	0.811	0.774	0.734	0.690		0.001	0.000	0.000	0.000	0.000	0.000
1,000		0.910	0.910	0.870	0.850	0.830	0.820		0 .000	0.000	0.000	0.000	0.000	0.000

Frequency of retaining risk factors using BF for different samples size (T=200, 600, and 1,000) across 2,000 simulations of a misspecified model with strong and useless factors. A factor is retained if its posterior probability, $\Pr(\gamma_i = 1|data)$, is greater than a given threshold: 55%, 57%, 59%, 61%, 63%, and 65%. Returns and factors are standardized. Panel A reports results for the flat prior. Panels B and C use the spike-and-slab approach of Section III.1.3 with demeaned correlations, $r = 0.001$ and $\psi = 1$ or 10, mapping into the corresponding monthly Sharpe ratios, $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]}$, listed in the table. The prior for each factor inclusion in Panels B and C is a $Beta(1, 1)$, yielding a prior expectation for factor inclusion of 50%.

We consider a misspecified model with both strong and useless factors and compute Bayes factors for each of the potential sources of risk. Table 2 reports the empirical frequency of variable retention in the model across 2,000 simulations of different sample sizes ($T = 200$, 600, and 1,000). We first report the probability of retaining a factor under a flat prior, which is standard in the literature. Second, we use the continuous spike-and-slab prior for the price of risk and compute the marginal probability of each factor as the posterior mean of γ_j. The decision rule is based on a range of critical values, 55%–65%, such that when the posterior factor probability ($\Pr[\gamma_j = 1|data]$) is above a particular threshold, we retain the factor.

The difference generated by the two priors is drastic in the presence of useless factors. As discussed in Section III.1.1, under a flat prior for the price of risk, the posterior probability of including a useless factor in the model converges to 1 asymptotically. Table 2 makes it clear that the same holds even for a very short sample, making the overall process of model selection completely invalid. In turn, factor selection via the spike-and-slab prior approach of Section III.1.3 is reliable in both retaining strong factors and excluding useless ones (even with a very small sample size). As Panels B and C indicate, our results are robust to different prior values for the factor Sharpe ratio.

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 useless factors from the model and identifies the true sources of risk.

V Empirical Analysis

In this section we apply our hierarchical Bayesian method to a large set of factors proposed in the previous literature. First, 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 risk prices (Section V.1). Second, based on the results of this estimation, in Section V.2 we construct an SDF via Bayesian Model Averaging and show its superior asset pricing properties. Following Martin and Nagel (2019), we consider not only in-sample but also out-of-sample performance (both in the time-series and cross-sectional dimension) and compare the BMA-SDF with both notable reduced-form models and the shrinkage-based approach to factor aggregation (Kozak, Nagel, and Santosh (2020)). Finally, in Sections V.3 and V.4 we study whether one can achieve an accurate representation of the SDF with low-dimensional (observable) factor models, and show that such conjecture is not supported by the data. Strikingly, our results indicate that there is scope for both selection and aggregation in linear factor models.

V.1 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 IA13 of the Internet Appendix.

As test assets we consider a cross-section of 60 asset returns that are meant to capture well-documented cross-sectional anomalies. These include all the (34) tradable (long-short) factors in Panel A of Table A1 in the Appendix, and an additional set of 26 long-short portfolios based on the univariate sorting of the characteristics listed in Panel B of the same table. The inclusion of the tradable factors among the test assets, and the usage of the non-spherical pricing error formulation (i.e., GLS), also imposes (asymptotically) the restriction of factors pricing themselves.[29]

Since we do not restrict the maximum number of factors to include, all the possible combinations of factors give us a total of 2^{51} possible specifications, that is 2.25 quadrillion models. We

[29]Note that we could also have enforced this pricing restriction in finite sample using an ad hoc prior for these factors – which is analogous to estimating the model via the GLS version of the beta representation of expected returns, and then inverting the estimates to obtain the price of risk of the SDF formulation.

use the continuous spike-and-slab approach of Section III.1.3 with non-spherical errors, since it easily handles a very large number of possible models while remaining valid in the presence of the most common identification failures. We report both posterior probabilities (given the data) of each factor (i.e., $\mathbb{E}[\gamma_j|\text{data}], \forall j$) as well as the posterior means of the factors' price of risk (i.e., $\mathbb{E}[\lambda_j|\text{data}], \forall j$) computed as the Bayesian Model Average (BMA) across the universe of models. We use the formulation of the penalty term ψ_j in equation (23) in order to also handle identification failures of factors' price of risk caused by level factors (see Remark 4).[30]

The posterior evaluation is performed and reported over a wide range for the parameter ψ (in equation (23)) that regulates the degree of shrinkage of potentially useless factors. This parameter controls the prior belief about the Sharpe ratio achievable with the pricing factors. We tabulate the results in units of *Sharpe ratio prior* defined as $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]}$, since this is a natural metric of beliefs. The lower value that we consider, a prior SR of 1, generates a strong shrinkage (small ψ), while the highest value reported, a prior SR of 3.5, makes the shrinkage virtually irrelevant. Since our prior gives non-zero probability to any SR value, these are *not* hard constraints.[31]

The prior probability for each factor inclusion is drawn from a $Beta(1,1)$ (i.e., 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.[32]

Figure 4 plots the posterior probabilities of the 51 factors as a function of the prior SR. The corresponding values are reported in Table 3.

First, there is particularly strong evidence for including the BEH_PEAD factor of Daniel, Hirshleifer, and Sun (2020), or (behavioral) post-earnings announcement drift anomaly, as the source of priced risk in the SDF. This factor is meant to capture investors' limited attention. The posterior probability of this factor being part of the SDF is over 70% for most prior values. This might not be too surprising, given that many anomaly portfolios seem to be associated with short-term market inefficiencies.

Second, the excess return on the market (MKT) appears as a likely source of priced risk with posterior probabilities significantly above the prior for a wide range of prior SR. This is both surprising and reassuring. Surprising, since the market return is rarely found to be significant

[30] In Internet Appendix IA.B.2 we report results based on the formulation in equation (22) as well as the Fisher transformation of the correlation coefficients. The findings therein are very similar to the ones discussed below. Table IA15 reports the values of the squared correlations, and their cross-sectionally demeaned version, of factors and test assets.

[31] We report results for an extended range of the SR prior, starting from a prior at 0 (corresponding to a 100% dogmatic shrinkage) in Figure IA2 and Table IA16 of the Internet Appendix.

[32] We obtain virtually identical results 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. Furthermore, using a prior in favor of more sparse factor models (a $Beta(1,9)$), the empirical findings are very similar to the ones reported. These additional results are reported in Section IA.B.2 of the Internet Appendix.

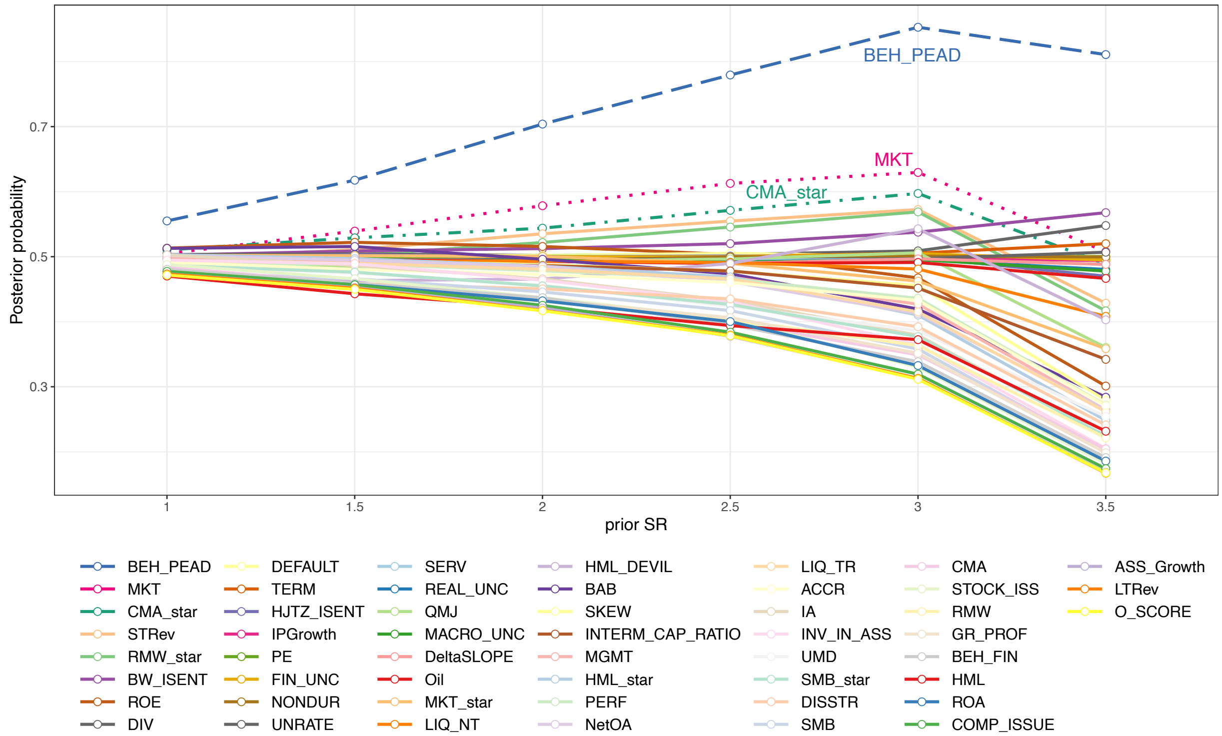


Figure 4: Posterior factor probabilities

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, computed using the continuous spike-and-slab approach of Section III.1.3 and 51 factors described in Table A1 of the Appendix. Sample: 1973:10–2016:12. Test assets: 60 anomaly 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 for different values of the prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR_f^2 \mid \sigma^2]}$, annualized.

for cross-sectional asset pricing. It is reassuring because Giglio and Xiu (2021) show that once inference is corrected for potential misspecification, the market factor appears to be priced. In our setting, estimation across all the universe of possible models is meant exactly to address the misspecification problem, and it seems to do so successfully.

Third, the CMA* factor of Daniel, Mota, Rottke, and Santos (2020) shows a non-trivial increase in the posterior probability of being part of the SDF. This is the investment factor of Fama and French (2015) without its unpriced component.

Fourth, there are three more factors (RMW*, STRev, and RMW*, described in Table A1) for which the posterior probability estimate provide some (albeit not strong) support.

Fifth, there is a substantial set of factors for which the posterior probability stays roughly equal to the prior one. That is, these factors are likely to be weakly identified at best. Finally, there is a large set of factors that is unlikely to be part of the SDF pricing our data (e.g., long-short portfolios sorted by the Ohlson O-score, long-term reversal, and asset growth).

Interestingly, the results are not very sensitive to the choice of prior maximum Sharpe ratio unless there is almost no shrinkage, that is, there is no protection against weakly identified factors. In this latter case, weakly identified factors seem to drive out the statistical support for likely components of the true SDF, which is consistent with the findings of Gospodinov, Kan, and Robotti (2014) for the frequentist estimation of linear factor models.

In addition to the posterior probabilities of the factors, Table 3 reports the posterior means of the price of risk computed as Bayesian Model Average (BMA), that is, 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), and Madigan and Raftery (1994)).

Several observations are in order. First, the price of risk estimates for factors that are more likely to be part of the SDF (top three to six factors) are relatively stable for non-extreme values of the prior SR. Second, for factors that are likely to be at best weakly identified the estimated price of risk is very close to zero but becomes large when the prior SR is very high, and therefore the estimation is no more robust to the weak factors. This is to be expected given the frequentist results on this issue. Third, for factors for which there is clear evidence that they should not be part of the SDF, the estimates of the price of risk are stably around zero. Furthermore, for these factors they are very close to zero even *conditional* on the factors being included in the SDF. This quantity can be easily computed by dividing the posterior mean of the price of risk by the factor posterior probability – both reported in Table 3.

As a reality check on the results in Table 3, in Table 6 of Section V.3 below, we expand our set of candidate priced factors to include artificially generated weak factors and show that our procedure successfully singles them out. Furthermore, in the above estimation we have allowed for a common cross-sectional intercept due to allowing for an average level of mispricing. In Tables IA18–IA19 of the Internet Appendix we repeat the estimation imposing a zero common intercept and obtain virtually identical results.[33]

Finally, since we sample the space of 2 quadrillion models instead of estimating them one-by-one, one might wonder whether the estimation is accurate. We address this formally with the standard Separated Partial Means test (see, e.g., Geweke (2005)) for both posterior probabilities and prices of risk, which clearly indicates fast and accurate convergence of the Markov Chain-based estimates.[34]

A natural question is whether the posterior probabilities and prices of risk estimates, sum-

[33]The fact that imposing the zero intercept restriction leaves the results virtually unchanged is not too surprising since, across all our estimates, the posterior mean of the common intercept is about 0.02–0.03 in monthly SR unit. Hence, since the average monthly variance of the baseline test assets is about 4.5%, the posterior mean of the common intercept is about 0.09%–0.135% in monthly returns units i.e. it is quite small.

[34]To implement the test we drop the first 50,000 draws and split our Markov Chain in five subsets. We compute the average frequency of rejection of posterior probability of factor inclusion and price of risk being the same for all the subsets for different values of the test size (i.e., 95%, 90%, and 80%). The corresponding empirical frequencies of rejection are 6.0%, 9.9%, and 20.2% for the posterior probability of factor inclusion and 4.1%, 9.1%, and 20.4% for the price of risk. In addition, we have repeated the estimation increasing the number of draws by a factor of 10, and found virtually identical parameter values.

Table 3: Posterior factor probabilities, $\mathbb{E}[\gamma_j|\text{data}]$, and risk prices: 2.25 quadrillion models

	Factor inclusion prob., $\mathbb{E}[\gamma_j\|\text{data}]$						Price of risk, $\mathbb{E}[\lambda_j\|\text{data}]$					
	Total prior SR						Total prior SR					
Factors:	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
BEH_PEAD	0.555	0.618	0.704	0.779	0.853	0.811	0.018	0.043	0.085	0.146	0.231	0.278
MKT	0.505	0.539	0.578	0.613	0.630	0.508	0.017	0.040	0.073	0.114	0.170	0.186
CMA*	0.510	0.529	0.544	0.571	0.597	0.488	0.011	0.023	0.041	0.067	0.106	0.117
STRev	0.496	0.511	0.535	0.555	0.572	0.428	0.007	0.018	0.036	0.060	0.093	0.090
RMW*	0.499	0.502	0.522	0.546	0.569	0.417	0.009	0.020	0.038	0.065	0.105	0.099
BW_ISENT	0.502	0.509	0.512	0.520	0.538	0.568	0.002	0.005	0.009	0.016	0.035	0.122
ROE	0.513	0.522	0.516	0.503	0.467	0.301	0.021	0.039	0.056	0.075	0.093	0.077
DIV	0.503	0.504	0.502	0.503	0.509	0.548	0.000	0.001	0.002	0.004	0.009	0.042
DEFAULT	0.501	0.501	0.502	0.505	0.501	0.500	0.000	0.001	0.001	0.003	0.006	0.022
TERM	0.501	0.498	0.498	0.500	0.505	0.520	0.000	-0.001	-0.002	-0.004	-0.008	-0.037
HJTZ_ISENT	0.499	0.503	0.500	0.501	0.499	0.470	0.001	0.002	0.003	0.005	0.009	0.029
IPGrowth	0.501	0.501	0.500	0.496	0.498	0.494	0.000	0.000	-0.001	-0.002	-0.004	-0.014
PE	0.497	0.497	0.500	0.498	0.500	0.500	0.000	-0.001	-0.002	-0.003	-0.007	-0.029
FIN_UNC	0.494	0.491	0.500	0.500	0.505	0.495	0.001	0.002	0.003	0.007	0.016	0.050
NONDUR	0.494	0.493	0.495	0.499	0.501	0.500	0.001	0.001	0.003	0.005	0.012	0.051
UNRATE	0.496	0.494	0.496	0.495	0.497	0.507	0.000	0.001	0.002	0.003	0.008	0.038
SERV	0.493	0.495	0.494	0.495	0.495	0.488	0.000	0.000	0.001	0.001	0.003	0.018
REAL_UNC	0.496	0.495	0.493	0.492	0.495	0.480	0.000	0.000	0.001	0.002	0.005	0.010
QMJ	0.492	0.484	0.493	0.496	0.506	0.360	0.016	0.030	0.050	0.081	0.132	0.128
MACRO_UNC	0.496	0.493	0.495	0.491	0.496	0.478	0.000	0.000	0.001	0.001	0.003	0.001
DeltaSLOPE	0.494	0.495	0.493	0.490	0.497	0.488	0.000	0.001	0.001	0.002	0.004	0.016
Oil	0.498	0.495	0.493	0.490	0.491	0.467	0.000	0.000	0.001	0.002	0.005	0.021
MKT*	0.502	0.502	0.500	0.490	0.462	0.358	0.007	0.015	0.024	0.034	0.043	0.057
LIQ_NT	0.492	0.493	0.493	0.491	0.481	0.408	0.000	0.001	0.000	-0.002	-0.010	-0.026
HML_DEVIL	0.471	0.463	0.466	0.490	0.543	0.403	0.008	0.017	0.036	0.073	0.152	0.163
BAB	0.513	0.516	0.496	0.474	0.419	0.284	0.015	0.027	0.037	0.046	0.052	0.049
SKEW	0.493	0.494	0.488	0.478	0.455	0.279	0.013	0.027	0.043	0.061	0.082	0.061
INTERM_CAP_RATIO	0.496	0.491	0.486	0.478	0.452	0.342	0.006	0.013	0.021	0.027	0.028	0.016
MGMT	0.498	0.494	0.479	0.469	0.427	0.264	0.020	0.032	0.044	0.061	0.077	0.062
HML*	0.503	0.497	0.485	0.469	0.410	0.248	0.010	0.020	0.031	0.041	0.045	0.033
PERF	0.489	0.489	0.478	0.466	0.436	0.272	0.012	0.022	0.034	0.047	0.065	0.053
NetOA	0.502	0.495	0.485	0.462	0.413	0.265	0.006	0.013	0.019	0.026	0.030	0.027
LIQ_TR	0.494	0.490	0.481	0.466	0.415	0.262	0.003	0.007	0.012	0.018	0.023	0.019
ACCR	0.491	0.480	0.473	0.460	0.433	0.271	0.004	0.008	0.016	0.028	0.041	0.034
IA	0.503	0.486	0.466	0.432	0.379	0.224	0.018	0.028	0.037	0.044	0.051	0.041
INV_IN_ASS	0.495	0.489	0.464	0.431	0.365	0.205	0.009	0.015	0.021	0.025	0.026	0.018
UMD	0.486	0.475	0.456	0.424	0.386	0.254	0.007	0.010	0.011	0.011	0.015	0.023
SMB*	0.487	0.476	0.455	0.426	0.377	0.224	0.005	0.009	0.014	0.019	0.025	0.020
DISSTR	0.474	0.459	0.451	0.435	0.392	0.241	-0.002	-0.009	-0.020	-0.034	-0.047	-0.040
SMB	0.476	0.466	0.446	0.417	0.358	0.199	0.010	0.019	0.029	0.036	0.037	0.025
CMA	0.484	0.459	0.435	0.400	0.349	0.204	0.011	0.012	0.009	0.000	-0.015	-0.015
STOCK_ISS	0.488	0.466	0.437	0.404	0.330	0.182	0.011	0.017	0.021	0.024	0.021	0.015
RMW	0.471	0.455	0.432	0.403	0.363	0.221	0.005	0.005	0.002	-0.006	-0.023	-0.019
GR_PROF	0.475	0.454	0.434	0.406	0.352	0.198	0.001	0.002	0.004	0.006	0.007	0.001
BEH_FIN	0.480	0.459	0.437	0.396	0.338	0.191	0.014	0.018	0.020	0.018	0.012	0.012
HML	0.470	0.443	0.422	0.394	0.372	0.232	0.005	0.001	-0.006	-0.019	-0.044	-0.042
ROA	0.472	0.457	0.432	0.400	0.333	0.186	0.009	0.013	0.015	0.014	0.009	0.003
COMP_ISSUE	0.477	0.457	0.425	0.384	0.319	0.174	0.006	0.007	0.007	0.005	0.002	0.004
A_Growth	0.474	0.452	0.421	0.378	0.312	0.168	0.007	0.008	0.006	0.002	-0.002	-0.003
LTRev	0.473	0.451	0.417	0.379	0.313	0.167	0.004	0.005	0.005	0.004	0.001	0.001
O_SCORE	0.472	0.450	0.417	0.378	0.311	0.168	-0.004	-0.006	-0.006	-0.005	-0.007	-0.005

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, and posterior mean of factors' risk prices, $\mathbb{E}[\lambda_j|\text{data}]$, are computed using the continuous spike-and-slab approach of Section III.1.3 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. Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12. Results are tabulated for different values of the (annualized) prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]}$. Light-shaded grey rows denote non-tradable factors.

marized in Table 3, deliver a good representation of the true latent SDF.

V.2 Cross-Sectional Performance

We now focus on the cross-sectional asset pricing performance of our BMA estimates of the Stochastic Discount Factor (BMA-SDF), both in- and out-of-sample, and compare it with traditional popular reduced-form factor models. Table 4 reports root mean squared pricing error (RMSE), mean absolute pricing errors (MAPE), and OLS and GLS cross-sectional R^2 for a variety of models and test assets. For a benchmark comparison, we consider the CAPM, Fama-French five-factor model, Carhart four-factor model, and the q4 model of Hou, Xue, and Zhang (2015). Finally, we also present results for the 51 factor model that includes all the candidate risk factors considered in our analysis, as well as the shrinkage-based approach of Kozak, Nagel, and Santosh (2020) (KNS) with the optimal shrinkage level and number of factors chosen by three-fold cross-validation.[35] Results for the Bayesian (GLS) SDFs are reported for a wide range of SR priors. All the frequentist SDFs are estimated via a GLS version of the GMM (i.e., imposing the tradability restriction on the model-implied price of risk, whenever factors are tradable).[36]

Panel A reports in-sample asset pricing statistics for the baseline set of assets used in our estimation (60 anomaly portfolios).[37] It is striking that the Bayesian SDF tends to outperform conventional models across a wide range of metrics, and this result is stable across the whole set of SR priors. Furthermore, unlike the benchmark models, the BMA-SDF delivers cross-sectional OLS and GLS R^2 measures that are consistent with each other – without explicitly targeting any of them at the SDF estimation stage. The only model that seems to perform better than the BMA-SDF is the one using 51 factors to price 60 assets and is very likely to be overfitting the cross-section (as we show below). One might wonder whether part of the Bayesian SDF success could also be due to overfitting. We address this issue by analyzing its OOS performance, in both cross-sectional and time-series dimensions.

Panels B and C summarize the performance of SDFs estimated on a set of 60 anomaly portfolios ($\widehat{M_t}$) but then used to price a different cross-section – 25 portfolios sorted by size and value (Panel B), and 49 industry portfolios (Panel C). Since we shrink away level factors in the BMA-SDF, to put different models on equal footing we focus on cross-sectionally demeaned pricing errors. Our findings make it clear that the superior performance of the BMA-SDF observed

[35] When applied to our sample of 60 portfolios, three-fold CV selects a model with 11 factors and the root expected SR^2 of 1.2.

[36] We have also obtained virtually identical results using time-series regressions (with tradable factors) instead of GMM, as well as other cross-sections not reported in Table 4.

[37] The table reports the following measures: $RMSE \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}\alpha_i^2}$, $MAPE \equiv \frac{1}{N}\sum_{i=1}^{N}|\alpha_i|$, $R^2_{ols} \equiv 1 - \frac{\left(\boldsymbol{\alpha}-\frac{1}{N}\boldsymbol{\alpha}^\top\mathbf{1}_N\right)^\top\left(\boldsymbol{\alpha}-\frac{1}{N}\boldsymbol{\alpha}^\top\mathbf{1}_N\right)}{(\boldsymbol{\mu}_{\boldsymbol{R}}-\frac{1}{N}\boldsymbol{\mu}_{\boldsymbol{R}}^\top\mathbf{1}_N)^\top(\boldsymbol{\mu}_{\boldsymbol{R}}-\frac{1}{N}\boldsymbol{\mu}_{\boldsymbol{R}}^\top\mathbf{1}_N)}$, and $R^2_{gls} \equiv 1 - \frac{\boldsymbol{\alpha}^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{\alpha}}{\boldsymbol{\mu}_{\boldsymbol{R}}^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{\mu}_{\boldsymbol{R}}}$.

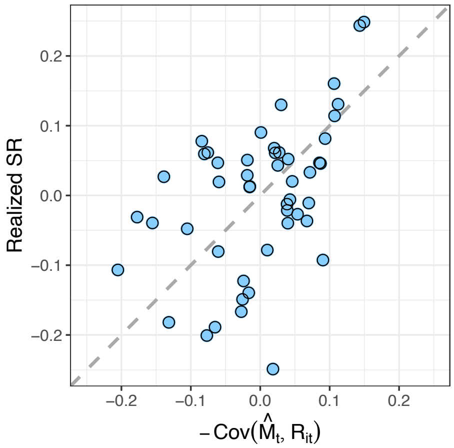
Table 4: Cross-sectional asset pricing

Model		RMSE	MAPE	R^2_{ols}	R^2_{gls}	Model	RMSE	MAPE	R^2_{ols}	R^2_{gls}
Panel A: In-sample pricing, test assets: 60 anomalies										
BMA-SDF:	$SR_{pr}=1$	0.287	0.227	39.2%	24.2%	51 factors	0.041	0.022	98.1%	97.7%
	$SR_{pr}=1.5$	0.253	0.197	49.8%	30.3%	CAPM	0.418	0.338	-29.4%	16.8%
	$SR_{pr}=2$	0.223	0.170	59.1%	37.4%	FF5	0.301	0.223	24.5%	23.2%
	$SR_{pr}=2.5$	0.193	0.148	68.2%	45.5%	Carhart	0.317	0.244	21.5%	21.2%
	$SR_{pr}=3$	0.162	0.128	76.6%	54.7%	q4	0.267	0.189	37.5%	28.1%
	$SR_{pr}=3.5$	0.157	0.128	78.4%	58.8%	KNS_{CV_3}	0.296	0.237	53.7%	19.6%
Panel B: Cross-sectional out-of-sample pricing, test assets: 25 size-value portfolios										
BMA-SDF:	$SR_{pr}=1$	0.108	0.082	42.1%	17.5%	51 factors	0.200	0.163	-98.5%	-1653%
	$SR_{pr}=1.5$	0.094	0.070	55.7%	24.5%	CAPM	0.145	0.112	-4.6%	5.2%
	$SR_{pr}=2$	0.085	0.063	64.5%	30.2%	FF5	0.079	0.059	69.2%	28.0%
	$SR_{pr}=2.5$	0.077	0.058	70.5%	34.9%	Carhart	0.086	0.063	63.2%	27.1%
	$SR_{pr}=3$	0.073	0.054	73.9%	38.4%	q4	0.083	0.065	66.1%	28.2%
	$SR_{pr}=3.5$	0.075	0.055	72.3%	36.8%	KNS_{CV_3}	0.096	0.074	54.4%	28.0%
Panel C: Cross-sectional out-of-sample pricing, test assets: 49 industry portfolios										
BMA-SDF:	$SR_{pr}=1$	0.097	0.080	15.6%	11.8%	51 factors	0.420	0.310	-1474.3%	-1694%
	$SR_{pr}=1.5$	0.097	0.082	15.3%	12.8%	CAPM	0.111	0.082	-10.6%	20.9%
	$SR_{pr}=2$	0.097	0.082	15.7%	15.8%	FF5	0.123	0.103	-35.8%	3.6%
	$SR_{pr}=2.5$	0.098	0.081	14.9%	18.5%	Carhart	0.117	0.089	-22.1%	13.7%
	$SR_{pr}=3$	0.100	0.083	10.9%	19.7%	q4	0.134	0.105	-60.5%	-10.9%
	$SR_{pr}=3.5$	0.100	0.083	11.5%	20.9%	KNS_{CV_3}	0.100	0.082	10.9%	14.0%

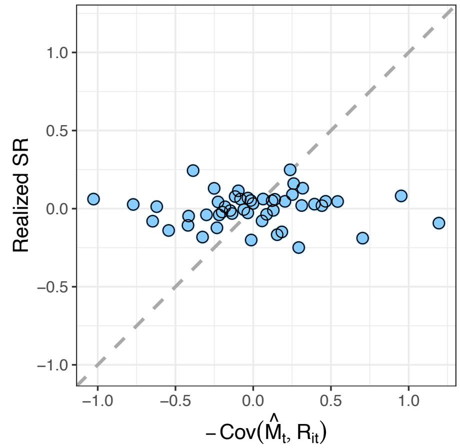
This table compares in-sample and cross-sectional out-of-sample asset pricing performance of the BMA-SDF and notable frequentist factor models. We use GMM-GLS to estimate factor prices of risk for the CAPM, FF5 model of Fama and French (2015), Carhart (1997) model, q4 model of Hou, Xue, and Zhang (2015), and the model including all 51 factors. KNS stands for the SDF estimation of Kozak, Nagel, and Santosh (2020), with tuning parameter and number of factors chosen by three-fold cross-validation. For the BMA-SDF, we report results with risk prices under a range of (annualized) prior Sharpe ratio values: $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]} \in \{1\text{–}3.5\}$. In the cross-sectional OOS the models are first estimated using the baseline test assets of Panel A and then used to price (without additional parameters estimation), the test assets listed in Panels B and C. All the data is standardized, that is, pricing errors are in SR units. We report the annualized RMSE and MAPE.

in-sample is not due to overfitting. While the 51-factor model has a disastrous cross-sectional OOS performance, this is not the case for the BMA-SDF. Consistent with our in-sample results, the performance of the Bayesian SDF is stable across priors and metrics. Furthermore, it is either on par with that of the best reduced-form benchmark model (the FF5 model when focusing on size-value portfolios) or better. The BMA-SDF pricing ability is particularly striking in the case of industry portfolios that have long been considered a challenge for asset pricing and often advocated as an appropriate testing ground for models (e.g., Lewellen, Nagel, and Shanken (2010), and Daniel and Titman (2012)).

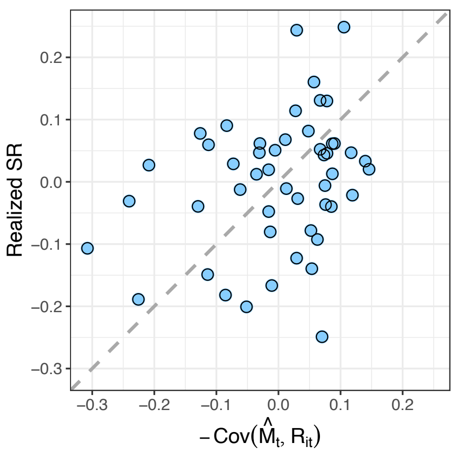
Figure 5 further illustrates the performance of different SDFs estimated on the baseline cross-section and then used to price the 49 industry portfolios. The BMA-SDF is the only model that generates predicted Sharpe ratios close to the observed ones and has positive (OLS and GLS) cross-sectional R^2s. Note that while some of the models yield predictions that have positive correlation with the actual return realizations, they are still characterized by a substantially negative R^2 since we impose the theoretical pricing restriction of $\mathbb{E}[\boldsymbol{R_t}] = -Cov(M_t, \boldsymbol{R_t})$ (using



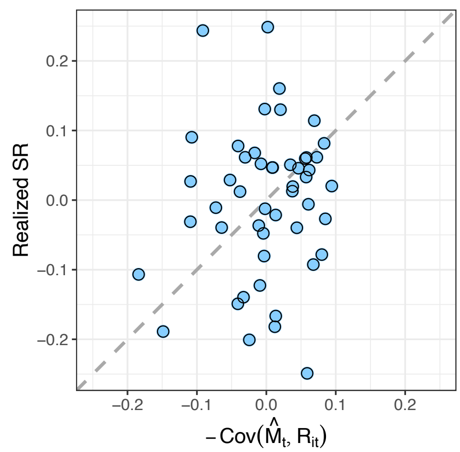
(a) B-SDF, $SR_{pr} = 2$ ($R^2_{ols} = 16\%$)



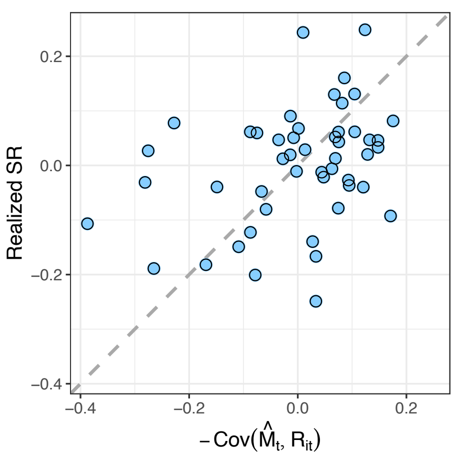
(b) 51-factor model ($R^2_{ols} = -1474\%$)



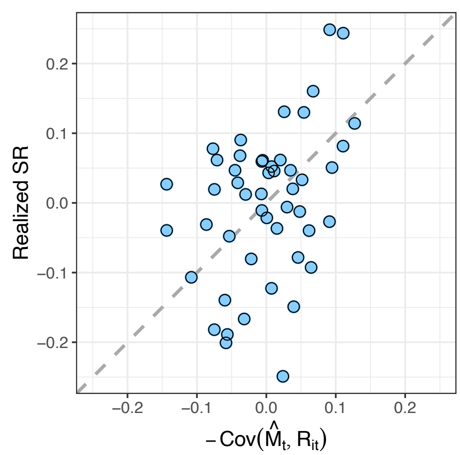
(c) FF5 ($R^2_{ols} = -36\%$)



(d) Carhart ($R^2_{ols} = -22\%$)



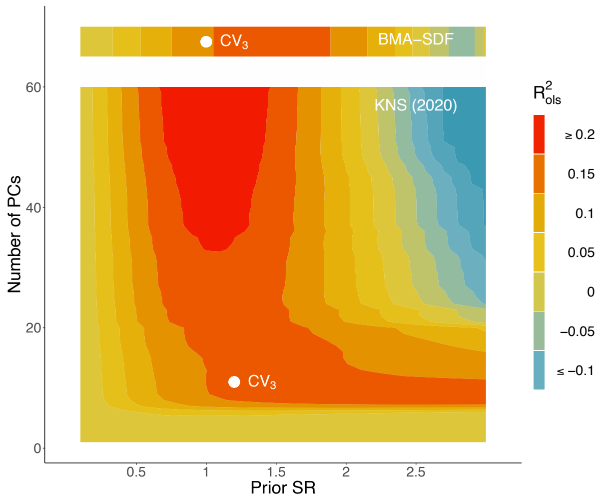
(e) q4 ($R^2_{ols} = -61\%$)



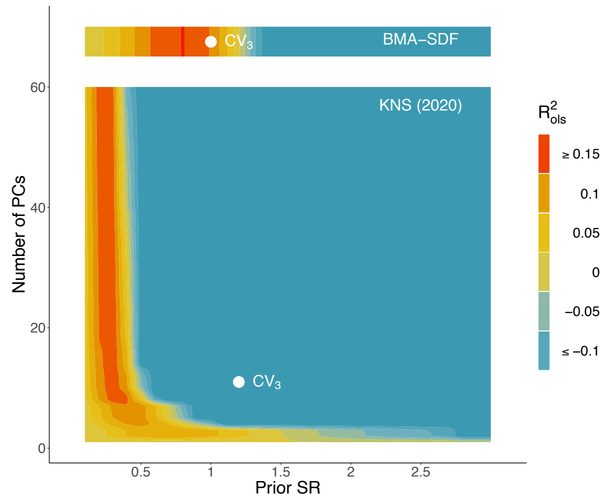
(f) KNS$_{CV_3}$ ($R^2_{ols} = 10\%$)

Figure 5: Out-of-sample cross-sectional pricing of 49 industry portfolios

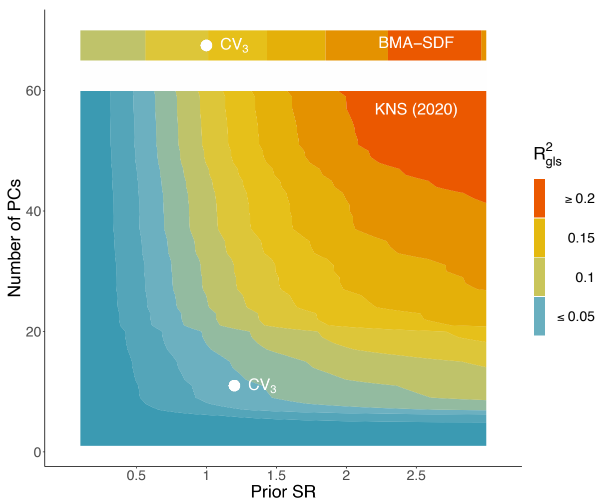
For each model, the figure depicts the out-of-sample performance of the SDF, obtained by using 60 anomaly portfolios as test assets, and applied to pricing 49 industry portfolios without re-estimation. All the data are standardized; that is, pricing errors are in SR units. The 45-degree line corresponds to the theoretical relationship of $\mathbb{E}[\boldsymbol{R_t}] = -Cov(M_t, \boldsymbol{R_t})$, where SDFs are normalized to have unit mean.



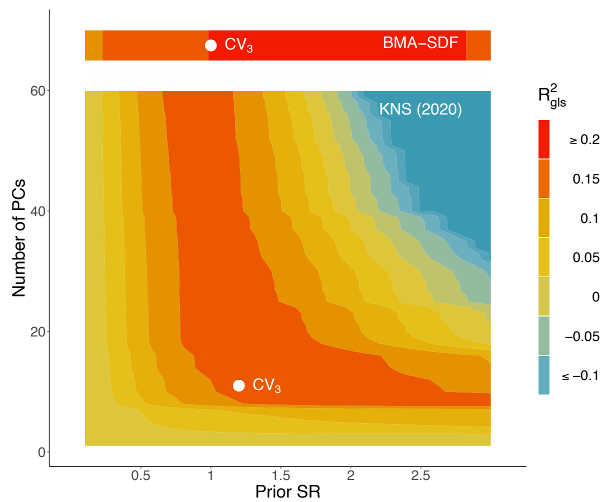
(a) R^2_{ols}, estimation period: 1995/06-2016/12, evaluation period: 1973/10-1995/05



(b) R^2_{ols}, estimation period: 1973/10-1995/05, evaluation period: 1995/06-2016/12



(c) R^2_{gls}, estimation period: 1995/06-2016/12, evaluation period: 1973/10-1995/05



(d) R^2_{gls}, estimation period: 1973/10-1995/05, evaluation period: 1995/06-2016/12

Figure 6: Out-of-sample cross-sectional pricing (different time samples)

The figure depicts out-of-sample performance of the SDF (R^2_{ols} and R^2_{gls}), obtained by using both BMA and Kozak, Nagel, and Santosh (2020) approaches using a time series subsample of 60 anomaly portfolios. We use half of the time-series sample for the model estimation and SDF recovery and evaluate its cross-sectional pricing ability on the other subsample. Results are reported for a range of annualized SR prior, and in the case of KNS (2020) for different number of PCs used, as well as a combination of tuning parameters and priors chosen by a three-fold cross-validation (CV_3) applied to the estimation period.

the innocuous normalization $\mathbb{E}[M_t] = 1$).

We now turn to the time-series out-of-sample performance of the BMA-SDF.[38] According to

[38] We follow the approach canonical in the literature of performing time series OOS via a split-sample (see, e.g., Linnainmaa and Roberts (2018), Chen, Pelger, and Zhu (2019), Gu, Kelly, and Xiu (2020)). Nevertheless, ideally, one might want to focus on the post publication sample of the factors. This is unfortunately unfeasible in our empirical setting since a large share of the factors that we analyze have been only very recently documented.

Table 4, only the shrinkage-based approach of Kozak, Nagel, and Santosh (2020) comes close to matching the performance of our Bayesian approach overall. Hence, we use it as a benchmark model for the time-series out-of-sample performance. Figure 6 reports out-of-sample model performance, based on the time-series difference between estimation and prediction periods. Following Martin and Nagel (2019), we use half of the time-series sample for the model estimation and SDF recovery and evaluate its cross-sectional pricing ability on the other subsample. Thus, we consider out-of-sample performance of the model, going into both future and past, without re-estimating any of the parameters. For the same value of the prior SR, BMA-SDF tends to outperform the cross-validated estimates (CV_3) of KNS, despite the fact that cross-validation was carried out on the full data sample. Furthermore, for a wide range of prior SR, our Bayesian approach performs either as well as the ex-post best combination of tuning parameters in KNS or better. This is particularly evident when recent data is used as the evaluation subsample. Finally, to compare the BMA-SDF and KNS-SDF on similar footing, albeit this is not natural in a Bayesian setting, we select the prior hyper-parameters for the former using the same three-fold cross-validation as for the latter (yielding a prior SR of 1, singled out in Figure 6). The resulting OOS performance of the BMA-SDF is in the same ballpark, or better, than the KNS-SDF.[39]

V.3 Model Uncertainty: Selection or Aggregation?

In the previous section we have shown that averaging across the space of possible models yields an accurate representation of the SDF. A natural question is whether in the universe of models there is a *single* best model.

For consistency, frequentist model selection *demands* the existence of a unique first-best model that can be reliably distinguished from the alternatives. This is a key assumption underlying reliable factor selection via t- and χ^2-tests, LASSO, and many other approaches.

In contrast, the existence of such a dominant model can be formally assessed within the Bayesian paradigm. For instance, Giannone, Lenza, and Primiceri (2021) study the sparsity assumption in popular empirical economic applications (using, like us, a spike-and-slab prior approach for model and variable selection). They find that the posterior distribution does not typically concentrate on a single sparse model but rather supports a wide set of models that often include a large number of predictors.

Figure 7 presents the model posterior probabilities of the 2,000 most likely specifications (with the annualized prior SR of 2).[40] The first thing to notice is that even the most likely spec-

[39] The cross-sectional measures of fit in Figure 6 for the cross-validated BMA and KNS SDFs are, respectively: Panel (a), 14.3% and 15.9%; Panel (b), 11.8% and -234%; Panel (c), 12.6% and 5.6%; Panel (d), 17.9% and 15.2%.

[40] Note that the posterior model probabilities decay in a step-like manner due to numerical rounding. This

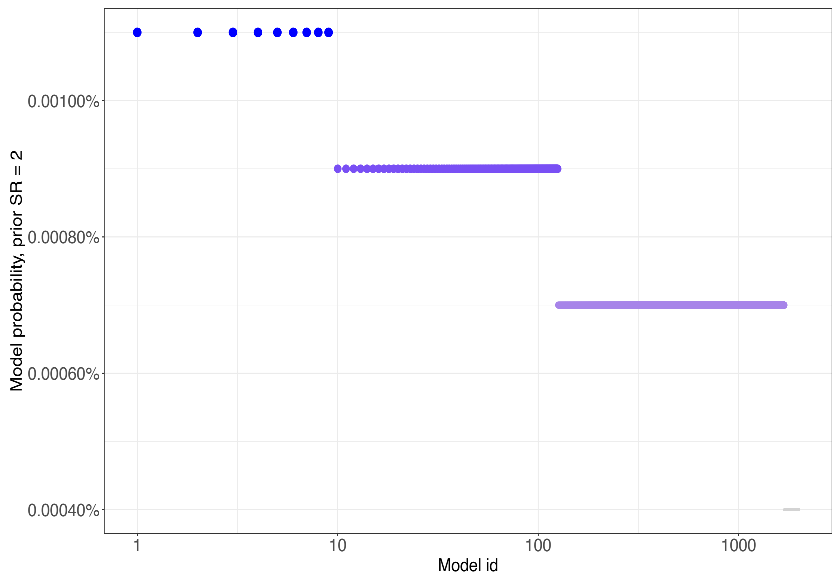


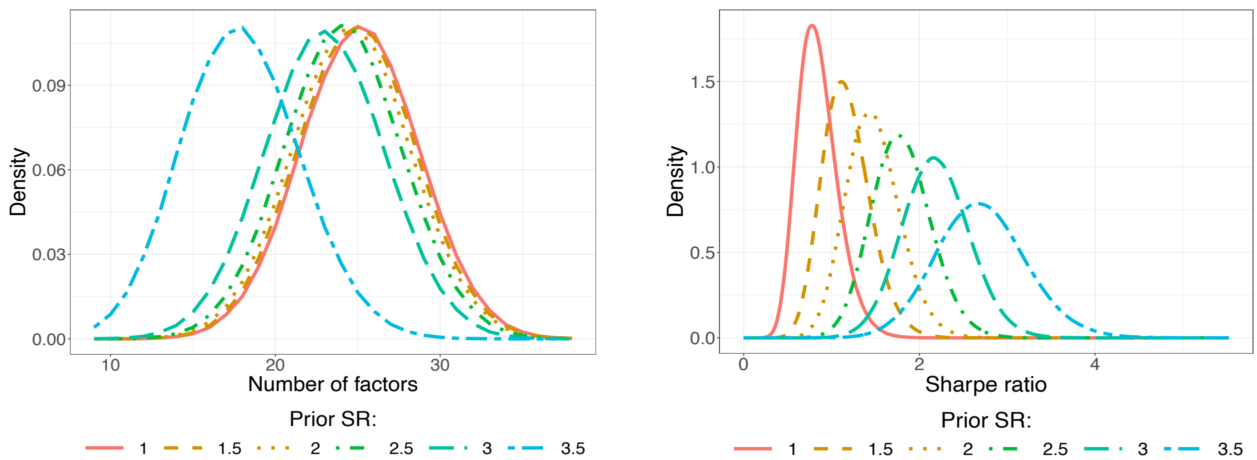
Figure 7: Posterior model probabilities of the 2,000 most likely models

Posterior model probabilities of the 2,000 most likely models computed using the continuous spike-and-slab of Section III.1.3, 51 factors and an annualized prior SR: $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]} = 2$ The horizontal axis uses a log scale. Sample: 1973:10–2016:12. Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12.

ification(s) is not a clear winner within the set of all possible models – its posterior probability is only about 0.011%. 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, we have 10 specifications with basically the same posterior probability, and the posterior model probability decays very slowly as we move down the list of most likely models: Moving from the best model, it takes more than a thousand models to reach the relative odds of 2:1 (i.e., to reduce the posterior probability by 50%). That is, to a first-order approximation, the frequentist likelihood ratio test of the best performing model versus the 1000^{th} one would yield a p-value of 30% at best (and a p-value of 15% after 2,000 models).

But how many of the factors proposed in the literature does it really take to price the cross-section? Thanks to our Bayesian method, this question can be easily answered. In fact, our framework is ideally suited for evaluating the assumption of sparsity (in observable factors) in cross-sectional asset pricing. 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) SDF being a function with that

arises due to the fact that the estimated model probability is simply the number of times that a given model is sampled by the Markov chain, divided by the total number of sampled models. Hence, models selected exactly the same number of times have identical posterior probability.



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

Figure 8: Posterior densities of model dimensionality and its implied Sharpe ratio

Left panel: posterior density of the true model having the number of factors listed on the horizontal axis. Right panel: posterior density of the (annualized) Sharpe ratio implied by the linear factor model for various values of the (annualized) prior Sharpe ratio. Sample: 1973:10–2016:12. Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12. 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.

number of factors.

Figure 8a reports the posterior distributions of the model dimensionality for various values of prior SR. These distributions are also summarized in Table IA20 of Internet Appendix IA.B.2.

For the most salient values of the prior SR (1–3), the posterior mean of the number of factors in the true model is in the 23–25 range, and the 95% posterior credible intervals are contained in the 16–32 factors range. That is, there is substantial evidence that the SDF is *dense* in the space of observable factors: 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.

It is worth noticing that, as Figure 8a shows, for very large prior SR, that is, with basically a flat prior for factors' price of risk, 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 price of risk can generate a "Bartlett Paradox" (see the discussion in Section III.1.1).

Our method allows to assess not only *how many* but also *which type* of factors we need to characterise the SDF in the economy. Table 5 reports the posterior numbers of tradable and non-tradable factors, as well as the associated estimation uncertainty around them. Strikingly, about one third (5-12) of the selected factors are non-tradable while the remaining ones (5-17)

Table 5: Posterior model dimensionality: Tradable versus non-tradable factors

	(a) Number of non-tradable factors						**(b)** Number of tradable factors					
	Total prior SR:						Total prior SR:					
	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
mean	8.45	8.45	8.45	8.45	8.46	8.25	16.68	16.48	16.17	15.67	14.56	9.66
median	8	8	8	8	8	8	17	16	16	16	15	10
5%	5	5	5	5	5	5	12	12	11	11	10	5
95th	12	12	12	12	12	12	21	21	21	21	20	15

Summary statistics for posterior distribution of the number of non-tradable (Panel (a)) and tradable (Panel (b)) factors in the SDF. Results tabulated for different values of the (annualized) prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2]}$. All the parameters are estimated over the 1973:10-2016:12 sample using a cross-section of 34 tradable factors plus 26 investment anomalies, computed using the continuous spike-and-slab approach of Section III.1.3 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.

are portfolio-based, suggesting complementarity between the two groups of factors in explaining asset returns.

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 (2020)). Since, given a model, the SDF-implied maximum Sharpe ratio is merely a function of the factors' price of risk and covariance matrix, our Bayesian method allows us 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 8b (and Table IA20b in Internet Appendix IA.B.2) reports the posterior distribution of the SDF-implied maximum Sharpe ratio (annualized) for several values of the prior SR. Except when a very strong shrinkage (small prior SR) is imposed (hence, Sharpe ratios are shrunk) the posterior distributions of the Sharpe ratio are quite similar for all prior values. 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 a prior SR $\in [1.5,\ 3]$ its posterior mean is about 1.17–2.19, and the 95% posterior credible intervals are in the 0.70–2.96 range.

Note that a model that is dense in the space of observable factors might be in principle sparse in the space of latent factors, for example, principle components. We address this issue by directly including principal components in the set of candidate factors. In particular, we consider the first five principal components of our cross-section of test assets, followed by a set of five "Risk Premia" principal components (RP-PC) of Lettau and Pelger (2020). In addition, to confirm that our method successfully handles weak identification, we add two artificially generated useless factors (independent of returns and i.i.d. distributed). Table 6 reports our findings.

Panel A of Table 6 shows that the first five principal components do not seem to capture priced risk: Their posterior probability is substantially lower than their prior probability, and their estimated market price of risk is zero (despite them explaining 61% of the time-series variation of returns). This is quite expected since standard principal components are not designed to capture cross-sectional pricing information.

Clearly, the artificially generated useless factors are successfully handled by the estimation procedure: As expected, their posterior probability remains at the prior level (50%), and their estimated price of risk is basically zero.

In Panel B we replace the canonical PCs with RP-PCs. We find strong support for two of them (first and third) capturing priced risk, while the other three have posterior probability below the prior value and prices of risk close to zero. Interestingly, even though some of the RP-PCs seem to successfully aggregate pricing information from the cross-section of returns (and factors, since the tradable ones are part of the test assets), they do not drive out the relevance of the robust stand-alone factors we identified earlier: BEAH_PEAD, CMA*, RMW*, among others. Consequently, the underlying SDF would be best described by a combination of both observable factors and (some) latent variables. Hence, the results in Panel B highlight that, in the quest of describing the sources of priced risk, there is scope for both selection and aggregation. This is confirmed by Figure IA3 in Internet Appendix IA.B.2, which shows that the most likely SDF is dense in the combined space of observable factors and principal components.

V.4 A Quest for Sparsity

The previous subsections suggest that only a small number of observable factors – BEH_PEAD, MKT, CMA*, and, to a lesser extent, STRev, RMW*, and BW_ISENT – are likely stand-alone explanators of the cross-section of asset returns. A natural question is whether the Bayesian factor posterior probabilities of Table 3 can help identify a low-dimensional benchmark model for pricing asset returns.

Table 7 reports the model posterior probabilities, that is, the probability of any of these models being the true data-generating process, for the SDFs built with the most likely factors and notable linear factor models. Posterior model probabilities (for all models) are computed using the closed-form solutions for the Dirac spike-and-slab prior method of Section III.1.2, giving us very precise estimates.

Strikingly, for *any* value of SR_{pr}, the best performing model is the one based on the most likely factors: Just three most likely factors (see Panel A), BEH_PEAD, MKT, and CMA*, are

Table 6: Observable factors versus Principal Components

	Factor inclusion prob., $\mathbb{E}[\gamma_j \vert \text{data}]$						Price of risk, $\mathbb{E}[\lambda_j \vert \text{data}]$					
	Total prior SR						Total prior SR					
Factors:	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
Panel A: Principal Components as Factors												
BEH_PEAD	0.547	0.602	0.678	0.766	0.840	0.814	0.015	0.036	0.073	0.132	0.220	0.287
MKT	0.508	0.542	0.573	0.598	0.607	0.504	0.015	0.035	0.064	0.100	0.149	0.182
CMA⋆	0.509	0.523	0.539	0.564	0.597	0.516	0.009	0.020	0.037	0.061	0.101	0.124
BW_ISENT	0.499	0.502	0.509	0.514	0.528	0.555	0.002	0.004	0.008	0.014	0.030	0.105
RMW⋆	0.500	0.499	0.514	0.537	0.568	0.450	0.007	0.017	0.032	0.057	0.097	0.107
STRev	0.495	0.503	0.522	0.546	0.555	0.435	0.006	0.016	0.030	0.052	0.083	0.089
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Useless I	0.499	0.499	0.501	0.498	0.498	0.497	0.000	0.000	0.000	0.000	0.001	0.006
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Useless II	0.496	0.495	0.495	0.494	0.498	0.500	0.000	0.000	0.001	0.001	0.002	0.010
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PC5	0.489	0.490	0.488	0.482	0.459	0.336	0.000	0.000	0.000	0.000	0.000	0.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PC4	0.497	0.487	0.480	0.471	0.451	0.322	0.000	0.000	0.000	0.000	0.000	0.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PC3	0.483	0.477	0.467	0.449	0.420	0.280	0.000	0.000	0.000	0.000	0.000	0.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PC1	0.478	0.467	0.457	0.437	0.399	0.248	0.000	0.000	0.000	0.000	-0.001	0.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PC2	0.473	0.455	0.444	0.429	0.397	0.249	0.000	0.000	0.000	0.000	0.000	0.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
LTRev	0.477	0.464	0.437	0.402	0.347	0.204	0.003	0.005	0.004	0.002	-0.002	-0.003
COMP_ISSUE	0.485	0.462	0.438	0.399	0.338	0.191	0.006	0.007	0.008	0.007	0.003	0.002
A_Growth	0.481	0.462	0.436	0.399	0.337	0.189	0.007	0.008	0.006	0.003	-0.002	-0.003
O_SCORE	0.473	0.450	0.425	0.385	0.323	0.186	-0.003	-0.005	-0.004	-0.002	-0.002	-0.003
Panel B: RP-Principal Components (Lettau and Pelger (2020)) as Factors												
RP-PC1	0.600	0.631	0.640	0.634	0.592	0.448	-0.016	-0.030	-0.043	-0.056	-0.066	-0.067
RP-PC3	0.548	0.597	0.645	0.661	0.651	0.529	-0.004	-0.009	-0.017	-0.024	-0.032	-0.035
BEH_PEAD	0.540	0.585	0.628	0.681	0.709	0.630	0.014	0.032	0.058	0.097	0.149	0.185
CMA⋆	0.510	0.523	0.542	0.571	0.616	0.531	0.009	0.020	0.037	0.062	0.104	0.129
RMW⋆	0.500	0.504	0.517	0.547	0.583	0.466	0.007	0.017	0.033	0.059	0.101	0.112
MKT	0.507	0.518	0.525	0.516	0.493	0.391	0.013	0.028	0.044	0.061	0.081	0.103
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Useless I	0.499	0.499	0.500	0.500	0.499	0.497	0.000	0.000	0.000	0.000	0.001	0.007
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Useless II	0.495	0.495	0.495	0.498	0.496	0.499	0.000	0.000	0.000	0.001	0.002	0.010
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
RP-PC5	0.481	0.487	0.488	0.484	0.459	0.338	0.001	0.003	0.005	0.008	0.011	0.012
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
RP-PC4	0.494	0.487	0.479	0.459	0.433	0.303	0.002	0.003	0.004	0.005	0.005	0.005
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
RP-PC2	0.479	0.464	0.458	0.439	0.403	0.267	0.000	-0.001	-0.001	-0.001	-0.001	0.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
COMP_ISSUE	0.483	0.464	0.438	0.406	0.338	0.193	0.006	0.008	0.010	0.009	0.004	0.002
A_Growth	0.483	0.466	0.443	0.396	0.337	0.196	0.007	0.007	0.005	0.000	-0.005	-0.007
LTRev	0.483	0.461	0.438	0.404	0.358	0.222	0.003	0.003	0.000	-0.006	-0.014	-0.015
O_SCORE	0.472	0.456	0.426	0.386	0.331	0.189	-0.003	-0.002	0.001	0.005	0.005	0.001

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, and posterior mean of factors' risk prices, $\mathbb{E}[\lambda_j|\text{data}]$, are computed using the continuous spike-and-slab approach of Section III.1.3 and 58 factors yielding 2^{58} models. The factors included are the 51 factors described in Table A1 of the Appendix plus two artificial i.i.d. useless factors, and five principal components. Panel A uses simple time-series principal components while Panel B uses the RP-PCs of Lettau and Pelger (2020). Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12. Results tabulated for different values of the (annualized) prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]}$.

Table 7: Posterior probabilities of notable models versus most likely factors

model: SR_{pr}:	*Panel A*: 3 most likely factors					*Panel B*: 6 most likely factors					*Panel C*: Most likely 5-factor model				
	1	1.5	2	2.5	3	1	1.5	2	2.5	3	1	1.5	2	2.5	3
Most likely factors	17.5%	24.9%	36.0%	48.8%	59.1%	17.8%	27.0%	44.0%	66.5%	83.7%	23.0%	35.3%	57.0%	77.6%	88.1%
CAPM	12.7%	12.5%	11.8%	11.3%	13.1%	12.7%	12.1%	10.3%	7.3%	5.2%	11.9%	10.8%	8.0%	5.0%	3.9%
FF3	10.3%	7.9%	5.3%	3.2%	1.7%	10.3%	7.7%	4.7%	2.1%	0.7%	9.6%	6.8%	3.6%	1.4%	0.5%
FF5	9.9%	7.0%	4.2%	2.1%	0.7%	9.8%	6.8%	3.7%	1.3%	0.3%	9.2%	6.0%	2.8%	0.9%	0.2%
Carhart	10.2%	7.8%	5.2%	2.9%	1.3%	10.2%	7.6%	4.6%	1.9%	0.5%	9.6%	6.7%	3.5%	1.3%	0.4%
q4	15.7%	17.8%	17.9%	14.9%	9.6%	15.6%	17.3%	15.7%	9.9%	3.9%	14.6%	15.3%	11.9%	6.4%	2.7%
Liq-CAPM	12.5%	12.0%	10.9%	9.6%	9.0%	12.5%	11.7%	9.5%	6.2%	3.6%	11.7%	10.4%	7.4%	4.3%	2.7%
FF3-QMJ	11.2%	10.1%	8.8%	7.4%	5.5%	11.1%	9.8%	7.7%	4.8%	2.1%	10.4%	8.6%	5.8%	3.1%	1.5%

Posterior model probabilities for the specifications in the first column, for different (annualized) prior Sharpe ratio values, computed using the Dirac spike-and-slab prior method of Section III.1.2. Panel A includes the factors BEH_PEAD, MKT, CMA*, while Panel B considers in addition STRev, RMW*, and BW_ISENT. Panel C uses the most likely 5-factor model according to the posterior probability. Factors are: MKT, MGMT, BAB, BEH_PEAD, CMA⋆ for $SR_{pr} = 1$; STRev, BAB, BEH_PEAD, RMW⋆, CMA⋆ for $SR_{pr} = 1.5$ to 2.5; BW_ISENT, BEH_PEAD, MKT⋆, RMW⋆, CMA⋆ for $SR_{pr} = 3$. Factors are described in Table A1 of the Appendix. Liq-CAPM stands for the liquidity-adjusted model of Pástor and Stambaugh (2003) and FF3-QMJ corresponds to the 4-factor model of Asness, Frazzini, and Pedersen (2019). Sample: 1973:10 to 2016:12. Test assets: 60 anomaly portfolios.

enough to outperform the most widely used empirical SDFs. This outperformance becomes even more pronounced when we consider the six most likely factors (see Panel B). Note that this drastic difference in performance understates the true power of our Bayesian approach to factor and model selection. Indeed, a subset of the most (individually) likely factors does not necessarily create the *most likely model*. Luckily, our approach can also be used to select the most likely model of any dimension. In particular, in Panel C we run the horse race between the most likely five-factor model that emerges using the Dirac Spike-and-Slab approach of Section III.1.2. Clearly, for all the values of prior SR, the best five-factor model outperforms not only all the notable models but also the combination of six overall most likely factors (from Panel B). While different prior SR may lead to different most likely low-dimensional models, the subset of selected factors is quite stable: All the specifications include BEH_PEAD and CMA*, while RMW* and BAB are selected four times out of five, and STRev is part of the most likely model three times out of five.

Our approach can also be used to formally evaluate the space of sparse factor models. In particular, in Table 8 we consider the universe of all the possible models that include no more than five factors, that is, 2.6 mln models. We evaluate *all* of those models individually, computing each of their marginal likelihoods following the Dirac spike-and-slab approach of Section III.1.2 (instead of sampling models, as in Section V.1). The table reports both posterior probabilities of the factor inclusion and their posterior price of risk. For simplicity, we consider the prior probability of a factor being included into the model being equal to 9.58% (since we have 51 factors total and each model with up to five factors is given equal ex ante probability).

First, three factors clearly stand out in Table 8: BEH_PEAD, BW_SENT, CMA⋆, all of which were also among the most likely factors in the SDF identified in the whole model space

Table 8: Posterior factor probabilities, $\mathbb{E}[\gamma_j|\text{data}]$, and risk prices: 2.6 million models

	Factor inclusion prob., $\mathbb{E}[\gamma_j\|\text{data}]$						Price of risk, $\mathbb{E}[\lambda_j\|\text{data}]$					
	Total prior SR:						Total prior SR:					
Factors:	0.5	1	1.5	2	2.5	3	0.5	1	1.5	2	2.5	3
BEH_PEAD	0.124	0.206	0.309	0.389	0.430	0.421	0.005	0.024	0.059	0.095	0.122	0.130
BW_ISENT	0.099	0.109	0.128	0.161	0.225	0.343	0.001	0.002	0.007	0.016	0.041	0.111
CMA⋆	0.104	0.122	0.136	0.141	0.137	0.120	0.003	0.010	0.017	0.023	0.026	0.025
BAB	0.112	0.133	0.140	0.136	0.125	0.105	0.005	0.014	0.021	0.024	0.025	0.022
DIV	0.097	0.102	0.109	0.121	0.141	0.183	0.000	0.000	0.001	0.002	0.005	0.016
HJTZ_ISENT	0.097	0.102	0.109	0.119	0.134	0.156	0.000	0.001	0.002	0.005	0.010	0.021
NONDUR	0.097	0.101	0.108	0.118	0.133	0.161	0.000	0.001	0.002	0.004	0.008	0.020
TERM	0.097	0.101	0.108	0.118	0.133	0.161	0.000	0.000	-0.001	-0.002	-0.005	-0.012
PE	0.097	0.101	0.108	0.117	0.132	0.160	0.000	0.000	-0.001	-0.002	-0.004	-0.012
FIN_UNC	0.097	0.101	0.108	0.117	0.131	0.149	0.000	0.001	0.002	0.004	0.008	0.017
UNRATE	0.097	0.101	0.107	0.116	0.130	0.154	0.000	0.000	0.001	0.002	0.005	0.013
DeltaSLOPE	0.097	0.101	0.107	0.116	0.129	0.154	0.000	0.000	0.001	0.002	0.003	0.010
IPGrowth	0.097	0.101	0.107	0.115	0.127	0.148	0.000	0.000	0.000	-0.001	-0.002	-0.006
DEFAULT	0.097	0.101	0.107	0.115	0.127	0.146	0.000	0.000	0.001	0.001	0.003	0.007
SERV	0.096	0.101	0.106	0.114	0.126	0.146	0.000	0.000	0.000	0.001	0.002	0.006
REAL_UNC	0.096	0.100	0.106	0.114	0.125	0.141	0.000	0.000	0.000	0.001	0.002	0.004
STRev	0.095	0.098	0.105	0.116	0.123	0.109	0.001	0.005	0.010	0.016	0.022	0.022
MACRO_UNC	0.096	0.100	0.106	0.113	0.122	0.136	0.000	0.000	0.000	0.000	0.001	0.000
Oil	0.096	0.100	0.105	0.111	0.119	0.129	0.000	0.000	0.000	0.000	0.001	0.002
MKT⋆	0.097	0.101	0.104	0.105	0.103	0.105	0.002	0.005	0.009	0.013	0.015	0.020
RMW⋆	0.096	0.098	0.102	0.106	0.103	0.083	0.002	0.006	0.011	0.015	0.018	0.016
LIQ_NT	0.096	0.098	0.100	0.102	0.101	0.096	0.000	0.000	0.001	0.001	0.002	0.003
MKT	0.094	0.099	0.103	0.103	0.095	0.080	0.003	0.009	0.014	0.018	0.020	0.018
ROE	0.107	0.113	0.106	0.093	0.078	0.060	0.006	0.013	0.016	0.017	0.015	0.012
MGMT	0.109	0.109	0.101	0.092	0.080	0.061	0.007	0.014	0.017	0.018	0.017	0.013
NetOA	0.098	0.102	0.101	0.094	0.084	0.068	0.002	0.005	0.008	0.010	0.010	0.009
IA	0.108	0.108	0.099	0.089	0.077	0.060	0.006	0.013	0.015	0.016	0.015	0.012
HML⋆	0.099	0.101	0.096	0.087	0.075	0.058	0.003	0.007	0.010	0.011	0.011	0.009
LIQ_TR	0.095	0.095	0.093	0.087	0.078	0.063	0.001	0.002	0.004	0.006	0.006	0.005
INTERM_CAP_RATIO	0.093	0.090	0.087	0.083	0.075	0.062	0.001	0.004	0.006	0.008	0.009	0.008
INV_IN_ASS	0.098	0.097	0.090	0.079	0.067	0.051	0.003	0.006	0.009	0.009	0.008	0.007
PERF	0.096	0.091	0.082	0.071	0.059	0.044	0.003	0.007	0.009	0.009	0.008	0.006
STOCK_ISS	0.098	0.092	0.081	0.070	0.058	0.043	0.004	0.008	0.009	0.009	0.008	0.006
ACCR	0.093	0.087	0.079	0.070	0.060	0.048	0.001	0.002	0.004	0.004	0.004	0.004
BEH_FIN	0.099	0.089	0.077	0.067	0.057	0.043	0.005	0.009	0.010	0.010	0.009	0.007
QMJ	0.095	0.086	0.076	0.066	0.055	0.040	0.004	0.008	0.010	0.010	0.009	0.007
UMD	0.094	0.087	0.076	0.065	0.055	0.043	0.002	0.004	0.005	0.005	0.004	0.003
SMB⋆	0.092	0.083	0.073	0.063	0.052	0.039	0.001	0.003	0.004	0.004	0.004	0.003
HML_DEVIL	0.085	0.073	0.067	0.066	0.062	0.050	0.002	0.004	0.006	0.009	0.011	0.010
CMA	0.095	0.084	0.071	0.059	0.049	0.037	0.004	0.007	0.007	0.006	0.005	0.004
SKEW	0.089	0.081	0.071	0.060	0.048	0.036	0.002	0.005	0.006	0.006	0.005	0.004
ASS_Growth	0.093	0.081	0.068	0.057	0.047	0.035	0.003	0.005	0.005	0.005	0.004	0.003
COMP_ISSUE	0.091	0.077	0.065	0.055	0.046	0.034	0.003	0.004	0.005	0.004	0.004	0.003
LTRev	0.089	0.076	0.064	0.053	0.043	0.032	0.001	0.003	0.003	0.003	0.003	0.002
RMW	0.088	0.074	0.062	0.052	0.043	0.032	0.002	0.003	0.004	0.004	0.003	0.003
ROA	0.089	0.075	0.062	0.051	0.041	0.030	0.002	0.004	0.004	0.004	0.003	0.002
GR_PROF	0.087	0.073	0.061	0.051	0.042	0.031	0.000	0.001	0.000	0.000	0.000	0.000
SMB	0.086	0.073	0.061	0.051	0.040	0.029	0.002	0.004	0.004	0.004	0.004	0.002
DISSTR	0.084	0.069	0.059	0.051	0.043	0.033	0.001	0.000	-0.001	-0.002	-0.003	-0.002
HML	0.086	0.070	0.057	0.048	0.039	0.030	0.002	0.003	0.003	0.003	0.003	0.002
O_SCORE	0.084	0.069	0.056	0.045	0.036	0.027	-0.001	-0.002	-0.002	-0.002	-0.001	-0.001

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, and posterior mean of factors' risk prices, $\mathbb{E}[\lambda_j|\text{data}]$, are computed using the the Dirac spike-and-slab approach of Section III.1.2 and 51 factors described in Table A1 of Appendix. Sample: 1973:10-2016:12. Test assets: 34 tradable factors and 26 investment anomalies. Prior probability of a factor being included is about 9.58% since we give each possible model equal prior probability and a factor could be included in a model with up to four other variables. Results tabulated for different values of the (annualized) prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR^2_{\boldsymbol{f}} \mid \sigma^2]}$. Light-shaded grey rows denote non-tradable factors.

of Table 3. Second, and strikingly, there is a large set of factors that have posterior probability of inclusion above the prior, providing support for them being included in a low-dimensional model. This group includes not only the other robust factors identified in Section V.1 but also 40% of both tradable and non-tradable macro-factors, such as nondurable consumption, unemployment, and industrial production growth. This second finding is consistent with our results in Section V.3, where we showed that many factors seem to load on the same underlying sources of economic risks: Sparse models, therefore, tend to rely on them almost interchangeably. This is further illustrated in Figure IA4 of the Internet Appendix, which depicts posterior probabilities for the top 2,000 sparse models under the (annualized) SR prior of 2. Similar to our findings in Section V.3, the space of best performing models is quite flat, with their corresponding posterior probability decaying slowly. In fact, up to a first-order approximation, the frequentist likelihood ratio test of the best performing model versus the 100^{th} (1000^{th}) specification would yield a p-value of 19.0% (9.2%) at best. Interestingly, as outlined in Table IA22 of the Internet Appendix, non-tradable factors are very salient for low-dimensional models: the overwhelming majority, 67%-99%, of the (top 10%) best-performing sparse SDFs include at least one non-tradable factor.

Our findings indicate that low-dimensional models with observable factors are likely to be severely misspecified, and in many cases reflect noisy measures of the same underlying economic risks. While some of the factors still stand alone as significant drivers of the cross-section of asset returns, the true latent SDF is still best approximated by an efficient aggregation of many underlying variables, provided by the BMA. To further validate this point, we have performed an OOS analysis (in both time-series and cross-sectional dimension) of the BMA versus the best low-dimensional models and found that the former strongly outperforms the latter.

VI Conclusions and Extensions

We develop a novel (Bayesian) method for the analysis of linear factor models in asset pricing. This approach can handle quadrillions of models generated by the zoo of traded and non-traded factors and delivers inference that is robust to the common identification failures caused by weak and level factors.

We apply our approach to the study of more than 2 quadrillion factor model specifications and find that: 1) only a handful of factors seem to be robust explanators of the cross-section of asset returns; 2) jointly, the three to six robust factors provide a model that substantially outperforms notable benchmarks; 3) nevertheless, with very high probability the "true" latent SDF is dense in the space of factors proposed in the previous literature, likely containing 23–25 observable factors; and 4) a BMA over the universe of possible models delivers a novel benchmark SDF for in- and out-of-sample empirical asset pricing.

Our method can be feasibly modified to accommodate several salient extensions. First, one might want to bound the maximum price of risk (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 centred) 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 weak (and level) factors based on their correlation with asset returns, could also be 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).[41] The rationale for 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, the posterior eventually reverts toward the prior. Nevertheless, note that this mechanism (first pointed out by Jeffreys (1961)) is actually *desirable* in our settings in order to shrink the price of risk of useless factors toward zero.[42]

Second, thanks to its hierarchical structure, our approach can formally handle the statistical uncertainty caused by generated factors, for example, mimicking portfolios, and provides valid inference in their presence. Furthermore, it can accommodate a wide range of both priced and unpriced latent factors.

Third, thanks to the hierarchical structure of our method, time-varying expected returns and SDF factor loadings could be accommodated by adopting the time-varying parameter approach of Primiceri (2005). Furthermore, although this would significantly increase the numerical complexity of the cross-sectional inference step, the time-varying parameters formulation could also be used for modeling time-varying factor risk prices.

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.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61, 259–299.

Asness, C. S., A. Frazzini, and L. H. Pedersen (2019): "Quality minus Junk," *Review of Accounting Studies*, 24(1), 34–112.

Avramov, D. (2002): "Stock Return Predictability and Model Uncertainty," *Journal of Financial Economics*, 64, 423–458.

——— (2004): "Stock Return Predictability and Asset Pricing Models," *Review of Financial Studies*, 17, 699–738.

[41]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 equations (22) and (23).

[42]Since useless factors tend to generate heavy-tailed likelihoods (in the limit, the likelihood is an improper "uniform" on $\mathbb{R}$), with peaks for price of risk that deviate toward infinity, the posterior price of risk of such factors is shrunk toward the prior (zero) mean if the prior has thin tails.

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

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.

Basu, S. (1977): "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratio: A Test of the Efficient Market Hypothesis," *Journal of Finance*, 32, 663–82.

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.

Bhandari, L. (1988): "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence," *The Journal of Finance*, 43, 507–528.

Bondt, W. F. M. D., and R. Thaler (1985): "Does the Stock Market Overreact?," *The Journal of Finance*, 40, 793–805.

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

Bryzgalova, S., J. Huang, and C. Julliard (2022): "Bayesian Fama-MacBeth," mimeo.

Bustamante, M. C., and A. Donangelo (2017): "Product Market Competition and Industry Returns," *The Review of Financial Studies*, 30, 4216–4266.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52(1), 57–82.

Chen, L., M. Pelger, and J. Zhu (2019): "Deep Learning in Asset Pricing," working paper.

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*, 75(1), 551–577.

Chung, K. H., and H. Zhang (2014): "A Simple Approximation of Intraday Spreads Using Daily Data," *Journal of Financial Markets*, 17, 94–120.

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.

D'Acunto, F., R. Liu, C. Pflueger, and M. Weber (2018): "Flexible Prices and Leverage," *The Journal of Financial Economics*, 129, 46–68.

Daniel, K., D. Hirshleifer, and L. Sun (2020): "Short- and Long-Horizon Behavioral Factors," *Review of Financial Studies*, 33, 1673–1736.

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

Daniel, K., and S. Titman (2012): "Testing Factor-Model Explanations of Market Anomalies," *Critical Finance Review*, 1(1), 103–139.

Desai, H., S. Rajgopal, and M. Venkatachalam (2004): "Value-Glamour and Accruals Mispricing: One Anomaly or Two?," *The Accounting Review*, 79, 355–385.

Fama, E. F., and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Fama, E. F., and K. R. French (2015): "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116, 1–22.

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

Freyberger, J., A. Neuhierl, and M. Weber (2020): "Dissecting Characteristics Nonparametrically," *The Review of Financial Studies*, 33, 2326–2377.

Garfinkel, J. (2009): "Measuring Investors' Opinion Divergence," *Journal of Accounting Research*, 47, 1317–1348.

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.

GEORGE, T. J., AND C.-Y. HWANG (2004): "The 52-Week High and Momentum Investing," *The Journal of Finance*, 59, 2145–2176.

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

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons.

GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2021): "Economic Predictions with Big Data: The Illusion of Sparsity," *Econometrica*, forthcoming.

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 (2020): "Taming the Factor Zoo: A Test of New Factors," *Journal of Finance*, 75, 1327–1370.

GIGLIO, S., AND D. XIU (2021): "Asset Pricing with Omitted Factors," *Journal of Political Economy*, forthcoming.

GIGLIO, S., D. XIU, AND D. ZHANG (2021): "Test Assets and Weak Factors," working paper, available at SSRN: https://ssrn.com/abstract=3768081.

GORODNICHENKO, Y., AND M. WEBER (2016): "Are Sticky Prices Costly? Evidence from the Stock Market," *American Economic Review*, 106(1), 165–99.

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.

GOSPODINOV, N., AND C. ROBOTTI (2021a): "Common Pricing across Asset Classes: Empirical Evidence Revisited," *Journal of Financial Economics*, 140, 292–324.

——— (2021b): "Common Pricing across Asset Classes: Empirical Evidence Revisited," working paper.

GU, S., B. KELLY, AND D. XIU (2020): "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, 33(5), 2223–2273.

HANSEN, L. P. (1982): "Large Sample Properties of Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

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.

HAUGEN, R. A., AND N. L. BAKER (1996): "Commonality in the determinants of expected stock returns," *Journal of Financial Economics*, 41(3), 401–439.

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.

HOETING, J. A., D. MADIGAN, A. E. RAFTERY, AND C. T. VOLINSKY (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14(4), 382–401.

HOU, K., G. KAROLYI, AND B.-C. KHO (2011): "What Factors Drive Global Stock Returns?," *Review of Financial Studies*, 24(8), 2527–2574.

HOU, K., C. XUE, AND L. ZHANG (2015): "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, 28(3), 650–705.

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.

JAROCIŃSKI, M., AND A. MARCET (2019): "Priors about Observables in Vector Autoregressions," *Journal of Econometrics*, 209(2), 238–255.

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

KALDOR, N. (1966): "Marginal Productivity and the Macro-Economic Theories of Distribution: Comment on Samuelson and Modigliani," *The Review of Financial Studies*, 33, 309–319.

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.

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-squared's," *Journal of Econometrics*, 189(1), 101–116.

——— (2020): "Robust Inference for Consumption-Based Asset Pricing," *Journal of Finance*, 75(1), 507–550.

Kozak, S., S. Nagel, and S. Santosh (2020): "Shrinking the Cross-Section," *Journal of Financial Economics*, 135, 271–292.

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

Lettau, M., and M. Pelger (2020): "Factors That Fit the Time Series and Cross-Section of Stock Returns," *The Review of Financial Studies*, 33(5), 2274–2325.

Lewellen, J. (2015): "The Cross-Section of Expected Stock Returns," *Critical Finance Review*, 4, 1–44.

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

Linnainmaa, J. T., and M. R. Roberts (2018): "The History of the Cross-Section of Stock Returns," *The Review of Financial Studies*, 31(7), 2606–2649.

Litzenberger, R. H., and K. Ramaswamy (1979): "The Effect of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical evidence," *The Journal of Financial Economics*, 7, 163–195.

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

Lyandres, E., L. Sun, and L. Zhang (2008): "The new issues puzzle: Testing the investment- based explanation," *The Review of Financial Studies*, 21, 2925–2855.

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.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F.Engle, and D. McFadden, vol. 4. Elsevier Press.

Novy-Marx, R. (2011): "Operating Leverage," *The Review of Finance*, 15, 103–134.

——— (2012): "Is Momentum Really Momentum?," *The Journal of Financial Economics*, 103, 429–453.

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

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.

Pástor, L., and R. F. Stambaugh (2003): "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 111(3), 642–685.

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.

Raftery, A. E., and Y. Zheng (2003): "Discussion: Performance of Bayesian Model Averaging," *Journal of the American Statistical Association*, 98, 931–938'.

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

Schervish, M. J. (1995): *Theory of Statistics*, Springer Series in Statistics. Springer-Verlag.

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

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

Soliman, M. T. (2008): "The Use of DuPont Analysis by Market Participants," *The Accounting Review*, 83, 823–853.

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.

A Appendix

A.1 Additional Derivations and Proofs

A.1.1 Derivation of the posterior distributions in Section III

Let's consider first the time series layer of our hierarchical model. We assume that $\boldsymbol{Y_t} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$. The likelihood function of the observed time-series data $\mathbf{Y} = \{\boldsymbol{Y_t}\}_{t=1}^T$ is

$$
\begin{aligned}
p(\mathbf{Y}|\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) &\propto |\boldsymbol{\Sigma_Y}|^{-\frac{T}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma_Y}^{-1}\sum_{t=1}^T(\boldsymbol{Y_t}-\boldsymbol{\mu_Y})(\boldsymbol{Y_t}-\boldsymbol{\mu_Y})^\top]} \\
&\propto |\boldsymbol{\Sigma_Y}|^{-\frac{T}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma_Y}^{-1}\sum_{t=1}^T(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})^\top + T\boldsymbol{\Sigma_Y}^{-1}(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})^\top]},
\end{aligned}
$$

where $\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}} = \frac{1}{T}\sum_{t=1}^T \boldsymbol{Y_t}$. After assigning a diffuse prior for $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$, $\pi(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{p+1}{2}}$, the posterior distribution function of $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$ is

$$
p(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}|\mathbf{Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{T+p+1}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma_Y}^{-1}\sum_{t=1}^T(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})^\top + T\boldsymbol{\Sigma_Y}^{-1}(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})^\top]}.
$$

Hence, the posterior distribution of $\boldsymbol{\mu_Y}$ conditional on $\boldsymbol{Y}$ and $\boldsymbol{\Sigma_Y}$ is

$$
p(\boldsymbol{\mu_Y}|\mathbf{Y}, \boldsymbol{\Sigma_Y}) \propto e^{-\frac{1}{2}tr[T\boldsymbol{\Sigma_Y}^{-1}(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})^\top]},
$$

and the above is the kernel of the multivariate normal in equation (6). If we further integrate out $\boldsymbol{\mu_Y}$, it is easy to show that $p(\boldsymbol{\Sigma_Y}|\mathbf{Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{T+p}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma_Y}^{-1}\sum_{t=1}^T(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}})^\top]}$. Therefore, the posterior distribution of $\boldsymbol{\Sigma}$ is the inverse-Wishart in equation (7).

Recall that $\boldsymbol{C} = (\boldsymbol{1_N}, \boldsymbol{C_f})$, $\boldsymbol{\lambda}^\top = (\lambda_c, \boldsymbol{\lambda_f}^\top)$. Assuming $\alpha_i \sim$ iid $\mathcal{N}(0, \sigma^2)$, the cross-sectional likelihood function conditional on the time-series parameters ($\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$), $p(data|\boldsymbol{\lambda}, \sigma^2)$, is given in equation (10), where "data" in this cross-sectional (second) step include the observed time-series $\mathbf{Y} = \{\boldsymbol{Y_t}\}_{t=1}^T$, as well as $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ drawn from the time-series step.

Assuming the diffuse prior $\pi(\boldsymbol{\lambda}, \sigma^2) \propto \sigma^{-2}$, the posterior distribution of $(\boldsymbol{\lambda}, \sigma^2)$ is

$$
\begin{aligned}
p(\boldsymbol{\lambda}, \sigma^2|data) &\propto (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\lambda})^\top(\boldsymbol{\mu_R}-\boldsymbol{C\lambda})} = (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}}+\boldsymbol{C}(\boldsymbol{\hat{\lambda}}-\boldsymbol{\lambda}))^\top(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}}+\boldsymbol{C}(\boldsymbol{\hat{\lambda}}-\boldsymbol{\lambda}))} \\
&= (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})^\top(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})-\frac{1}{2\sigma^2}(\boldsymbol{\lambda}-\boldsymbol{\hat{\lambda}})^\top\boldsymbol{C}^\top\boldsymbol{C}(\boldsymbol{\lambda}-\boldsymbol{\hat{\lambda}})}, \text{ and} \\
\therefore\ \ p(\boldsymbol{\lambda}|\sigma^2, data) &\propto e^{-\frac{(\boldsymbol{\lambda}-\boldsymbol{\hat{\lambda}})^\top\boldsymbol{C}^\top\boldsymbol{C}(\boldsymbol{\lambda}-\boldsymbol{\hat{\lambda}})}{2\sigma^2}},
\end{aligned}
$$

where $\boldsymbol{\hat{\lambda}} = (\boldsymbol{C}^\top\boldsymbol{C})^{-1}\boldsymbol{C}^\top\boldsymbol{\mu_R}$, $\hat{\sigma}^2 = \frac{(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})^\top(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})}{N}$, and the above is the kernel of a Gaussian distribution. Note that sending $\sigma^2 \to 0$ the posterior $p(\boldsymbol{\lambda}|\sigma^2, data)$ is proportional to a Dirac at $\boldsymbol{\hat{\lambda}}$ as per Definition 1. For non-degenerate values of σ^2, the conditional posterior of $\boldsymbol{\lambda}$ is instead the one in equation (11). We derive the posterior of σ^2 by integrating out $\boldsymbol{\lambda}$ in the joint posterior, $p(\sigma^2|data) = \int p(\boldsymbol{\lambda}, \sigma^2|data)d\boldsymbol{\lambda} \propto (\sigma^2)^{-\frac{N-K+1}{2}} e^{-\frac{N\hat{\sigma}^2}{2\sigma^2}}$, hence, obtaining equation (12).

Under the GLS distributional assumption, $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{0_N}, \sigma^2\boldsymbol{\Sigma_R})$, where $\boldsymbol{\Sigma_R}$ is the covariance matrix of returns $\boldsymbol{R_t}$. The posterior of $(\boldsymbol{\lambda}, \sigma^2)$ is then

$$
\begin{aligned}
p(\boldsymbol{\lambda}, \sigma^2|data) &\propto (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\lambda})^\top\boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R}-\boldsymbol{C\lambda})} \\
&= (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})^\top\boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})-\frac{1}{2\sigma^2}(\boldsymbol{\lambda}-\boldsymbol{\hat{\lambda}})^\top\boldsymbol{C}^\top\boldsymbol{\Sigma_R^{-1}}\boldsymbol{C}(\boldsymbol{\lambda}-\boldsymbol{\hat{\lambda}})}, \text{ and} \\
\therefore\ \ p(\boldsymbol{\lambda}|\sigma^2, data) &\propto e^{-\frac{(\boldsymbol{\lambda}-\boldsymbol{\hat{\lambda}})^\top\boldsymbol{C}^\top\boldsymbol{\Sigma_R^{-1}}\boldsymbol{C}(\boldsymbol{\lambda}-\boldsymbol{\hat{\lambda}})}{2\sigma^2}},
\end{aligned}
$$

where $\boldsymbol{\hat{\lambda}} = (\boldsymbol{C}^\top\boldsymbol{\Sigma_R^{-1}}\boldsymbol{C})^{-1}\boldsymbol{C}^\top\boldsymbol{\Sigma_R^{-1}}\boldsymbol{\mu_R}$ and the above is the kernel of a Gaussian distribution. Note that sending $\sigma^2 \to 0$, the posterior $p(\boldsymbol{\lambda}|\sigma^2, data)$ is proportional to a Dirac at $\boldsymbol{\hat{\lambda}}$ as per Definition 2. For non-degenerate values of σ^2 the conditional posterior of $\boldsymbol{\lambda}$ is instead the one in equation (13). Further integrating out $\boldsymbol{\lambda}$, we obtain $p(\sigma^2|data) = \int p(\boldsymbol{\lambda}, \sigma^2|data)d\boldsymbol{\lambda} \propto (\sigma^2)^{-\frac{N-K+1}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})^\top\boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})}$. Hence, the posterior of σ^2 is as in equation (14).

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

Following the derivation in Section A.1.1, the cross-sectional likelihood is given by equation (10). Assigning a flat prior to the parameters[43] $(\boldsymbol{\lambda}, \sigma^2)$, the marginal cross-sectional likelihood function conditional on model index $\boldsymbol{\gamma}$ is

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

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

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

A.1.3 Proof of Proposition 2

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

$$
\begin{aligned}
p(\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}) &\propto p(data|\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma}) \\
&\propto (2\pi)^{-\frac{p_\gamma}{2}} |\boldsymbol{D}_\gamma|^{\frac{1}{2}} (\sigma^2)^{-\frac{N+p_\gamma}{2}} e^{-\frac{1}{2\sigma^2}[(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma)^\top (\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma) + \boldsymbol{\lambda}_\gamma^\top \boldsymbol{D}_\gamma \boldsymbol{\lambda}_\gamma]} \\
&= (2\pi)^{-\frac{p_\gamma}{2}} |\boldsymbol{D}_\gamma|^{\frac{1}{2}} (\sigma^2)^{-\frac{N+p_\gamma}{2}} e^{-\frac{(\boldsymbol{\lambda}_\gamma - \hat{\boldsymbol{\lambda}}_\gamma)^\top (\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)(\boldsymbol{\lambda}_\gamma - \hat{\boldsymbol{\lambda}}_\gamma)}{2\sigma^2}} e^{-\frac{SSR_\gamma}{2\sigma^2}},
\end{aligned}
$$

where $SSR_\gamma = \boldsymbol{\mu}_{\boldsymbol{R}}^\top \boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{\mu}_{\boldsymbol{R}}^\top \boldsymbol{C}_\gamma (\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1} \boldsymbol{C}_\gamma^\top \boldsymbol{\mu}_{\boldsymbol{R}} = \min_{\boldsymbol{\lambda}_\gamma}\{(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma)^\top (\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma) + \boldsymbol{\lambda}_\gamma^\top \boldsymbol{D}_\gamma \boldsymbol{\lambda}_\gamma\}$. Hence, defining $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1} \boldsymbol{C}_\gamma^\top \boldsymbol{\mu}_{\boldsymbol{R}}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma) = \sigma^2 (\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1}$, we obtain the posterior distribution in (16).

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

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

Sampling σ^2. From Bayes theorem, the posterior of σ^2 is $p(\sigma^2|data, \boldsymbol{\gamma}) \propto p(data|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2) \propto (\sigma^2)^{-\frac{N}{2}-1} e^{-\frac{SSR_\gamma}{2\sigma^2}}$. 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^2 as follows:

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

where $SSR_\gamma = \boldsymbol{\mu}_{\boldsymbol{R}}^\top \boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{\mu}_{\boldsymbol{R}}^\top \boldsymbol{C}_\gamma (\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1} \boldsymbol{C}_\gamma^\top \boldsymbol{\mu}_{\boldsymbol{R}}$.

A.1.4 Proof of Proposition 3

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

$$
\begin{aligned}
p(\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}) &\propto p(data|\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma}) \\
&\propto (2\pi)^{-\frac{p_\gamma}{2}} |\boldsymbol{D}_\gamma|^{\frac{1}{2}} (\sigma^2)^{-\frac{N+p_\gamma}{2}} e^{-\frac{1}{2\sigma^2}[(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma)^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} (\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma) + \boldsymbol{\lambda}_\gamma^\top \boldsymbol{D}_\gamma \boldsymbol{\lambda}_\gamma]} \\
&= (2\pi)^{-\frac{p_\gamma}{2}} |\boldsymbol{D}_\gamma|^{\frac{1}{2}} (\sigma^2)^{-\frac{N+p_\gamma}{2}} e^{-\frac{(\boldsymbol{\lambda}_\gamma - \hat{\boldsymbol{\lambda}}_\gamma)^\top (\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)(\boldsymbol{\lambda}_\gamma - \hat{\boldsymbol{\lambda}}_\gamma)}{2\sigma^2}} e^{-\frac{SSR_\gamma}{2\sigma^2}},
\end{aligned}
$$

where $SSR_\gamma = \min_{\boldsymbol{\lambda}_\gamma}\{(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma)^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} (\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_\gamma \boldsymbol{\lambda}_\gamma) + \boldsymbol{\lambda}_\gamma^\top \boldsymbol{D}_\gamma \boldsymbol{\lambda}_\gamma\}$. Hence, defining $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1} \boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{\mu}_{\boldsymbol{R}}$, $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma) = \sigma^2 (\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1} \boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma)^{-1}$, we obtain the posterior distribution in (19).

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

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

Obviously, the posterior distribution of σ^2 is identical to that in equation (20).

Finally, we obtain the marginal likelihood of the data in (21) by integrating out σ^2 as follows:

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

A.1.5 Proof of Corollary 1

To begin with, we introduce the following matrix notations:

$$\boldsymbol{C_\gamma} = (\boldsymbol{C_{\gamma'}},\boldsymbol{C_p}),\;\; \boldsymbol{D_\gamma} = \begin{pmatrix} \boldsymbol{D_{\gamma'}} & \boldsymbol{0} \\ \boldsymbol{0} & \psi_p^{-1}\end{pmatrix},$$

where $\boldsymbol{0}$ denotes conformable matrices of zeros.

Under the spherical (OLS) distributional assumption for pricing errors $\boldsymbol{\alpha}$,

$$\boldsymbol{C_\gamma^\top C_\gamma} + \boldsymbol{D_\gamma} = \begin{pmatrix} \boldsymbol{C_{\gamma'}^\top C_{\gamma'}} + \boldsymbol{D_{\gamma'}} & \boldsymbol{C_{\gamma'}^\top C_p} \\ \boldsymbol{C_p^\top C_{\gamma'}} & \boldsymbol{C_p^\top C_p} + \psi_p^{-1}\end{pmatrix},$$

$|\boldsymbol{C_\gamma^\top C_\gamma} + \boldsymbol{D_\gamma}| = |\boldsymbol{C_{\gamma'}^\top C_{\gamma'}} + \boldsymbol{D_{\gamma'}}| \times |\boldsymbol{C_p^\top C_p} + \psi_p^{-1} - \boldsymbol{C_p^\top C_{\gamma'}}(\boldsymbol{C_{\gamma'}^\top C_{\gamma'}} + \boldsymbol{D_{\gamma'}})^{-1}\boldsymbol{C_{\gamma'}^\top C_p}|$, and $|\boldsymbol{D_\gamma}| = |\boldsymbol{D_{\gamma'}}| \times \psi_p^{-1}$. Equipped with the above, we have by direct calculation

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

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

Similar to the above, in the non-spherical (GLS) pricing errors case we have

$$\boldsymbol{C_\gamma^\top \Sigma_R^{-1} C_\gamma} + \boldsymbol{D_\gamma} = \begin{pmatrix} \boldsymbol{C_{\gamma'}^\top \Sigma_R^{-1} C_{\gamma'}} + \boldsymbol{D_{\gamma'}} & \boldsymbol{C_{\gamma'}^\top \Sigma_R^{-1} C_p} \\ \boldsymbol{C_p^\top \Sigma_R^{-1} C_{\gamma'}} & \boldsymbol{C_p^\top \Sigma_R^{-1} C_p} + \psi_p^{-1}\end{pmatrix},$$

$|\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma| = |\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'} + \boldsymbol{D}_{\gamma'}| \times |\boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}} + \frac{1}{\psi_p} - \boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'}(\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'} + \boldsymbol{D}_{\gamma'})^{-1}\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}}|$, and $|\boldsymbol{D}_{\boldsymbol{\gamma}}| = |\boldsymbol{D}_{\boldsymbol{\gamma'}}| \times \psi_p^{-1}$. Equipped with the above, we have by direct calculation

$$\frac{p(data|\gamma_j=1,\boldsymbol{\gamma_{-j}})}{p(data|\gamma_j=0,\boldsymbol{\gamma_{-j}})} = \frac{|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}}{|\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_\gamma + \boldsymbol{D}_\gamma|^{\frac{1}{2}}}\frac{1}{(SSR_\gamma/2)^{\frac{N}{2}}} \Bigg/ \frac{|\boldsymbol{D}_{\boldsymbol{\gamma'}}|^{\frac{1}{2}}}{|\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'} + \boldsymbol{D}_{\gamma'}|^{\frac{1}{2}}}\frac{1}{(SSR_{\gamma'}/2)^{\frac{N}{2}}}$$
$$= \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \psi_p^{-\frac{1}{2}} \left|\boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}} + \frac{1}{\psi_p} - \boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'}\left(\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'} + \boldsymbol{D}_{\gamma'}\right)^{-1}\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}}\right|^{-\frac{1}{2}}$$
$$= \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \left|1 + \psi_p\left[\boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}} - \boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'}\left(\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'} + \boldsymbol{D}_{\gamma'}\right)^{-1}\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}}\right]\right|^{-\frac{1}{2}},$$

where $\boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}} - \boldsymbol{C}_{\boldsymbol{p}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'}\left(\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma'} + \boldsymbol{D}_{\gamma'}\right)^{-1}\boldsymbol{C}_{\boldsymbol{\gamma'}}^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{p}} = \min_{\boldsymbol{b}}\{(\boldsymbol{C}_{\boldsymbol{p}} - \boldsymbol{C}_{\boldsymbol{\gamma'}}\boldsymbol{b})^\top \boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{C}_{\boldsymbol{p}} - \boldsymbol{C}_{\boldsymbol{\gamma'}}\boldsymbol{b}) + \boldsymbol{b}^\top \boldsymbol{D}_{\boldsymbol{\gamma'}}\boldsymbol{b}\}$, which is the minimal value of the penalized sum of squared errors when we use $\boldsymbol{C}_{\boldsymbol{\gamma'}}$ to predict $\boldsymbol{C}_{\boldsymbol{p}}$, but the prediction errors are weighted by $\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}$.

A.1.6 Proof of Propositions 5 and 6

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

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

Therefore, defining $\hat{\boldsymbol{\lambda}} = (\boldsymbol{C}^\top\boldsymbol{C}+\boldsymbol{D})^{-1}\boldsymbol{C}^\top\boldsymbol{\mu}_{\boldsymbol{R}}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}) = \sigma^2(\boldsymbol{C}^\top\boldsymbol{C}+\boldsymbol{D})^{-1}$, we have the posterior in equation (28).

Sampling $\{\gamma_j\}_{j=1}^K$. Given a ω_j, the conditional Bayes factor for the j-th risk factor is[44]

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

Sampling $\boldsymbol{\omega}$. From Bayes' theorem we have $p(\omega_j|data,\boldsymbol{\lambda},\boldsymbol{\gamma},\sigma^2) \propto \pi(\omega_j)\pi(\gamma_j|\omega_j) \propto \omega_j^{\gamma_j}(1-\omega_j)^{1-\gamma_j}\omega_j^{a_\omega-1}(1-\omega_j)^{b_\omega-1} \propto \omega_j^{\gamma_j+a_\omega-1}(1-\omega_j)^{1-\gamma_j+b_\omega-1}$. Therefore, the posterior distribution of ω_j is the Beta in equation (30).

Sampling σ^2. Finally, $p(\sigma^2|data,\boldsymbol{\omega},\boldsymbol{\lambda},\boldsymbol{\gamma}) \propto (\sigma^2)^{-\frac{N+K+1}{2}-1}e^{-\frac{1}{2\sigma^2}[(\boldsymbol{\mu}_{\boldsymbol{R}}-\boldsymbol{C}\boldsymbol{\lambda})^\top(\boldsymbol{\mu}_{\boldsymbol{R}}-\boldsymbol{C}\boldsymbol{\lambda})+\boldsymbol{\lambda}^\top\boldsymbol{D}\boldsymbol{\lambda}]}$. Hence, the posterior distribution of σ^2 is the inverse-Gamma in equation (31). The proof of Proposition 6 follows the same identical steps, and is therefore omitted for brevity.

[44]If we had instead imposed $\omega_j = 0.5$, as in Section III.1.2, the Bayes factor would simply be $\frac{p(\lambda_j|\gamma_j=1,\sigma^2)}{p(\lambda_j|\gamma_j=0,\sigma^2)}$.

A.2 Data

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

Panel A: asset pricing factors

Factor ID	Reference	Factor ID	Reference
MKT	Sharpe (1964, JF), Lintner (1965, JF)	HML_DEVIL	Asness and Frazzini (2013, JPM)
SMB	Fama and French (1992, JF)	QMJ	Asness, Frazzini, and Pedersen (2019, RAS)
HML	Fama and French (1992, JF)	FIN_UNC	Jurado, Ludvigson, and Ng (2015, AER), Ludvigson, Ma, and Ng (2019, AEJ-M)
RMW	Fama and French (2015, JFE)	REAL_UNC	Jurado, Ludvigson, and Ng (2015, AER), Ludvigson, Ma, and Ng (2019, AEJ-M)
CMA	Fama and French (2015, JFE)	MACRO_UNC	Jurado, Ludvigson, and Ng (2015, AER), Ludvigson, Ma, and Ng (2019)
UMD	Carhart (1997, JF), Jegadeesh and Titman (1993, JF)	TERM	Chen, Ross and Roll (1986, JB), Fama and French (1993, JFE)
STRev	Jegadeesh and Titman (1993, JF)	DELTA_SLOPE	Ferson and Harvey (1991, JPE)
LTRev	Jegadeesh and Titman (2001, JF)	CREDIT	Chen, Ross and Roll (1986, JB), Fama and French (1993, JFE)
q_IA	Hou, Xue, Zhang (2015, RFS)	DIV	Campbell (1996, JPE)
q_ROE	Hou, Xue, Zhang (2015, Review of Financial Studies)	PE	Basu (1977, JF), Ball (1978, JFE)
LIQ_NT	Pastor and Stambaugh (2003, JPE)	BW_INV_SENT	Baker and Wurgler (2006, JF)
LIQ_TR	Pastor and Stambaugh (2003, JPE)	HJTZ_INV_SENT	Huang, Jiang, Tu, and Zhou (2015, RFS)
MGMT	Stambaugh and Yuan (2016, RFS)	BEH_PEAD	Daniel, Hirshleifer, and Sun (2019, RFS)
PERF	Stambaugh and Yuan (2016, RFS)	BEH_FIN	Daniel, Hirshleifer, and Sun (2019, RFS)
ACCR	Sloan (1996, AR)	MKT*	Daniel, Mota, Rottke, and Santos (2020, RFS)
DISSTR	Campbell, Hilscher, and Szilagyi (2008, JF)	SMB*	Daniel, Mota, Rottke, and Santos (2020, RFS)
A_Growth	Cooper, Gulen, and Schill (2008, JF)	HML*	Daniel, Mota, Rottke, and Santos (2020, RFS)
COMP_ISSUE	Daniel and Titman (2006, JF)	RMW*	Daniel, Mota, Rottke, and Santos (2020, RFS)
GR_PROF	Novy-Marx (2013, JFE)	CMA*	Daniel, Mota, Rottke, and Santos (2020, RFS)
INV_IN_ASSETS	Titman, Wei, and Xie (2004, JFQA)	SKEW	Langlois (2019, JFE)
NetOA	Hirshleifer, Kewei, Teoh, and Zhang (2004, JAE)	NONDUR	Chen, Ross and Roll (1986, JB), Breeden, Gibbons, and Litzenberger (1989, JF)
OSCORE	Ohlson (1980, JAR)	SERV	Breeden, Gibbons, and Litzenberger (1989, JF), Hall (1978, JPE)
ROA	Chen, Novy-Marx, and Zhang (2010, working paper)	UNRATE	Gertler and Grinols (1982, JMCB)
STOCK_ISS	Ritter (1991, JF), Fama and French (2008, JF)	IND_PROD	Chan, Chen, and Hsieh (1985, JFE), Chen, Ross and Roll (1986, JB)
INTERM_CR	He, Kelly, and Manela (2017, JFE)	OIL	Chen, Ross and Roll (1986, JB)
BAB	Frazzini and Pedersen (2014, JFE)		

Panel B: additional anomalies used for the construction of test assets

Anomaly ID	Reference	Anomaly ID	Reference
CashAssets	Palazzo (2012, JFE)	Volume	Garfinkel (2009, RAS)
FCFBook	Hou, Karolyi, and Kho (2011, RFS)	SGASales	Freyberger, Neuhierl, and Weber (2020, RFS)
CFPrice	Desai, Rajgopal, and Venkatachalam (2004, AR)	Q	Kaldor (1996, REStud)
CapTurnover	Haugen and Baker (1996, JFE)	IVolCAPM	Ang, Hodrick, Xing, and Zhang (2006, JF)
CapIntens	Gorodnichenko and Weber (2016, AER)	IVolFF3	Ang, Hodrick, Xing, and Zhang (2006, JF)
DP_tr	Litzenberger and Ramaswamy (1979, JFE)	DayVariance	Ang, Hodrick, Xing, and Zhang (2006, JF)
PPE_delta	Lyandres, Sun, and Zhang (2008, RFS)	ProfMargin	Soliman (2008, AR)
Lev	Lewellen (2015, CFR)	PriceCostMargin	Bustamante and Donangelo (2017, RFS)
SalesPrice	Lewellen (2015, CFR)	OperLev	Novy-Marx (2011, RF)
IntermMom	Novy-Marx (2012, JFE)	FixedCostSale	D'Acunto, Liu, Pflueger, and Weber (2018, JFE)
YearHigh	George and Hwang (2004, JF)	LTMom	Bondt and Thaler (1985, JF)
PE_tr	Basu (1983, JFE)	NetSalesNetOA	Soliman (2008, AR)
BidAsk	Chung and Zhang (2014, JFM)	AssetsMarket	Bhandari (1988, JF)

The table presents the list of factors and anomalies used in Section V.1. 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, anomalies, sources, and references can be found in Tables IA13 and IA14 of the Internet Appendix. The journal acronyms used in the table are: AEJ-M = AEJ: Macroeconomics; AER = American Economy Review; AR = Accounting Review; CFR = Critical Finance Review; JAE = Journal of Accounting and Economics; JAR = Journal of Accounting Research; JB = Journal of Business; JF= Journal of Finance; JFE = Journal of Financial Economics; JFM = Journal of Financial Markets; JFQA = Journal of Financial and Quantitative Analysis; JMCB = Journal of Money, Credit, and Banking; JMP = Journal of Portfolio Management; JPE = Journal of Political Economy; RAS = Review of Accounting Studies; REStud = Review of Economic Studies; RF = Review of Finance; RFS = Review of Financial Studies.