THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

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

Svetlana Bryzgalova
Jiantao Huang
Christian Julliard

SRC Discussion Paper No 93
January 2020

Systemic Risk Centre

Discussion Paper Series

ISSN 2054-538X

Abstract
We propose a novel, and simple, Bayesian estimation and model selection procedure for crosssectional asset pricing. Our approach, that allows for both tradable and non-tradable factors, and is applicable to high dimensional cases, has several desirable properties. First, weak and spurious factors lead to diffuse, and centered at zero, posteriors for their market price of risk, making such factors easily detectable. Second, posterior inference is robust to the presence of such factors. Third, we show that flat priors for risk premia lead to improper marginal likelihoods, rendering model selection invalid. Therefore, we provide a novel prior, that is diffuse for strong factors but shrinks away useless ones, under which posterior probabilities are well behaved, and can be used for factor and (non necessarily nested) model selection, as well as model averaging, in large scale problems. We apply our method to a very large set of factors proposed in the literature, and analyse 2.25 quadrillion possible models, gaining novel insights on the empirical drivers of asset returns.

JEL codes: G12, C11, C12, C52, C58.

Keywords: Cross-sectional asset pricing, factor models, model evaluation, multiple testing, data mining, p-hacking, Bayesian methods.

This paper is published as part of the Systemic Risk Centre's Discussion Paper Series. The support of the Economic and Social Research Council (ESRC) in funding the SRC is gratefully acknowledged [grant number ES/R009724/1].

Svetlana Bryzgalova, London Business School and Systemic Risk Centre, London School of Economics and Political Science

Jiantao Huang, Department of Finance, London School of Economics and Political Science

Christian Julliard, Department of Finance, Financial Markets Group and Systemic Risk Centre, London School of Economics and Political Science, and Centre for Economic Policy Research

Published by
Systemic Risk Centre
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

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

Svetlana Bryzgalova† Jiantao Huang‡ Christian Julliard§

January 10, 2020

Abstract

We propose a novel, and simple, Bayesian estimation and model selection procedure for cross-sectional asset pricing. Our approach, that allows for both tradable and non-tradable factors, and is applicable to high dimensional cases, has several desirable properties. First, weak and spurious factors lead to diffuse, and centered at zero, posteriors for their market price of risk, making such factors easily detectable. Second, posterior inference is robust to the presence of such factors. Third, we show that flat priors for risk premia lead to improper marginal likelihoods, rendering model selection invalid. Therefore, we provide a novel prior, that is diffuse for strong factors but shrinks away useless ones, under which posterior probabilities are well behaved, and can be used for factor and (non necessarily nested) model selection, as well as model averaging, in large scale problems. We apply our method to a very large set of factors proposed in the literature, and analyse 2.25 quadrillion possible models, gaining novel insights on the empirical drivers of asset returns.

Keywords: Cross-sectional asset pricing, factor models, model evaluation, multiple testing, data mining, p-hacking, Bayesian methods.
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 Mikhail Chernov, Pierre Collin-Dufresne, Aureo de Paula, Marcelo Fernandes, Rodrigo Guimaraes, Alexander Michaelides, Olivier Scaillet, and George Tauchen.

†London Business School, sbryzgalova@london.edu

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

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

I Introduction

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

As stressed by Harvey (2017) in his AFA presidential address, the first observation naturally calls for a Bayesian solution – and we develop one. Furthermore, we show that the two fundamental problems above are tightly connected, and a naive Bayesian approach to model selection may fail in the presence of spurious factors. Hence, we correct it, and apply our method to the zoo of traded and non-traded factors proposed in the literature, jointly evaluating 2.25 quadrillion models, and gaining novel insights on the empirical drivers of asset returns. In particular, we find that only a handful of factors proposed in the previous literature are robust explanators of the cross-section of asset returns, and a four *robust* factor model easily outperforms canonical factor models. Nevertheless, we also show that the 'true' latent stochastic discount factor (SDF) is dense is the space of empirical asset pricing factors i.e. a large set of factors is needed to fully capture its pricing implications. Nonetheless, the SDF-implied maximum Sharpe ratio in the economy is not unrealistically high.

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

Second, the main intent of this paper is to provide a method for handling inference on the *entirety* of the factor zoo at once. This naturally calls for the use of model posterior probabilities. But, we show that, under flat priors for risk premia, model and factors selection based on marginal likelihoods (i.e. on posterior model probabilities or Bayes factors) is unreliable: asymptotically, useless factors get selected with probability one. This is due to the fact that lack of identification generates an unbounded manifold for the risk premia parameters, over which the likelihood surface is totally flat.[1] Hence, integration applied directly to the likelihood, as if it were an unnormalized probability distribution function (pdf), produces improper marginal "posteriors." As a result, in the presence of identification failure, naive Bayesian inference has the same weakness as the frequentist one. This observation, however, not only illustrates the nature of the problem, but also suggests how to restore inference: use suitable, non-informative but yet non-flat, priors.

Third, building upon the literature on predictor selection (see e.g. Ishwaran, Rao, et al. (2005) and Giannone, Lenza, and Primiceri (2018)), we provide a novel (continuous) "spike-and-slab" prior that restores the validity of model selection based on posterior model probabilities and Bayes factors. The prior is uninformative (the "slab") for strong factors, but shrinks away (the "spike") useless factors. This approach is similar in spirit to a ridge regression, and acts as a (Tikhonov-Phillips) regularization of the likelihood function of the cross-sectional regression needed to estimate risk premia. A distinguishing feature of our prior is that the prior variance of a factor's risk premium is proportional to its correlation with the test asset returns. Hence, when a useless factor is present, the prior variance of its risk premium converges to zero, so the shrinkage dominates and forces its posterior distribution to concentrated around zero. Not only this prior restores integrability, but also: *i*) makes it computationally feasible to analyse quadrillions of alternative factor models; *ii*) allows the researcher to encode prior beliefs about the sparsity of the true SDF without imposing hard thresholds; and *iii*) shrinks the estimate of useless factors' risk premia toward zero. We regard this novel spike-and-slab prior approach as a solution for the high-dimensional inference problem generated by the factor zoo.

Our method is easy to implement and, in all of our simulations, has good finite sample properties, even when the cross-section of test assets is large. We investigate its performance for risk premia estimation, model evaluation and factor selection, in a range of simulation designs that mimic the stylized features of returns. Our simulations account for potential model misspecification and the presence of either strong or useless factors in the model. The use of posterior sampling naturally allows to build credible confidence intervals not only for risk premia, but also other statistics of interest, such as the cross-sectional R^2, that is notoriously hard to estimate precisely (Lewellen, Nagel, and Shanken (2010)).

We show that whenever risk premia are well identified, both our method and the frequentist approach provide valid confidence intervals for model parameters, with empirical coverage being close to its nominal size. The posterior distribution for useless factors, however, is reliably centered around zero and quickly revels them even in a relatively short sample. We find that the posterior of strong factors is largely unaffected by the identification failure, with the posterior coverage corresponding to its nominal size as well. In other words, the Bayesian approach restores reliable statistical inference in the model.

We also demonstrate the pitfalls of flat priors for risk premia with the same simulation design: their use leads to selecting useless factor with probability approaching 1 for all the sample sizes. However, our spike-and-slab prior seems to successfully eliminate them from the model, while

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

retaining the true sources of risk.

Our results have important empirical implications for the estimation of popular linear factor models, and their comparison. We jointly evaluate 51 factors proposed in the previous literature, yielding a total of 2.25 quadrillion possible models to analyze, and find that only a handful of factors are robust explanators of the cross-section of asset returns (the Fama and French (1992) "high-minus-low" proxy for the value premium, as well the adjusted versions of both market and "small-minus-big" size factors of Daniel, Mota, Rottke, and Santos (2018)).

Jointly, the four robust factors provide a model that is, compared to the previous empirical literature, one order of magnitude more likely to have generated the observed asset returns (it's posterior probability is about 90%). However, we show that with very high probability the "true" latent SDF is dense in the space of factors i.e. capturing its characteristics requires the use of 24-25 factors. Nevertheless, the SDF-implied maximum Sharpe ratio is not excessive, suggesting a high degree of commonality, in terms of captured risks, among the factors in the zoo.

Furthermore, we apply our useless factors detection method to a selection of popular linear SDFs. We find that many non-traded factors, such as consumption proxies, labour factors, or the consumption-to-wealth ratio, *cay*, are only weakly identified at best, and are characterised by a substantial degree of model misspecification and uncertainty.

I.1 Closely related literature

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

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

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

Shanken (1987) and Harvey and Zhou (1990) were probably the first contributions to the literature that adapted the Bayesian framework to the analysis of portfolio choice, and developed GRS-type tests (cf. Gibbons, Ross, and Shanken (1989)) for mean-variance efficiency. While

Shanken (1987) was the first to examine the posterior odds ratio for portfolio alphas in the linear factor model, Harvey and Zhou (1990) set the benchmark by imposing the priors, both diffuse and informative, directly on the deep model parameters. Using the data on 12 industry portfolios, they reject the mean-variance efficiency of the market return with a high degree of certainty.

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

Barillas and Shanken (2018) extend the aforementioned prior to derive a closed-form solution for the Bayes' factor, and use it to compare different linear factor models exploiting the time series dimension of the data.[2] In a recent paper, Goyal, He, and Huh (2018) also extend the notion of distance between alternative model specifications, and highlight the tension between the power of the GRS-type tests, and the absolute return-based measures of mispricing.

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

Naturally, our paper also contributes to the very active (and growing) body of work that critically evaluates existing findings in the empirical asset pricing literature and tries to develop a robust methodology. There is ample empirical evidence that most linear asset pricing models are misspecified (e.g. Chernov, Lochstoer, and Lundeby (2019), He, Huang, and Zhou (2018), Giglio and Xiu (2018)). Gospodinov, Kan, and Robotti (2014) develop a general approach for misspecification-robust inference that provides valid confidence interval for the pseudo-true values of the risk premia. Giglio and Xiu (2018) exploit the invariance principle of the PCA, and recover the risk premia, asso-

[2]Chib, Zeng, and Zhao (forthcoming) show that the *improper* prior specification of Barillas and Shanken (2018) is problematic and propose a new class of priors that leads to valid model comparison.

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

ciated with the given factor, from the projection on the span of latent factors driving a cross-section of asset returns. Uppal, Zaffaroni, and Zviadadze (2018) adopt a similar approach by recovering latent factors from the residuals of the asset pricing model, effectively completing the span of the SDF. Daniel, Mota, Rottke, and Santos (2018), instead, focus on the construction of the cross-sectional factors, and note that many well-established tradable portfolios, such as HML and SMB, can be substantially improved in asset pricing tests by hedging their unpriced component (that does not carry a risk premium). We do not take a stand on the origin of the factors or the completion of the model space. Instead, we consider the whole universe of potential models that can be created from the set of observable candidates factors proposed in the empirical literature. As such, our analysis explicitly takes into account both standard specifications that have been successfully used in numerous papers (e.g. Fama-French 3-factor models, or nondurable consumption growth), as well as the combinations of the factors that have never been explicitly tested.

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

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

II Inference in Linear Factor Models

This section introduces the notation and reviews the main results of the Fama-MacBeth (FM) regression method (see Fama and MacBeth (1973)). We focus on classic linear factor models for cross-sectional asset returns. Suppose that there are K factors, $\boldsymbol{f}_t = (f_{1t} \dots f_{Kt})^\top$, $t = 1, \dots T$, which could be either tradable or non-tradable. To simplify exposition, we consider mean zero factors that have also been demeaned in sample, so that we have both $\mathbb{E}[\boldsymbol{f}_t] = \boldsymbol{0_K}$ and $\bar{\boldsymbol{f}} = \boldsymbol{0_K}$ where $\mathbb{E}[.]$ denotes the unconditional expectation and the upper bar denotes the sample mean operator. The returns of N test assets, in excess of the risk free rate, are denoted by $\boldsymbol{R}_t = (R_{1t} \dots R_{Nt})^\top$.

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

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

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

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

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

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{f}}$ denotes the time series estimates, λ_c is a scalar average mispricing that should be equal to zero under the null of the model being correctly specified, $\boldsymbol{1_N}$ 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. If the model is correctly specified, it implies the parameter restriction: $\boldsymbol{a} = \mathbb{E}[\boldsymbol{R}_t] = \lambda_c \boldsymbol{1_N} + \boldsymbol{\beta_f} \boldsymbol{\lambda_f}$. Therefore, we can rewrite the two-step Fama-MacBeth regression into one equation as

$$\boldsymbol{R}_t = \lambda_c \boldsymbol{1_N} + \boldsymbol{\beta_f} \boldsymbol{\lambda_f} + \boldsymbol{\beta_f} \boldsymbol{f}_t + \boldsymbol{\epsilon}_t. \tag{3}$$

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

Let $\boldsymbol{B}^\top = (\boldsymbol{a}, \boldsymbol{\beta_f})$ and $\boldsymbol{F}_t^\top = (1, \boldsymbol{f}_t^\top)$, and consider the matrices of stacked time series observations

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_1^\top \\ \vdots \\ \boldsymbol{R}_T^\top \end{pmatrix}, \quad \boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_1^\top \\ \vdots \\ \boldsymbol{F}_T^\top \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1^\top \\ \vdots \\ \boldsymbol{\epsilon}_T^\top \end{pmatrix}.$$

The regression in (1) can then be rewritten as $\boldsymbol{R} = \boldsymbol{F}\boldsymbol{B} + \boldsymbol{\epsilon}$, yielding the time-series estimates of

$(\boldsymbol{a}, \boldsymbol{\beta_f})$ and $\boldsymbol{\Sigma}$:

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

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

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

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

$$\widehat{\boldsymbol{\sigma}}^2(\widehat{\boldsymbol{\lambda}}) = \frac{1}{T} \left[(\widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\beta}})^{-1} \widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\beta}} (\widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\beta}})^{-1} (1 + \widehat{\boldsymbol{\lambda}}_{\boldsymbol{f}}^\top \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^{-1} \widehat{\boldsymbol{\lambda}}_{\boldsymbol{f}}) \right], \tag{5}$$

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

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

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

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

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

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

[4] An alternative way (see e.g. Cochrane (2005), Page 242) to account for the uncertainty from "generated regressors," such as $\boldsymbol{\beta_f}$, is to estimate the whole system in GMM. The moments are

$$\boldsymbol{g}_T(\boldsymbol{a}, \boldsymbol{\beta_f}, \boldsymbol{\lambda}) = \begin{pmatrix} \boldsymbol{I_N} \otimes \boldsymbol{I_{K+1}} & \\ & \mathbb{A}^\top \end{pmatrix} \begin{pmatrix} \mathbb{E}[\boldsymbol{R_t} - \boldsymbol{a} - \boldsymbol{\beta_f f_t}] \\ \mathbb{E}[(\boldsymbol{R_t} - \boldsymbol{a} - \boldsymbol{\beta_f f_t}) \otimes \boldsymbol{f_t}^\top] \\ \mathbb{E}[\boldsymbol{R_t} - \lambda_c \boldsymbol{1_N} - \boldsymbol{\beta_f \lambda_f}] \end{pmatrix} = \boldsymbol{0}.$$

where $\boldsymbol{\beta} = (\boldsymbol{1_N} \ \boldsymbol{\beta_f})$, $\mathbb{A}^\top = \boldsymbol{\beta}^\top$ for OLS and $\mathbb{A}^\top = \boldsymbol{\beta}^\top \boldsymbol{\Sigma^{-1}}$ for GLS. Also note that GLS estimation is not the same as efficient GMM estimation.

[5] For example, Kleibergen and Zhan (2015) derive the asymptotic distribution of the R^2 under the assumption that a few unknown factors are able to explain expected asset returns, and show that, in the presence of a useless

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

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

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

II.1 Bayesian Fama-MacBeth

This section introduces our hierarchical Bayesian Fama-MacBeth (BFM) estimation method. A formal derivation is presented in Appendix A.1. To start with, let's consider the time-series regression. We assume that the time series error terms follow an iid multivariate Gaussian distribution (the approach, at the cost of analytical solutions, could be generalized to accommodate different distributional assumptions), i.e. $\boldsymbol{\epsilon} \sim \mathcal{MVN}(\boldsymbol{0_{T\times N}}, \boldsymbol{\Sigma} \otimes \boldsymbol{I_T})$. The likelihood of the data $(\boldsymbol{R}, \boldsymbol{F})$ is then

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

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

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

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

where $\widehat{\boldsymbol{B}}_{ols}$ and $\hat{\boldsymbol{\Sigma}}$ denote the canonical OLS based estimates, and $\mathcal{W}^{-1}$ is the inverse-Wishart distribution (a multivariate generalization of the inverse-gamma distribution). From the above,

factor, the OLS R^2 is more likely to be inflated than its GLS counterpart.

we can sample the posterior distribution of the parameters $(\boldsymbol{B}, \boldsymbol{\Sigma})$ by first drawing the covariance matrix $\boldsymbol{\Sigma}$ from the inverse-Wishart distribution conditional on the data, and then drawing $\boldsymbol{B}$ from a multivariate normal distribution conditional on the data and the draw of $\boldsymbol{\Sigma}$.

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

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

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

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

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

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

However, realistically, the models are rarely true. Therefore, one might want to allow for the presence of pricing errors, $\boldsymbol{\alpha}$, in the cross-sectional regression.[6] This can be easily accommodated

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

within our Bayesian framework since in this case the data generating process in the second stage becomes $\boldsymbol{a} = \boldsymbol{\beta}\boldsymbol{\lambda} + \boldsymbol{\alpha}$. If we further assume that pricing error α_i follows an independent and identical normal distribution $\mathcal{N}(0, \sigma^2)$, the likelihood function in the second step becomes[7]

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

In the cross-sectional regression the "data" are the expected risk premia, $\boldsymbol{a}$, and the factor loadings, $\boldsymbol{\beta}$, albeit these quantities are not directly observable to the researcher. Hence, in the above, we are conditioning on the knowledge of these quantities, which can be sampled from the first step Normal-inverse-Wishart posterior distribution (8)-(9). Conceptually, this is not very different from the Bayesian modeling of latent variables. In the benchmark case, we assume a Jeffreys' diffuse prior[8] for $(\boldsymbol{\lambda}, \sigma^2)$: $\pi(\boldsymbol{\lambda}, \sigma^2) \propto \sigma^{-2}$. In Appendix A.1, we show that the posterior distribution of $(\boldsymbol{\lambda}, \sigma^2)$ is then

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

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

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

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

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

Definition 2 (Bayesian Fama-MacBeth GLS (BFM-GLS)) *The posterior distribution of $\boldsymbol{\lambda}$ conditional on $\boldsymbol{B}$, $\boldsymbol{\Sigma}$ and the data, is a Dirac distribution at $(\boldsymbol{\beta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}$. A draw $(\boldsymbol{\lambda}_{(j)})$*

[7] We derive a formulation with non-spherical cross-sectional pricing errors, that leads to a GLS type estimator, in Appendix A.2.

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

from the posterior distribution of $\boldsymbol{\lambda}$ conditional on the data only is obtained by drawing $\boldsymbol{B}_{(j)}$ and $\boldsymbol{\Sigma}_{(j)}$ from the Normal-inverse-Wishart in equations (8)-(9) and computing $(\boldsymbol{\beta}_{(j)}^{\top}\boldsymbol{\Sigma}_{(j)}^{-1}\boldsymbol{\beta}_{(j)})^{-1}\boldsymbol{\beta}_{(j)}^{\top}\boldsymbol{\Sigma}_{(j)}^{-1}\boldsymbol{a}_{(j)}$.

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

$$R_{gls}^2 = 1 - \frac{(\boldsymbol{a} - \boldsymbol{\beta}\boldsymbol{\lambda}_{\boldsymbol{gls}})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{a} - \boldsymbol{\beta}\boldsymbol{\lambda}_{\boldsymbol{gls}})}{(\boldsymbol{a} - \bar{a}\mathbf{1}_{\boldsymbol{N}})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{a} - \bar{a}\mathbf{1}_{\boldsymbol{N}})}. \tag{14}$$

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

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

II.2 Model selection

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

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

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

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

In this subsection, we show first that flat priors for risk premia (as the Jeffreys' priors used in section II.1 for illustrative purposes), are not suitable for model selection in the presence of spurious factors. Given the close analogy between frequentist testing and Bayesian inference with flat priors, this is not too surprising. But the novel insight is that the problem arises exactly because of the use of flat priors, and can therefore be fixed by using non-flat, yet non-informative, priors. Second, we introduce "spike-and-slab" priors that are robust to the presence of spurious factors, and particularly powerful in high-dimensional model selection i.e. when one wants, as in our empirical application, to test all factors in the zoo.

II.2.1 Pitfalls of Flat Priors for Risk Premia

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

In order to perform model selection, we relax the (null) hypothesis that models are correctly specified and allow instead for the presence of cross-sectional pricing errors. That is, we consider the cross-sectional regression $\boldsymbol{a} = \boldsymbol{\beta}\boldsymbol{\lambda} + \boldsymbol{\alpha}$. For illustrative purposes, we focus on spherical errors, but all the results in this and the following subsections can be generalized to the non-spherical error setting in Appendix A.2.

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

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

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

Recall that when considering multiple models, say wlog 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.3 we show that, when using a Jeffreys' prior (that is flat for $\boldsymbol{\lambda}$), the marginal likelihood is

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

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

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

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

Note also that flat, hence improper, priors for the risk premia are not legitimate since they render the posterior model probabilities arbitrary. Suppose that we are testing the null $H_0 : \lambda_j = 0$. Under the null hypothesis, the prior for (λ, σ^2) is $\lambda_j = 0$ and $\pi(\lambda_{-j}, \sigma^2) \propto \frac{1}{\sigma^2}$. However, the prior under the alternative hypothesis is $\pi(\lambda_j, \lambda_{-j}, \sigma^2) \propto \frac{1}{\sigma^2}$. Since the marginal likelihoods of data, $p(data|H_0)$ and $p(data|H_1)$, are both undetermined, we cannot define the Bayes' factor $\frac{p(data|H_1)}{p(data|H_0)}$ (see e.g. Cremers (2002), Chib, Zeng, and Zhao (forthcoming)). 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.[10] Similarly, we can still assign improper priors for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in the first time series step.

The final reason why it might be undesirable to use Jeffreys' prior in the second step, is that it does not impose any shrinkage on the parameters. This is problematic given the large number of

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

members of the factor zoo, while we have only limited time-series observations of both factors and test asset returns.

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

II.2.2 Spike and Slab Prior for Risk Premia

In order to make sure that the integration of the marginal likelihood is well-behaved, we propose a novel prior specification for the factors' risk premia $\boldsymbol{\lambda}_{\boldsymbol{f}}^{\top} = (\lambda_1, ..., \lambda_K)$.[11] Since the inference in time-series regression is always valid, we only modify the priors of the cross-sectional regression parameters.

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

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

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 $\boldsymbol{\gamma}$. We can then express the prior for the risk factors, $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$, of model $\boldsymbol{\gamma}$ as

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

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

Given the above prior specification, we sample the posterior distribution by sequentially drawing from the conditional distributions of the parameters (i.e. we use a Gibbs sampling algorithm). The crucial steps, in addition to the sampling of the times series parameters from the posteriors in equations (8)-(9), are as follows.

Sampling $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$

Note that:

[11] We do not shrink the intercept λ_c.

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

where $SSR_\gamma = \boldsymbol{a}^\top\boldsymbol{a} - \boldsymbol{a}^\top\boldsymbol{\beta_\gamma}(\boldsymbol{\beta_\gamma^\top\beta_\gamma} + \boldsymbol{D_\gamma})^{-1}\boldsymbol{\beta_\gamma^\top a} = \min_{\boldsymbol{\lambda_\gamma}}\{(\boldsymbol{a} - \boldsymbol{\beta_\gamma\lambda_\gamma})^\top(\boldsymbol{a} - \boldsymbol{\beta_\gamma\lambda_\gamma}) + \boldsymbol{\lambda_\gamma^\top D_\gamma\lambda_\gamma}\}$. Note that SSR_γ is the minimized sum of squared errors with generalised ridge regression penalty term $\boldsymbol{\lambda_\gamma^\top D_\gamma\lambda_\gamma}$. That is, our prior modelling is analogous to introducing a Tikhonov-Phillips regularisation (see Tikhonov, Goncharsky, Stepanov, and Yagola (1995), 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 (that, in our settings, arises in the presence of useless factors). However, in our setting the shrinkage applied to the factors is heterogeneous, since we rely on the partial correlation between factors and test assets to set ψ_j as:

$$
\psi_j = \psi \times \boldsymbol{\rho_j^\top \rho_j}, \tag{16}
$$

where $\boldsymbol{\rho_j}$ is an $N \times 1$ vector of correlation coefficients between factor j and the test assets, and $\psi \in \mathbb{R}_+$ is a tuning parameter which controls the shrinkage over all the factors.[12] When the correlation between f_{jt} and $\boldsymbol{R_t}$ is very low, as in the case of a useless factor, the penalty for λ_j, which is the reciprocal of $\psi\boldsymbol{\rho_j^\top\rho_j}$, is very large and dominates the sum of squared errors.

Let $\boldsymbol{\hat{\lambda}_\gamma} = (\boldsymbol{\beta_\gamma^\top\beta_\gamma} + \boldsymbol{D_\gamma})^{-1}\boldsymbol{\beta_\gamma^\top a}$ and $\hat{\sigma}^2(\boldsymbol{\hat{\lambda}_\gamma}) = \sigma^2(\boldsymbol{\beta_\gamma^\top\beta_\gamma} + \boldsymbol{D_\gamma})^{-1}$, the posterior distribution of $\boldsymbol{\lambda_\gamma}$ is

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

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

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

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

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

Sampling σ^2

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

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

hence the posterior distribution of σ^2 is an inverse-Gamma: $\sigma^2|data, \boldsymbol{\gamma} \sim \Gamma^{-1}\left(\frac{N}{2}, \frac{SSR_\gamma}{2}\right)$.

Finally, we obtain the marginal likelihood of the data in model $\boldsymbol{\gamma}$ by integrating out σ^2:

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

When comparing two models, using posterior model probabilities is equivalent to simply using the ratio of the marginal likelihoods, i.e. the Bayes' factor 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 $\boldsymbol{\gamma'}$.

Remark 2 (Bayes Factor) *Consider two nested linear factor models, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$. The only difference between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ is γ_p: γ_p equals 1 in model $\boldsymbol{\gamma}$ but 0 in model $\boldsymbol{\gamma}'$. Let $\boldsymbol{\gamma}_{-p}$ denote a $(K-1) \times 1$ vector of model index excluding γ_p: $\boldsymbol{\gamma}^\top = (\boldsymbol{\gamma}_{-p}^\top, 1)$ and $\boldsymbol{\gamma}'^\top = (\boldsymbol{\gamma}_{-p}^\top, 0)$ where, without loss of generality, we have assumed that the factor p is ordered last. The Bayes' factor is then*

$$BF_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} = \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \left(1 + \psi_p \boldsymbol{\beta}_{\boldsymbol{p}}^\top \left[\boldsymbol{I_N} - \boldsymbol{\beta}_{\boldsymbol{\gamma}'}(\boldsymbol{\beta}_{\boldsymbol{\gamma}'}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'})^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma}'}^\top\right] \boldsymbol{\beta}_{\boldsymbol{p}}\right)^{-\frac{1}{2}}. \tag{17}$$

The above result is proved in Appendix A.4.

Since $\boldsymbol{\beta}_{\boldsymbol{p}}^\top[\boldsymbol{I_N} - \boldsymbol{\beta}_{\boldsymbol{\gamma}'}(\boldsymbol{\beta}_{\boldsymbol{\gamma}'}^\top \boldsymbol{\beta}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'})^{-1}\boldsymbol{\beta}_{\boldsymbol{\gamma}'}^\top]\boldsymbol{\beta}_{\boldsymbol{p}}$ 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 (17) dominates and the Bayes' factor will be much greater than 1, hence providing evidence in favour of model $\boldsymbol{\gamma}$.

Remember that $SSR_\gamma = \min_{\boldsymbol{\lambda_\gamma}}\{(\boldsymbol{a} - \boldsymbol{\beta_\gamma\lambda_\gamma})^\top(\boldsymbol{a} - \boldsymbol{\beta_\gamma\lambda_\gamma}) + \boldsymbol{\lambda}_{\boldsymbol{\gamma}}^\top \boldsymbol{D_\gamma \lambda_\gamma}\}$, 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 minimise SSR_γ, and $SSR_{\gamma'}/SSR_\gamma$ becomes much larger than 1; ii) large ψ_p decreases the second term in equation (17), lowering the Bayes' factor, and acting as a penalty for dimensionality.

A particularly interesting case is when the factor is useless: $\boldsymbol{\beta_p}$ converges to zero, but the penalty term $1/\psi_p \propto 1/\boldsymbol{\rho}_{\boldsymbol{p}}^\top \boldsymbol{\rho_p}$ goes to infinity. On the one hand, the first term in equation (17) will converge to 1; on the other hand, since $\psi_p \approx 0$ in large sample, the second term in equation (17) will also be around 1. Therefore, the Bayes' factor for a useless factor will go to 1 asymptotically.[13]

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

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.

Remark 3 (Level Factors) *Identification failure of factors' risk premia can arise in the presence of 'level factors,' exposure to which is non-zero, but lacks cross-sectional spread i.e.* $\boldsymbol{\beta}_j \to c\mathbf{1}_{\boldsymbol{N}}$ *with* $c \neq 0$. *These factors help explain the average level of returns, but not the their cross-sectional dispersion, and hence are collinear with the common cross-sectional intercept. Our approach can handle this case by using variance standardised variables in the estimation and replacing the penalty in (16) with* $\psi_j = \psi \times \widetilde{\boldsymbol{\rho}_j}^\top \widetilde{\boldsymbol{\rho}_j}$, *where* $\widetilde{\boldsymbol{\rho}_j}$ *is the cross-sectionally demeaned vector of correlations with asset returns, i.e.* $\widetilde{\boldsymbol{\rho}_j} = \boldsymbol{\rho}_j - \left(\frac{1}{N}\sum_{i=1}^{N}\rho_{j,i}\right) \times \mathbf{1}_{\boldsymbol{N}}$

II.2.3 Continuous Spike

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

$$\lambda_j|\gamma_j, \sigma^2 \sim \mathcal{N}(0, r(\gamma_j)\psi_j\sigma^2)$$

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

Redefine $\boldsymbol{D}$ as a diagonal matrix with elements c, $(r(\gamma_1)\psi_1)^{-1}, \ldots, (r(\gamma_K)\psi_K)^{-1}$ where ψ_j is given as before by equation (16). In matrix notation the prior for $\boldsymbol{\lambda}$ is: $\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2\boldsymbol{D}^{-1})$. The term $r(\gamma_j)\psi_j$ in $\boldsymbol{D}^{-1}$ is set to be small or large depending on whether $\gamma_j = 0$ or $\gamma_j = 1$. In the empirical implementation, we force r to be much less than 1 since we intend to shrink λ_j towards zero when γ_j is 0.[14] Hence the spike component concentrates the mass of $\boldsymbol{\lambda}$ towards zero, whereas the slab component allows $\boldsymbol{\lambda}$ to take values over a much wider range. Therefore, the posterior distribution of $\boldsymbol{\lambda}$ is very similar to the case of a Dirac spike in section II.2.2.

We use Gibbs sampling (i.e. sequential sampling from conditional distributions) to draw the posterior distribution of the parameters $(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \sigma^2)$, where, as explained below, $\boldsymbol{\omega}$ encodes our prior beliefs about the sparsity of the true model.

Sampling $\boldsymbol{\lambda}_\gamma$

Combining the likelihood and the prior for $\boldsymbol{\lambda}$ we have:

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

Therefore, defining $\hat{\boldsymbol{\lambda}} = (\boldsymbol{\beta}^\top\boldsymbol{\beta} + \boldsymbol{D})^{-1}\boldsymbol{\beta}^\top\boldsymbol{a}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}) = \sigma^2(\boldsymbol{\beta}^\top\boldsymbol{\beta} + \boldsymbol{D})^{-1}$, the posterior distribution

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

of $\boldsymbol{\lambda}$ can be expressed as: $\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega} \sim \mathcal{N}(\hat{\boldsymbol{\lambda}}, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}}))$.

Sampling $\{\gamma_j\}_{j=1}^K$

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

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

Different hyper-parameters a_ω and b_ω determine whether we favor more parsimonious models or not.[15]

Given a ω_j, the Bayes factor for the j-th risk then factor is

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

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

Sampling $\boldsymbol{\omega}$

From Bayes' theorem we have:

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

Therefore, the posterior distribution of ω_j is: $\omega_j|data, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^2 \sim Beta(\gamma_j + a_\omega, 1 - \gamma_j + b_\omega)$.

Sampling σ^2

Finally,

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

Hence the posterior distribution of σ^2 is

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

Note that, when ω_j is a constant 0.5 and r converges to 0, the continuous slab-and-spike prior is equivalent to the one with a Dirac spike in section II.2.2. However, the MCMC algorithm of the continuous spike setting is particularly useful in the high dimensional case. Imagine that there are 30 candidate factors in the factor zoo. In the Dirac spike-and-slab prior case we have to

[15] We set $a_\omega = b_\omega = 2$ in the benchmark case. However, we could assign $a_\omega = 1$ and $b_\omega = 2$ in order to select a sparser model.

calculate the posterior model probabilities for 2^{30} different models. Given that we update $(\boldsymbol{a}, \boldsymbol{\beta})$ in each sampling round, posterior probabilities for all models are necessarily re-computed for every new draw of these quantities, rendering the computational cost very large. In contrast, using our continuous spike approach, we can simply use the posterior mean of γ_j to approximate the posterior marginal probability of the j-th factor.

III Simulation

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

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

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

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

To better illustrate the properties of the frequentist and bayesian approaches, we consider 3 estimation setups:

(a) the model includes only a strong factor (HML);

(b) the model includes only a useless factor;

(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, 1000, and 20000. We compare the performance of the OLS/GLS standard frequentist and Bayesian Fama-MacBeth estimators (FM and BFM, correspondingly) with the focus on risk premia recovery, testing, and identification of strong and useless factors for model comparison.

III.1 Estimating risk premia via Bayesian Fama-MacBeth

Since it is unlikely that in most empirical settings a linear factor model is correctly specified, we focus our discussion on the case that allows for model misspecification. We report similar simulation results for the case of correct model specification in Appendix C.

Table 1: Tests of risk premia in a misspecified model with a strong factor

		λ_c			λ_{strong}			R^2_{adj}	
	T	10%	5%	1%	10%	5%	1%	5th	95th
				Panel A: OLS					
FM	100	0.107	0.052	0.007	0.115	0.076	0.013	-4.22%	45.30%
	200	0.094	0.067	0.006	0.118	0.072	0.015	-3.03%	56.21%
	600	0.097	0.053	0.006	0.098	0.047	0.011	3.89%	55.75%
	1,000	0.088	0.048	0.012	0.103	0.060	0.012	8.65%	53.28%
	20,000	0.109	0.056	0.010	0.110	0.060	0.008	24.86%	36.55%
BFM	100	0.063	0.026	0.006	0.032	0.013	0.001	-3.56%	36.09%
	200	0.086	0.041	0.012	0.079	0.039	0.008	-3.67%	46.89%
	600	0.087	0.043	0.013	0.087	0.047	0.009	-0.67%	52.61%
	1,000	0.095	0.046	0.009	0.093	0.046	0.009	4.40%	53.46%
	20,000	0.096	0.052	0.012	0.100	0.052	0.012	23.90%	36.01%
				Panel B: GLS					
FM	100	0.246	0.173	0.067	0.251	0.154	0.070	17.20%	74.64%
	200	0.165	0.107	0.031	0.171	0.105	0.035	53.37%	80.45%
	600	0.137	0.076	0.020	0.137	0.073	0.016	69.42%	83.87%
	1,000	0.131	0.072	0.019	0.132	0.072	0.015	73.41%	84.16%
	20,000	0.122	0.074	0.014	0.113	0.061	0.015	80.34%	82.83%
BFM	100	0.139	0.083	0.022	0.143	0.083	0.024	31.27%	68.15%
	200	0.131	0.072	0.014	0.121	0.069	0.019	48.58%	72.99%
	600	0.114	0.061	0.013	0.126	0.067	0.018	65.94%	80.09%
	1,000	0.100	0.048	0.009	0.106	0.058	0.008	70.63%	81.49%
	20,000	0.092	0.050	0.012	0.100	0.048	0.012	80.22%	82.67%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_i^* in a misspecified model with an intercept and a strong factor. The true value of the cross-sectional R^2_{adj} is 30.55% (81.75%) for the OLS (GLS) estimation. Fama-MacBeth estimates are constructed using OLS (GLS) two-step cross-sectional regressions, with standard errors including Shanken correction. Confidence intervals and their size for BFM estimates are constructed using posterior coverage of Fama-MacBeth estimates of λ. The last two columns report the 5th and 95th percentiles of cross-sectional R^2_{adj} across 1000 simulations, evaluated at the simulation point estimates for FM, and its posterior mode for BFM.

Table 1 presents the size of the tests for risk premia and confidence intervals for cross-sectional R^2 in probably the most relevant (and best-case) scenario for empirical applications. It compares the performance of frequentist and Bayesian Fama-MacBeth estimators for the case when the model is misspecified, and includes a single cross-sectional factor that is priced and strongly identified, proxied by HML. Since the model is misspecified, cross-sectional R^2 never reaches 100% even for $T = 20,000$, with the population value of 31% (82%) for OLS (GLS). As expected, both FM and BFM estimators are very similar to each other, and provide confidence intervals of correct size. In case of the standard FM approach, they are constructed using standard t-statistics, adjusted for Shanken correction, and in case of the BFM, we rely on the quantiles of the posterior distribution to form the credible confidence intervals for parameters. The last two columns also report the quantiles of the mode of the posterior distribution of R^2 across the simulations.

Table 2 summarizes risk premia estimation for the same cross-section of 25 expected returns, but using a useless (spurious) factor as a candidate cross-sectional factor. As expected, the standard Fama-MacBeth estimator fails to recognize the rank failure in the second stage, and conventional risk premia estimates and t-statistics are inflated. Indeed, it is widely known since Kan and Zhang (1999), that if the model is misspecified, then t-statistics of the spurious factors tend to infinity asymptotically. This is confirmed in Panels A and B: for $T = 20,000$, the probability of finding a useless factor to have a t-stat above 1.96, is almost 50% for FM-OLS, and over 80% for FM-GLS. Furthermore, cross-sectional measures of fit, such as R^2 and related quantities, are substantially inflated: even though it's true value is 0 for both OLS and GLS settings, in-sample estimates produced by the frequentist approach, not only have a much wider empirical support (for example, from -4% to over 40% for FM-OLS), but also uncertainty that does not decrease with the sample size.

Table 2: Tests of risk premia in a misspecified model with a useless factor

		λ_c			$\lambda_{useless}$			R^2_{adj}	
	T	10%	5%	1%	10%	5%	1%	5th	95th
Panel A: OLS									
FM	100	0.072	0.037	0.010	0.055	0.011	0.001	-4.29%	41.84%
	200	0.078	0.043	0.005	0.084	0.021	0.001	-4.18%	45.24%
	600	0.089	0.043	0.014	0.223	0.116	0.013	-4.28%	43.70%
	1000	0.093	0.052	0.014	0.333	0.187	0.027	-4.29%	45.01%
	20000	0.213	0.135	0.067	0.698	0.488	0.172	-4.27%	43.63%
BFM	100	0.035	0.011	0.001	0.001	0.000	0.000	-2.47%	-0.36%
	200	0.041	0.013	0.001	0.008	0.002	0.000	-2.61%	-0.16%
	600	0.071	0.03	0.003	0.031	0.006	0.002	-2.87%	0.19%
	1000	0.047	0.02	0.001	0.039	0.017	0.002	-2.98%	0.84%
	20000	0.034	0.013	0.000	0.091	0.043	0.012	-3.36%	11.40%
Panel B: GLS									
FM	100	0.238	0.144	0.066	0.305	0.199	0.062	-3.47%	38.57%
	200	0.152	0.091	0.028	0.292	0.189	0.067	-3.75%	19.53%
	600	0.126	0.066	0.017	0.407	0.314	0.148	-3.85%	16.43%
	1000	0.117	0.063	0.013	0.510	0.401	0.239	-4.05%	15.69%
	20000	0.104	0.039	0.005	0.864	0.847	0.768	-3.11%	13.33%
BFM	100	0.128	0.070	0.019	0.047	0.018	0.002	-2.18%	11.54%
	200	0.107	0.060	0.014	0.034	0.011	0.000	-2.75%	8.91%
	600	0.093	0.046	0.008	0.042	0.012	0.001	-3.25%	6.62%
	1000	0.083	0.031	0.004	0.061	0.028	0.004	-3.39%	5.19%
	20000	0.023	0.006	0.000	0.099	0.049	0.011	-3.04%	1.38%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true value of λ_c and $\lambda^*_{useless} = 0$ in a misspecified model with an intercept and a useless factor. The true value of the cross-sectional R^2 is zero.

The Bayesian approach to Fama-MacBeth regressions successfully overcomes the hurdle of useless factors. As Table 2 demonstrates, both BFM-OLS and BFM-GLS are able to identify the spurious factor, with the posterior distribution providing credible confidence bounds with the proper control for the size of the test (e.g. for $T = 20,000$, the probability to reject the null of no risk premia attached to a useless factor when using a test with the nominal size of 10%, is 9.1%). Recognizing a useless factor, cross-sectional measures of fit are more conservative, and overall tighter.

Why does the bayesian approach work, when the frequentist fails? The argument is probably best summarized by Figure 1, that plots a posterior distribution of $\hat{\lambda}_{useless}$ for BFM from one of

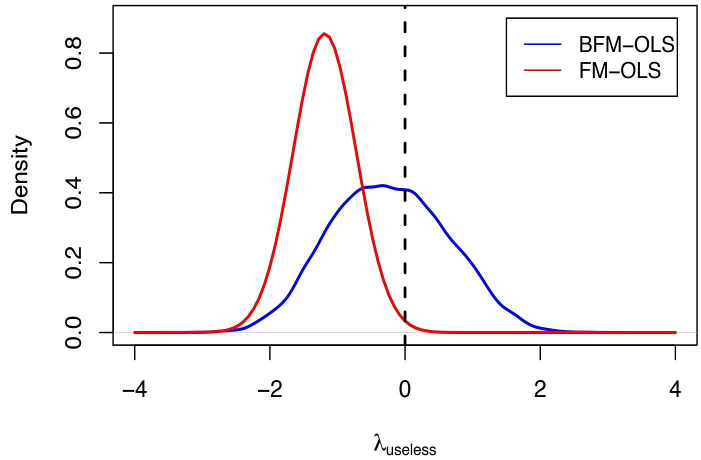


Figure 1: Risk premia estimates of a useless factor.

The graph presents the posterior distribution (blue line) of $\hat{\lambda}_{useless}$ from the BFM-OLS estimation in a misspecified model with a useless factor, based on a single simulation with $T = 1,000$. The red line depicts the asymptotic distribution of Fama-MacBeth estimate of risk premium under the normal approximation.

the simulations, along with the pseudo-true value of risk premium, defined as 0 in this case. In this particular simulation, Fama-MacBeth OLS estimate of $\lambda_{useless}$ is -1.19%, with Shanken-corrected t-statistics equal to -2.55, so according to traditional hypothesis testing, we would reject the null of $\lambda_{useless} = 0$ even at 1%. The posterior distribution of BFM estimates of risk premium (the blue line in Figure 1) behaves rather differently: it is centered around 0, and overall more spread out, with a confidence interval $(-1.603\%, 1.201\%)$. Intuitively, the main driving force behind it is the fact that in BFM $\boldsymbol{\beta}$ is updated continuously: when $\hat{\boldsymbol{\beta}}$ is close to zero, the posterior draws of $\boldsymbol{\beta}$ will be positive or negative randomly, which implies that the conditional expectation of $\boldsymbol{\lambda}$ in Equation 12 will also flip sign, depending on the draw. As a result, the posterior distribution of $\lambda_{useless}$ is centered around 0, and so is the confidence interval. The same logic applies to the case of BFM-GLS.

Finally, Table 3 combines the insights for true and irrelevant factors, and presents the simulation results for the most realistic model setup, that includes both useless and strong factors. As expected, in the conventional case of frequentist Fama-MacBeth estimation, the useless factor is often found to be a significant predictor of the asset returns: its OLS (GLS) t-statistic would be above a 5%-critical value in over 60% (80%) of the simulations. On the contrary, the bayesian confidence intervals have the right coverage, and reject the null of no risk premia attached to the spurious factor with frequency asymptotically approaching the size of the tests.

The crowding out of the true factors by the useless ones could also be an important empirical concern. When the model is misspecified, the presence of spurious factors can also bias the risk premia estimates for the strong ones, and often leads to their *crowding out* of the model. Panel A in Table 3 serves as a good illustration of this possibility, with risk premia estimates for the

Table 3: Tests of risk premia in a misspecified model with useless and strong factors

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

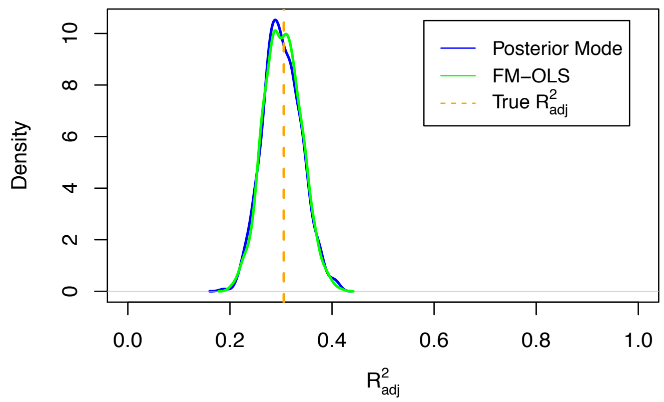
The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_c and λ_{strong}, $\lambda^*_{useless} \equiv 0$ in a misspecified model with an intercept, a strong and a useless factor. The true value of the cross-sectional R^2_{adj} is 30.55% (81.75%) for the OLS (GLS) estimation.

strong factor are clearly biased in the frequentist estimation by the identification failure in case of the frequentist approach. Again, in this case BFM provides reliable, albeit conservative, confidence bounds for model parameters.

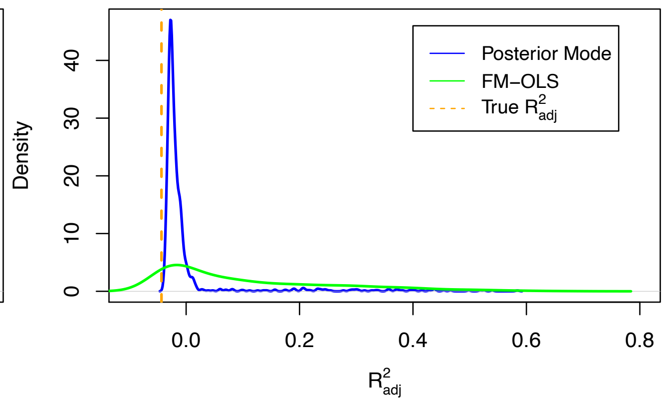
III.1.1 Evaluating cross-sectional fit

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

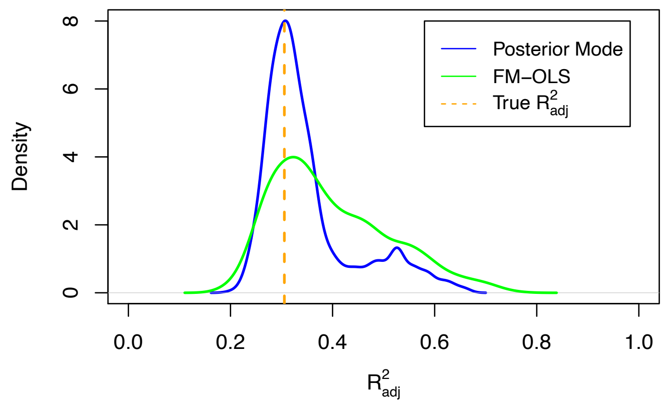
Figure 2 shows the distribution of cross-sectional OLS R^2 across a large number of simulations for the asymptotic case of $T = 20,000$ and a misspecified process for returns. As Panel (a) illustrates, if the model is strongly identified, the distribution of posterior mode of R^2 for BFM tends to coincide with that of conventional Fama-MacBeth procedure, as expected. The major difference emerges whenever a useless factor is included into the candidate set of variables. Indeed, it is well-known that in this case the distribution of conventional measures of fit is nonstandard and often inflated (Kleibergen and Zhan (2015)). This is further confirmed in Figures 2 b) and c) that show that under the presence of spurious factors, conventional Fama-MacBeth R^2 has an extremely spreadout right tail of the distribution, which makes it easy to find a substantial increase of fit whenever the model is simply not identified. This unfortunate property of the frequentist approach is not shared by the inference with BFM. Indeed, the mode of the posterior distribution of R^2 is generally tightly centered around the true values. The slight bump to the right tail of the distribution comes from the fact that whenever a spurious factor is included into the model with a small probability (based on t-statistic cut-off, this is equal to the size of the test, see e.g. Table 3,



(a) Model with a strong factor



(b) Model with a useless factor

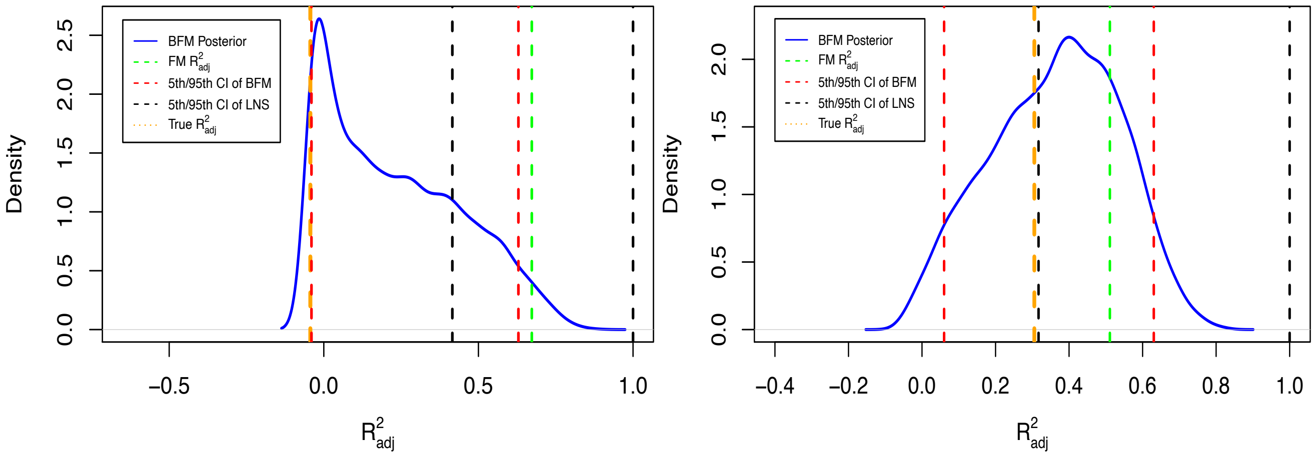


(c) Model with strong and useless factors

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

Figures (a)-(c) show the asymptotic distribution of cross-sectional R^2 under different model specifications across 1,000 simulations of sample size $T = 20,000$. Blue lines correspond to the distribution of the posterior mode for R^2_{adj}, while green lines depict the pointwise sample distribution of cross-sectional fit, evaluated at Fama-MacBeth risk premia estimates. The dark yellow dash line stands for the true value of R^2_{adj} in the model.

However, the pointwise distribution of cross-sectional R^2 across the simulations is only part of the story, as it does not reveal the in-sample estimation uncertainty and whether the confidence intervals are credible in reflecting it. While BFM incorporates this uncertainty directly into the shape of its posterior distribution, one needs to rely on bootstrap-like algorithms to build a similar analogue in the frequentist case. As frequentist benchmark, we use the approach Lewellen, Nagel, and Shanken (2010) to construct the confidence interval for R^2. Details on this procedure can be



(a) Model with a useless factor

(b) Model with strong and useless factors

Figure 3: The estimation uncertainty of cross-sectional R^2.
Figures (a) and (b) show the posterior density of the cross-sectional R^2 (blue) in one representative simulation, along with its 90% confidence interval (red). The dark yellow line denotes the true value of R^2, while the green line depicts its in-sample Fama-MacBeth estimate with the confidence interval constructed following Lewellen, Nagel, and Shanken (2010) (black).

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

In section A.6 of the Appendix we show that the performance of the Bayesian Fama-MacBeth method is robust to the use of a larger cross-sectional dimension – i.e. the above discussed properties hold in a larger N setting.

III.2 Bayes factors

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

Table 4: The probability of retaining risk factors using Bayes factors

		T	55%	57%	59%	61%	63%	65%
			Panel A: strong factors					
Jeffreys Prior	f_{strong}	200	0.813	0.784	0.758	0.722	0.693	0.662
		600	0.929	0.915	0.896	0.876	0.851	0.834
		1000	0.972	0.963	0.957	0.951	0.937	0.924
Spike-and-Slab Prior	f_{strong}	200	0.605	0.570	0.538	0.505	0.476	0.447
		600	0.853	0.830	0.807	0.784	0.762	0.735
		1000	0.954	0.940	0.928	0.912	0.896	0.875
			Panel B: useless factors					
Jeffreys Prior	$f_{useless}$	200	1.000	0.996	0.988	0.967	0.919	0.822
		600	0.998	0.998	0.995	0.988	0.977	0.943
		1000	1.000	1.000	1.000	0.994	0.983	0.965
Spike-and-Slab Prior	$f_{useless}$	200	0.325	0.188	0.106	0.057	0.024	0.013
		600	0.072	0.025	0.006	0.002	0.001	0.000
		1000	0.051	0.015	0.001	0.000	0.000	0.000
			Panel C: strong and useless factors					
Jeffreys Prior	f_{strong}	200	0.924	0.897	0.874	0.848	0.821	0.799
		600	0.988	0.985	0.976	0.974	0.965	0.958
		1000	0.998	0.996	0.996	0.995	0.992	0.987
	$f_{useless}$	200	0.984	0.960	0.910	0.811	0.702	0.584
		600	0.999	0.993	0.985	0.954	0.913	0.854
		1000	1.000	1.000	0.995	0.986	0.966	0.945
Spike-and-Slab Prior	f_{strong}	200	0.627	0.591	0.552	0.509	0.474	0.452
		600	0.860	0.830	0.802	0.783	0.758	0.727
		1000	0.956	0.942	0.927	0.911	0.895	0.875
	$f_{useless}$	200	0.260	0.128	0.071	0.031	0.019	0.010
		600	0.080	0.028	0.010	0.004	0.001	0.001
		1000	0.058	0.013	0.005	0.000	0.000	0.000

The table shows the frequency of retaining risk factors for different choice sets across 1,000 simulations of different size (T=200, 600, and 1,000). In Panel A, the candidate risk factor is truly cross-sectionally priced and strongly identified, while in Panel B they are not. Panel C summarizes the case of using both strong and useless candidate factors in the model. A candidate factor is retained in the model, if its marginal posterior probability, $p(\gamma_i = 1|data)$, is greater than a certain threshold, i.e. 55%, 57%, 59%, 61%, 63% and 65%.

Consider a cross-section of 25 portfolios that is actually loading on 2 systematic sources of risk, with the econometrician potentially observing at most only one of them, a strong (and priced) f_t. However, there is also a second candidate factor available, which is orthogonal to asset returns, and essentially useless. We compute Bayes factors, corresponding to each of the potential sources of risk, and document the empirical probability of retaining the variable in the model across 1000 simulations. Again, we consider models that contain either strong or useless factors, or a combination of both, and different sample sizes ($T = 200$, 600, and 1,000). In each case we run the Gibbs sampling algorithm derived using continuous spike-and-slab prior, and then approximate the marginal probability of each factor by the posterior mean of γ_j. The decision rule is based on a range of critical values, 55%-65%, such that whenever the posterior mean of γ_j is above a particular threshold, we retain the factor. Finally, we also compute the probability of retaining a factor under Jeffreys prior, which would be the standard in the literature.

Table 4 summarizes our findings. When only a true risk factor is included in the candidate set (Panel A), both Jeffrey's and spike-and-slab priors successfully identify it with a high probability, especially in large sample. For small sample sizes, however, Jeffreys prior clearly has a somewhat higher power of detecting a true risk factor. This outcome is not surprising, as the estimator does not impose additional shrinkage on the risk premium. The difference, however, vanishes, for larger sample sizes.

The difference between the two priors becomes drastic, whenever useless factors are included in the model (Panels B Panels B and C in Table 4). As discussed in Section II.2.1, since the matrix $\boldsymbol{\beta}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is nearly singular and its determinant goes to zero, under a flat prior the posterior probability of including a spurious factor in the model converges to 1 asymptotically. For example, the probability of misidentifying a spurious factor as being the true source of risk is almost 1 under Jeffreys prior, even for a very short sample. This in turn makes the overall process of model selection invalid.

Overall, we find the asymptotic behavior of the spike-and-slab prior encouraging for variable and model selection. While often somewhat conservative in very short samples, it successfully eliminates the impact of the spurious factors from the model, and identifies the true sources of risk.

IV Empirical Applications

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

IV.1 Some notable factor models

In this section we illustrate the differences between the frequentist and Bayesian FM estimation (both OLS and GLS) for several candidate models. In particular, we estimate a set of linear factor models on the returns of the standard 25 Fama-French portfolios, sorted by size and value, using frequentist and Bayesian Fama-MacBeth estimators. We use monthly data over the 1970:01-2017:12 sample for tradable factors and, whenever possible, nontradables. For factors available only at quarterly frequency, the sample is 1952:Q1-2017:Q3 (whenever possible). A full description of the data and models used can be found in Appendix B. Additional empirical results for other candidate factors and cross-sections (e.g. 25 Fama-French + 17 Industry portfolios) can be found in Appendix C.

Tables 5 and 6 summarize the performance of several leading factor models. For the classical FM approach, we report point estimates of risk premia with their Shanken-corrected t-statistics, and the cross-sectional R^2 along with its 90% confidence interval (constructed following the methodology of Lewellen, Nagel, and Shanken (2010)). For BFM, we report the posterior mean of risk premia estimates, and the posterior median and mode of R^2, along with the centered 90% posterior coverage. We chose to report both the median and the mode for cross-sectional fit, because the of the shape of its distribution, which is often heavily skewed.

Carhart (1997) 4-factor model. OLS and GLS Fama-MacBeth estimates of risk premia indicate that size, value, and momentum (SMB, HML, and UMD correspondingly) are significant drivers of the cross-section of test assets. The market factor does not command a significant risk premium, which is a typical finding for this model. Cross-sectional fit seems to be high, with R^2 over 70%, even though it comes with rather wide confidence bounds according to Lewellen, Nagel, and Shanken (2010) approach. The Bayesian estimation indicates that part of the model success is due to the fact that this cross section of test assets does not have much exposure to momentum, especially after one controls for the conventional Fama-French factors. While still marginally significant, its risk premium is substantially lower under both BFM and BFM-GLS estimation, with tighter bounds for R^2 too. On the contrary, both HML and SMB have virtually identical risk prices under both FM and Bayesian estimations.

Hou, Xue, and Zhang (2014) q-factor model emphasized the role of investment and profitability in matching the cross-section of equity returns, and true to the data, we find these factors strongly priced. Both estimation strategies give identical parameter estimates, and largely consistent measures of cross-sectional fit. This in turn implies that all the risk premia are strongly identified for this model, when estimated on a cross-section of 25 Fama-French portfolios. When industry portfolios are among the test assets (see Appendix C), risk premia and R^2 decline, and some of the parameters lose significance, but broadly speaking the model performs in a qualitatively similar way.

Liquidity-adjusted CAPM of Pastor and Stambaugh (2003): seems to suffer from identification failure, as the risk premium on the liquidity factor ceases to remain significant, when BFM is used in estimation. Wide confidence bounds and uncertain cross-sectional fit provide a stark difference to

Table 5: Tradable factors and 25 Fama-French portfolios, sorted by size and value

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

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

Table 6: Nontradable factors and 25 Fama-French portfolios, sorted by size and value

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

The table summarises risk premia estimates and cross-sectional fit for a selection of models with nontradable risk factors on a cross-section of 25 Fama-French quarterly excess returns. Each model is estimated via OLS and GLS. We report point estimates and 5% confidence intervals for risk premia, which are constructed based on the asymptotic normal distribution, and cross-sectional R^2 and its (5%, 95%) confidence level constructed as in Lewellen, Nagel, and Shanken (2010) for FM estimation. In Bayesian Fama-MacBeth estimation, we provide the posterior mean of λ, denoted by $\bar{\lambda}_j$, its (2.5%, 97.5%) credible intervals, the posterior mode and median of the cross-sectional R^2, as well as its (5%, 95%) credible intervals. *, ** and *** denote significance at the 90%, 95% and 99% level, respectively.

the pointwise estimates and their seemingly high significance levels under the standard frequentist approach.

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

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

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

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

IV.2 Sampling two quadrillion models

We now turn our attention to a large cross-section of candidate asset pricing factors. In particular, we focus on 51 (both tradable and non-tradable) monthly factors available from October 1973 to December 2016 (i.e. $T \simeq 600$). Factors are described in details in Table B.1 of Appendix B. In choosing the cross-section of assets to price we follow Lewellen, Nagel, and Shanken (2010) and employ 25 Fama-French size and book-to-market portfolios plus 30 Industry portfolio (i.e. $N = 55$). Since we do not restrict the maximum number of factors to be included, all the possible combinations of factors give us a total of 2^{51} possible specifications i.e. 2.25 quadrillion models. Note that each model involves 55 time series regressions and one cross-section regression i.e. we jointly evaluate the equivalent of 126 quadrillion regressions.

We employ the continuous spike-and-slab approach of section II.2.3, since it is the most suited for handling a very large number of possible models, and report both the posterior probability (given the data) of each factor (i.e. $\mathbb{E}[\gamma_j|\text{data}], \forall j$) as well as the posterior means of the factors' risk premia (i.e. $\mathbb{E}[\lambda_j|\text{data}], \forall j$) computed as the Bayesian Model Average[16] across all the models

[16]If we are interested in some quantity Δ that is well-defined for every model $j = 1, ..., M$ (e.g. risk premia,

considered.

The posterior evaluation is performed and reported over a wide range for the parameter (ψ in equation (16)) that controls the degree of shrinkage of potentially useless factors' risk premia: from $\psi = 1$ (i.e. very strong shrinkage) to $\psi = 100$ (making the shrinkage virtually irrelevant). We consider a value of ψ in the 10-20 range as a reasonable benchmark, since it is equivalent to a prior standard deviation for the (annualised) risk premia of about 9.1%-12.7% for the typical strong factor, implying that (annualized) factor Sharpe ratios as large as 0.92-2.62 are within the centered 95% prior coverage.

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% – i.e. a priori we have maximum uncertainty about whether a factor should be included or not.[17]

Figure 4 plots the posterior probabilities of the 51 factors as a function of the parameter ψ. The corresponding values are reported in Table 7. Overall, the inclusion of only three factors finds substantial support in our empirical analysis. First, the celebrated Fama-French HML (high-minus-low), designed to capture the so-called 'value premium', is a strong determinant of the cross-section of asset returns. For $\psi = 10$ (a reasonable benchmark) its posterior probability is about 92.1%, and only for very strong shrinkage ($\psi = 1$) the posterior probability gets reduced to 86.6%. Second, the market factor, in the version of Daniel, Mota, Rottke, and Santos (2018) (MKT*, that is meant to have hedged out the unpriced risk contained in the market index), has also high posterior probability (68.3% for $\psi = 10$). Instead, the simple market factor (MKT) seems to be driven out by MKT*. Third, albeit to a lesser extent, SMB*, the Daniel, Mota, Rottke, and Santos (2018) version of the small-minus-big Fama-French factor (meant to capture the so called 'size' premium), seems also to contain relevant information for pricing the cross-section of asset returns, with a posterior probability in the 50-60% range for small values of ψ. Beside the ones above mentioned, all other factors have posterior probabilities of about 50% or less for all values of ψ. Interestingly, the results are not very sensitive to the choice of ψ.

In addition to the posterior probabilities of the factors, Table 7 reports the posterior means of the factor risk premia computed as Bayesian Model Average i.e. 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),

---

maximum Sharpe ratio etc.), from Bayes' theorem we have

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

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

[17] Using a $Beta(2, 2)$, that still implies a prior probability of factor inclusion of 50%, but lower probabilities for very dense and very sparse models, we obtain virtually identical results. Furthermore, using a prior in favour of more sparse factor models (i.e. a $Beta(2, 8)$), the empirical findings are very similar to the ones reported. These additional results are reported in the Online Appendix available at `https://tinyurl.com/w5s5dy2`.

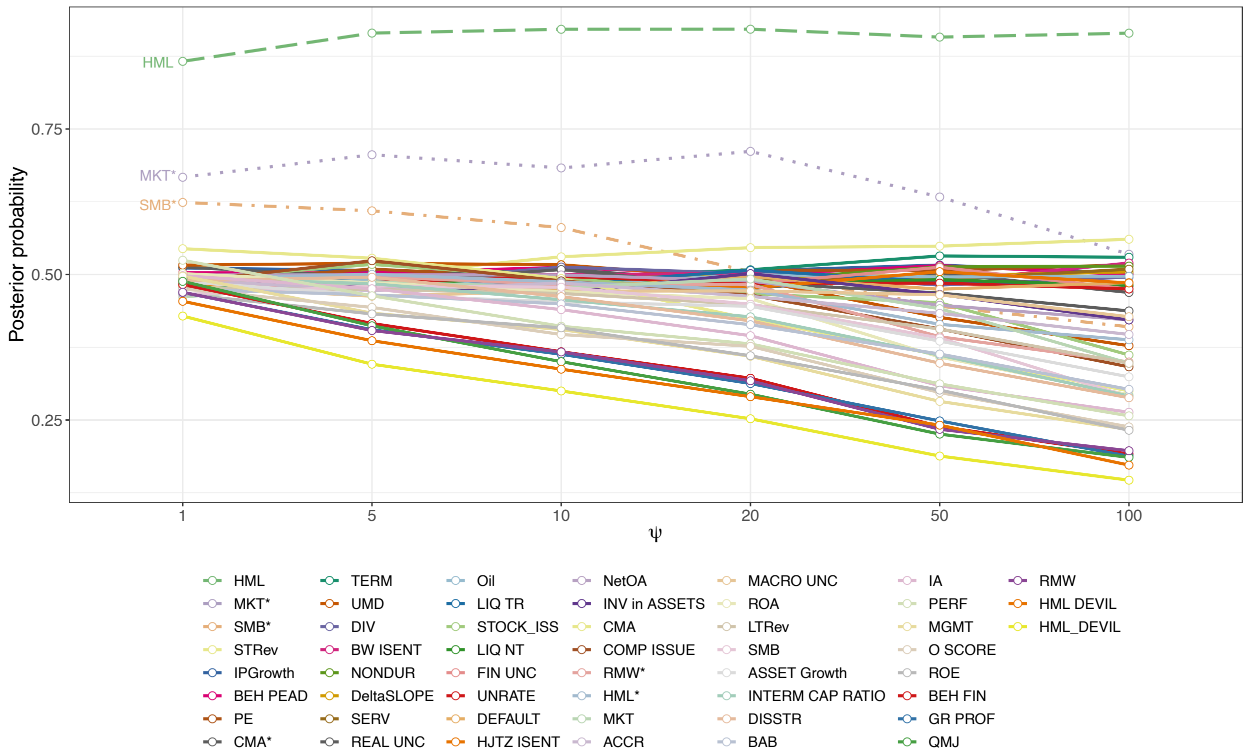


Figure 4: Posterior factor probabilities

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, estimated over the 1973:10-2016:12 sample using a cross-section of 25 Fama-French size and book-to-market and 30 Industry test asset portfolios, computed using the continuous spike and slab approach of section II.2.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The 51 factors considered are described in Table B.1 of Appendix B. The prior distribution for the j-th factor inclusion is a $Beta(1, 1)$, yielding a prior expectation of $\gamma_j = 50\%$. Posterior probabilities are plotted for $\psi \in [1,\ 100]$.

Geweke (1999), Madigan and Raftery (1994)). The results are not very sensitive to the choice of ψ, except when considering very small values of the shrinkage parameter ψ, since in this case posterior means are shrunk toward zero. Interestingly, the estimated price of risk for the market factor (MKT*) is positive, despite it being very often estimated as a negative quantity when considering multifactor models, and not dissimilar from the market excess return over the same period. More generally, there is a clear pattern in cross-sectionally estimated (i.e. ex post) factor risk premia and their simple time series average estimates (reported in the last column of Table 7): for the robust three factors (HML, MKT*, and SMB*) ex post risk premia are very similar to the time series estimnates multiplied by their posterior probabilities, while the opposite holds true for the other factors. In other words, robust factors seems to price themselves well (since theoretically their own beta is one), while other factors don't.

IV.3 Estimating 2.6 million sparse factor models

Instead of drawing unrestricted factor models specifications, we now constraint models to have a maximum of 5 factors – i.e. we are imposing sparsity on the linear factor models, as in most

Table 7: Posterior factor probabilities, $\mathbb{E}[\gamma_j|\text{data}]$, and risk premia in 2.25 quadrillion models

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

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

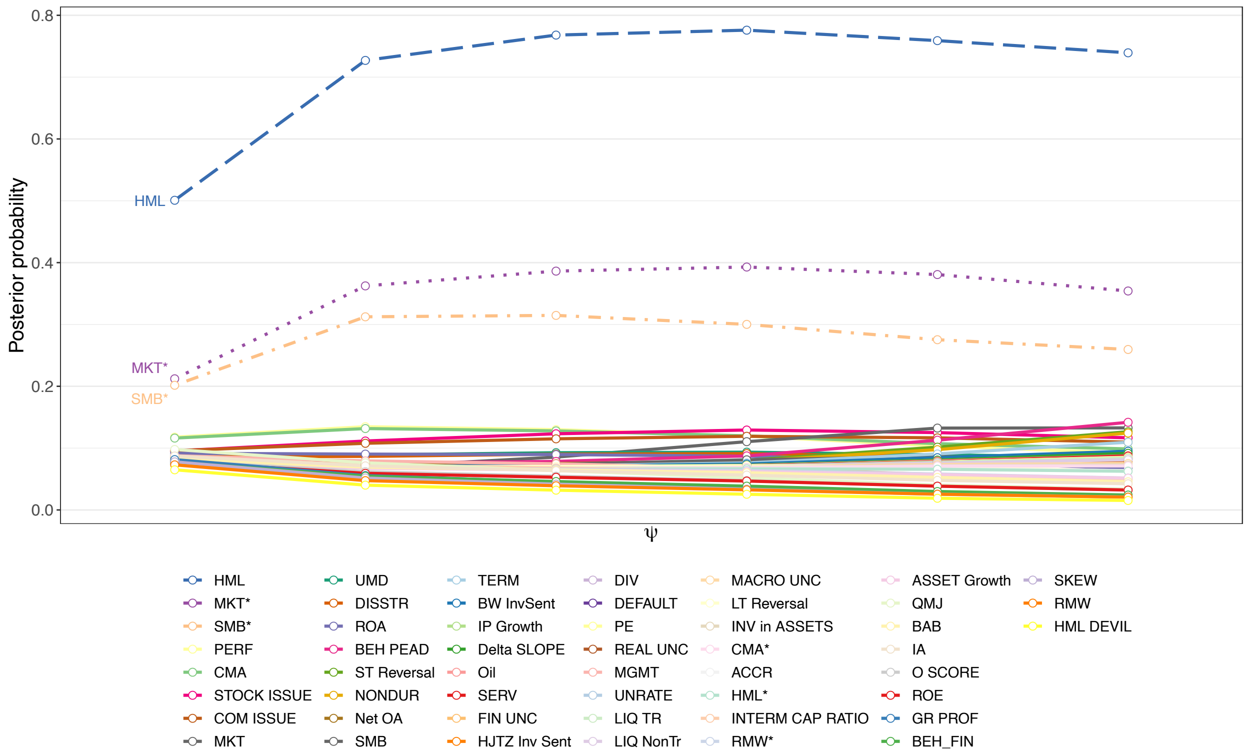


Figure 5: Posterior factor probabilities

Posterior probabilities of factors, $\Pr[\gamma_j = 1|\text{data}]$, estimated over the 1973:10-2016:12 sample using a cross-section of 25 Fama-French size and book-to-market and 30 Industry test asset portfolios, computed using the Dirac spike and slab approach of section II.2.2, 51 factors, and all possible models with up to 5 factors, yielding about 2.6 million candidate models. The prior probability of a factor being included is about 10.38%. The 51 factors considered are described in Table B.1 of Appendix B. Posterior probabilities are plotted for $\psi \in [1,\ 100]$.

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

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

where we have assigned equal prior probability to all possible specifications and $p(data|\boldsymbol{\gamma}_j)$ denotes the marginal likelihood of the j-th model. To both simplify the numerical computation, and to illustrate the qualities of the approach, we use the Dirac spike-and-slab prior of section II.2.2 since we can leverage its closed form solution for the marginal likelihoods.[18]

Posterior factor probabilities are reported in Figure 5 and, jointly with the Bayesian model averaging of risk premia across the sparse models, in Table C15. Note that in this case all factors

[18] Alternatively, one could use the continuous spike-and-slab approach employed in the previous subsection, and drop the draws of specifications with more than 5 factors, but this significantly reduces computational efficiency.

have an ex ante probability of being included equal to 10.38%. The results are strikingly similar to the ones in Table 7 and Figure 4: as before, only for three factors (HML, MKT* and SMB*) we observe a marked increase in the posterior probability of inclusion after observing the data. Furthermore, these factors seems to price themselves well – i.e. the BMA of their risk premia are very similar to the sample average of their excess returns – while other factors do not.

IV.4 A robust factors model

The previous subsections suggest that only a small number of factors – HML, MKT* and, to a lesser extent, SMB* – are robust explanators of the cross-section of asset returns. Furthermore, Table 8, that reports the ten factor model specifications with the highest posterior probabilities under continuous spike-and-slab approach, with ψ equal to 20 (a $\psi = 10$, as shown in Table C16 of the Appendix, yields very similar results). It shows that these robust factors tend to be almost always included in the most likely models: HML is featured in all ten specifications, and both MKT* and SMB* are included in seven or eight most likely models. Note that the posterior probabilities in Table 8 might appear small in absolute terms, but are actually of magnitude much larger than the prior model probabilities (equal to one over the number of models considered).

Therefore, a natural question is whether the three factors identified as robust in the above analysis do indeed deliver a significantly better cross-sectional asset pricing model. We answer this questions by comparing the performance of a three factor model with HML, MKT* and SMB* as factors, to the one of several notable factor models. In particular, Table 9 reports the model posterior probabilities for the specifications considered, i.e. the probability of any of these models being the true data generating process. Strikingly, for almost any value of ψ, the model posterior probabilities are in the single digits range for all models but the robust factors one: the probability of this specification is always higher than 85% except when using a very strong shrinkage (in which case it is reduced to 64%). Furthermore, for ψ in the most salient range (10-20), the posterior probability of the robust factors model is about 90%.

IV.5 The sparsity of the linear factor model

We have shown that a robust model with only three – robust – factors is much more likely to explain the cross-section of 25 Fama-French size-B/M portfolios and 30 industry portfolios than all the notable models considered (see Table 9). Nevertheless, are only three factors sufficient to deliver an accurate explanation in the cross section?

Thanks to our Bayesian method, this question can be easily answered. In particular, by using our estimations of about 2.25 quadrillion models and their posterior probabilities, we can compute the posterior distribution of the dimensionality of the 'true' model. That is, for any integer number between one and fifty-one, we can compute the posterior probability of the linear factor model being a function of that number of factors.

Figure 6 reports the posterior distributions of the model dimensionality for various values of ψ. These distributions are also summarized in Table 10. For the most salient values of ψ (10 and 20),

Table 8: Factor Models with highest posterior probability (continuous spike-and-slab, $\psi = 20$)

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

Factors and posterior model probabilities of ten most likely specifications computed using the continuous spike and slab approach of section II.2.3, $\psi = 20$, 51 factors, and all possible models with up to 5 factors, yielding about 2.25 quadrillion models and a model prior probability of the order of 10^{-16}. Specifications organised by columns with the symbol ✓ indicating that the factor in the corresponding row is included. The data is monthly, 1973:10 to 2016:12. Test assets: cross-section of 25 Fama-French size and book-to-market and 30 Industry portfolios. The 51 factors considered are described in Table B.1 of Appendix B.

the posterior mean of the number of factors in the true model is in the 24-25 range, and the 95% posterior credible intervals are contained in the 17 to 31 factors range. That is, there is substantial evidence that the linear factor model is *dense* in the space of factors considered: given the factors

Table 9: Posterior probabilities of notable models vs. robust factors

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

Posterior model probabilities for the specifications in the first column, for different values of ψ, computed using the Dirac spike-and-slab prior. The models and their factors are described in Appendix B. The model in the last row uses the HML, MKT* and SMB* factors described in Table B1. The data is monthly, 1973:10 to 2016:12. Test assets: cross-section of 25 Fama-French size and book-to-market and 30 Industry portfolios.

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 6 and Table 10 show, for very large ψ, i.e. with basically a flat prior for factor risk premia, the posterior dimensionality is reduced. This is due to two phenomena we have already outlined. First, if some of the factors are useless (and our analysis points in this direction), under a flat prior they do tend to have a higher posterior probability and drive out the true sources of priced risk. Second, a flat prior for the risk premia can generate a "Bartlett Paradox" (see the discussion in section II.2.1 and Bartlett (1957)).

Table 10: Posterior dimensionality of linear factor model

	ψ:					
	1	5	10	20	50	100
mean	25.62	25.02	24.36	23.52	21.70	19.88
median	26	25	24	23	22	20
2.5%	19	18	17	17	15	13
5%	20	19	19	18	16	14
95%	31	31	30	29	28	26
97.5%	32	32	31	30	29	27

Summary statistics of the posterior density of model dimensionality for various values of ψ. Estimated over the 1973:10-2016:12 sample using a cross-section of 25 Fama-French size and book-to-market and 30 Industry test asset portfolios, using the continuous spike and slab approach of section II.2.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1, 1)$.

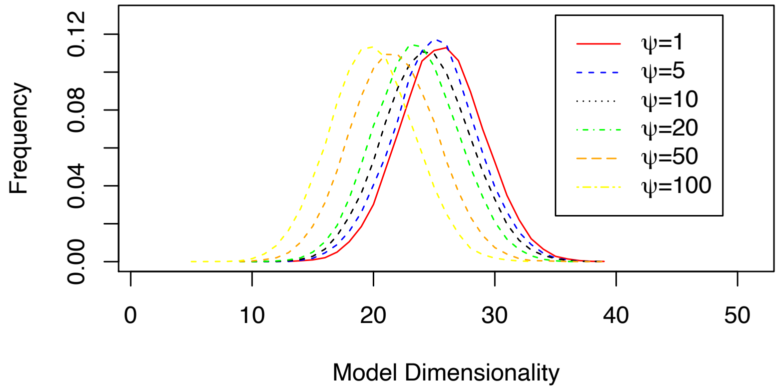


Figure 6: Posterior density of the dimensionality of the linear factor model

Posterior density of the true model having the number of factors listed on the horizontal axis. Estimated over the 1973:10-2016:12 sample using a cross-section of 25 Fama-French size and book-to-market and 30 industry test asset portfolios, computed using the continuous spike and slab approach of section II.2.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1, 1)$, yielding a prior expectation for γ_j equal to 50%. The 51 factors considered are described in Table B.1 of Appendix B. Posterior densities are plotted for $\psi \in [1,\ 100]$.

Note that if the factors proposed in the literature were to capture different and uncorrelated sources of risk, one might worry that a dense model in the space of factors could imply unrealistically high Sharpe ratios (see e.g. the discussion in Kozak, Nagel, and Santosh (2019)). Since, given a factor model, the SDF-implied maximum Sharpe ratio is just a function of the factors' risk premia and covariance matrix, our Bayesian method allows to construct the posterior distribution of the maximum Sharpe ratio for each of the 2 quadrilion models considered. Therefore, using the posterior probabilities of each possible model specification, we can actually construct the (Bayesian Model Averaging) posterior distribution of the SDF-implied maximum Sharpe ratio (conditional on the data only).

Figure 7 and table 11 report, respectively, the posterior distribution of the (annualized) SDF-implied maximum Sharpe ratio and its summary statistics for several values of the parameter ψ. Except when a very strong shrinkage (small ψ) is imposed (and hence risk premia, and consequently Sharpe ratios, are shrunk toward zero) the posterior distributions of the Sharpe ratio are quite similar for all values of ψ. Furthermore, despite the model being dense in the space of factors, the posterior maximum Sharpe ratio does not appear to be unrealistically high: e.g. for $\psi \in [10, 20]$ its posterior mean is about 0.65–0.72 and the 95% posterior credible intervals are in the 0.43–1.03 range. Interestingly, Ghosh, Julliard, and Taylor (2016, 2018) provides a non-parametric estimate of the pricing kernel, extracted using an information-theoretic approach and wide cross-sections of equity portfolios, and find SDF-implied maximum Sharpe ratios of very similar magnitude.

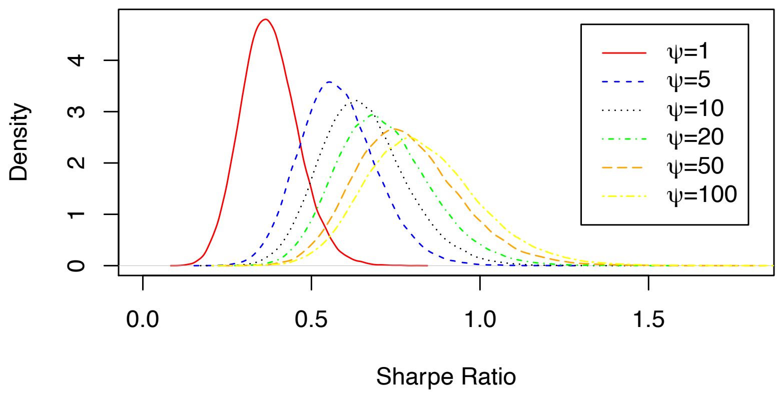


Figure 7: Posterior density of the Sharpe ratio implied by the linear factor model

Posterior density of the Sharpe ratio implied by the linear factor model for various values of $\psi \in [1,\ 100]$. Estimated over the 1973:10-2016:12 sample using a cross-section of 25 Fama-French size and book-to-market and 30 industry test asset portfolios, computed using the continuous spike and slab approach of section II.2.3, 51 factors described in Table B.1 of Appendix B, and Bayesian Model Averaging of the $2^{51} \approx 2.25$ quadrillion possible models. The prior for each factor inclusion is a $Beta(1, 1)$.

Table 11: Posterior distribution of the Sharpe ratio implied by the linear factor model

	ψ:					
	1	5	10	20	50	100
mean	0.38	0.58	0.65	0.72	0.79	0.83
median	0.37	0.57	0.65	0.71	0.77	0.82
2.5%	0.22	0.38	0.43	0.47	0.51	0.54
5%	0.24	0.41	0.46	0.51	0.55	0.58
95%	0.52	0.78	0.88	0.97	1.07	1.13
97.5%	0.55	0.83	0.94	1.03	1.13	1.20

Summary statistics of the posterior distribution of the maximal Sharpe ratio implied by the linear factor model for various values of $\psi \in [1,\ 100]$. Estimated over the 1973:10-2016:12 sample using a cross-section of 25 Fama-French size and book-to-market and 30 Industry test asset portfolios, computed using the continuous spike and slab approach of section II.2.3, 51 factors described in Table B.1 of Appendix B, and Bayesian Model Averaging of the $2^{51} \approx 2.25$ quadrillion possible models. The prior for each factor inclusion is a $Beta(1, 1)$.

V Extensions

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

First, based on economic considerations, one might possibly want to bound the maximum risk premia (or the maximum Sharpe ratios) associated with the factors. This can be achieved by replacing the Gaussian distributions in our spike-and-slab priors with (rescaled and centered) Beta distributions, since the latter have bounded support. Furthermore, for the sake of expositional simplicity and closed form solutions, we have focused on regularising spike and slab priors with exponential tails. Nevertheless, our approach, that shrinks useless factors based on their correlation with asset returns, could be as well implemented using polynomial tailed (i.e. heavy-tailed) mixing priors (see Polson and Scott (2011) for a general discussion of priors for regularisation and shrinkage).[19] The rational for using heavy tailed priors is that, when the likelihood has thick tails while the prior has thin tail, if the likelihood peak moves too far from the prior mean, the posterior eventually reverts toward the prior. This is illustrated in Figure D2 of the Appendix. Nevertheless, note that this mechanism (first pointed out in Jeffreys (1961)) is actually desirable in our settings in order to shrink the risk premia of useless factors toward zero.[20]

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

Third, again thanks to the hierarchical structure of our method, time varying time series betas could be accommodated by adopting the time varying parameters approach of Primiceri (2005) in the time series step. And since in our approach the asset specific expected risk premia are a parameter estimated in the time series step, this extension would also allow for time variation in asset risk premia. Furthermore, albeit this would increase the numerical complexity of the cross-sectional inference step, the time varying parameters formulation could also be used for the modelling of the factor risk premia.

[19]For example, albeit alternatives with more desirable properties exist, our spike and slab could be implemented using a Cauchy prior with location parameter set to zero and scale parameter proportional to ψ_j as defined in equation (16).

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

VI Conclusions

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

We have applied our approach to the study of more than two quadrillion factor model specifications and have found that: 1) only a handful of factors (the Fama and French (1992) "high-minus-low" proxy for the value premium, and the adjusted versions of both market and size factors of Daniel, Mota, Rottke, and Santos (2018)) seem to be robust explanators of the cross-sections of asset returns; 2) jointly, the four robust factors provide a model that is, compared to the previous empirical literature, one order of magnitude more likely to have generated the observed asset returns (it's posterior probability is about 90%); 3) with very high probability the "true" latent stochastic discount factor is dense in the space of factors proposed in the previous literature i.e. capturing its characteristics requires the use of 24-25 factors (at the posterior mean of the SDF sparsity); 4) despite being dense in the space of factors, the SDF-implied maximum Sharpe ratio is not excessive, suggesting a high degree of commonality, in terms of captured risks, among the factors in the zoo.

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

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

References

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

Asness, C., and A. Frazzini (2013): "The Devil in HMLs Details," *Journal of Portfolio Management*, 39, 49–68.

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

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

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

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

Ball, R. (1985): "Anomalies in Relationships between Securities' Yields and Yield-Surrogates," *Journal of Financial Economics*, 6, 103–126.

Barillas, F., and J. Shanken (2018): "Comparing Asset Pricing Models," *The 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," *The Review of Economic Studies*, 81, 608–650.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989): "Empirical Test of the Consumption-Oriented CAPM," *The Journal of Finance*, 44(2), 231–262.

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

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

Campbell, J. Y. (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104(2), 298–345.

Campbell, J. Y., J. Hilscher, and J. Szilagyi (2008): "In Search of Distress Risk," *Journal of Finance*, 63, 2899–2939.

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

Chan, K., N.-F. Chen, and D. A. Hsieh (1985): "An Exploratory Investigation of the Firm Size Effect," *Journal of Financial Economics*, 14, 451–71.

Chen, L., R. Novy-Marx, and L. Zhang (2010): "A New Three-Factor Model," working paper.

Chen, N.-F., S. A. Ross, and R. Roll (1986): "Economic Forces and the Stock Market," *Journal of Business*, 59, 383–403.

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 (forthcoming): "On Comparing Asset Pricing Models," *The Journal of Finance*.

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

Cooper, M. J., H. Gulen, and M. J. Schill (2008): "Asset Growth and the Cross-Section of Stock Returns," *Journal of Finance*, 63, 1609–52.

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

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

Daniel, K., L. Mota, S. Rottke, and T. Santos (2018): "The Cross-Section of Risk and Return," working paper.

Daniel, K., and S. Titman (2006): "Market reactions to tangible and intangible information," *Journal of Finance*, 61, 1605–43.

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

FAMA, E. F., AND K. FRENCH (2008): "Dissecting Anomalies," *Journal of Finance*, 63, 1653–78.

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

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *The 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.

——— (2016): "Dissecting Anomalies with a Five-Factor Model," *Review of Financial Studies*, 29, 69–103.

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

FERSON, W., AND C. HARVEY (1991): "The Variation of Economic Risk Premiums," *Journal of Political Economy*, 99, 385–415.

FRAZZINI, A., AND L. H. PEDERSEN (2014): "Betting against Beta," *Journal of Financial Economics*, 111(1), 1–25.

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.

GERTLER, M., AND E. GRINOLS (1982): "Unemployment, Inflation, and Common Stock Returns," *Journal of Money, Credit and Banking*, 14, 216–33.

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

GHOSH, A., C. JULLIARD, AND A. TAYLOR (2016): "What is the Consumption-CAPM missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models," *Review of Financial Studies*, 30, 442–504.

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

GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2018): "Economic Predictions with Big Data: The Illusion of Sparsity," FRB of New York Staff Report No. 847.

GIBBONS, M. R., S. A. ROSS, AND J. SHANKEN (1989): "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57(5), 1121–1152.

GIGLIO, S., G. FENG, AND D. XIU (2019): "Taming the factor Zoo," *Journal of Finance*, forthcoming.

GIGLIO, S., AND D. XIU (2018): "Asset Pricing with Omitted Factors," working paper.

GOSPODINOV, N., R. KAN, AND C. ROBOTTI (2014): "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors," *The Review of Financial Studies*, 27(7), 2139–2170.

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

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

HALL, R. E. (1978): "The Stochastic Implications of the Life Cycle Permanent Income Hypothesis," *Journal of Political Economy*, 86(6), 971–87.

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," *The Journal of Finance*, 72(4), 1399–1440.

HARVEY, C. R., Y. LIU, AND H. ZHU (2016): "... and the Cross-Section of Expected Returns," *The Review of Financial Studies*, 29(1), 5–68.

HE, A., D. HUANG, AND G. ZHOU (2018): "New Factors Wanted: Evidence from a Simple Specification Test," working paper, available at SSRN: https://ssrn.com/abstract=3143752.

He, Z., B. Kelly, and A. Manela (2017): "Intermediary Asset Pricing: New Evidence from Many Asset Classes," *Journal of Financial Economics*, 126, 1–35.

Hirshleifer, D., H. Kewei, S. Teoh, and Y. Zhang (2004): "Do Investors Overvalue Firms with Bloated Balance Sheets?," *Journal of Accounting and Economics*, 38, 297–331.

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., C. Xue, and L. Zhang (2015): "Digesting Anomalies: An Investment Approach," *The Review of Financial Studies*, 28(3), 650–705.

Huang, D., F. Jiang, J. Tu, and G. Zhou (2015): "Investor Sentiment Aligned: A Powerful Predictor of Stock Returns," *Review of Financial Studies*, 28, 791–837.

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," *The Annals of Statistics*, 33(2), 730–773.

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

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

Jegadeesh, N., and S. Titman (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance*, 48, 65–91.

——— (2001): "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations," *Journal of Finance*, 56, 699–720.

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

Jurado, K., S. Ludvigson, and S. Ng (2015): "Measuring Uncertainty," *American Economic Review*, 105, 1177–1215.

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," *the Journal of Finance*, 54(1), 203–235.

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

Kelly, B., S. Pruitt, and Y. Su (2019): "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, 134, 501–524.

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

Kleibergen, F., and Z. Zhan (2015): "Unexplained Factors and their Effects on Second Pass R-squareds," *Journal of Econometrics*, 189(1), 101–116.

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

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

Langlois, H. (2019): "Measuring Skewness Premia," *Journal of Financial Economics*.

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

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

Lintner, J. (1965): "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20, 587–615.

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

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.

Novy-Marx, R. (2013): “The Other Side of Value: The Gross Profitability Premium,” *Journal of Financial Economics*, 108, 1–28.

Ohlson, J. A. (1980): “Financial Ratios and the Probabilistic Prediction of Bankruptcy,” *Journal of Accounting Research*, 18, 109–131.

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.

Ritter, J. R. (1991): “The Long-Run Performance of Initial Public Offerings,” *Journal of Finance*, 46, 3–27.

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

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

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

Sharpe, W. F. (1964): “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk,” *Journal of Finance*, 19(3), 425–42.

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

Sloan, R. (1996): “Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?,” *The Accounting Review*, 71, 289–315.

Stambaugh, R. F., and Y. Yuan (2016): “Mispricing Factors,” *Review of Financial Studies*, 30, 1270–1315.

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

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

Titman, S., K. J. Wei, and F. Xie (2004): “Capital Investments and Stock Returns,” *Journal of Financial and Quantitative Analysis*, 39, 677–700.

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

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

A Additional Derivations

A.1 Derivation of the posterior distribution in section II.1

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

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

After assigning the Jeffreys' prior for $(\boldsymbol{B}, \boldsymbol{\Sigma})$: $\pi(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{N+1}{2}}$, we simplify the likelihood function by exploiting the fact that OLS estimated residuals are orthogonal to the regressors:

$$\begin{aligned}(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{B})^\top(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{B}) &= [\boldsymbol{R} - \boldsymbol{F}\boldsymbol{\hat{B}_{ols}} - \boldsymbol{F}(\boldsymbol{B} - \boldsymbol{\hat{B}_{ols}})]^\top[\boldsymbol{R} - \boldsymbol{F}\boldsymbol{\hat{B}_{ols}} - \boldsymbol{F}(\boldsymbol{B} - \boldsymbol{\hat{B}_{ols}})] \\ &= (\boldsymbol{R} - \boldsymbol{F}\boldsymbol{\hat{B}_{ols}})^\top(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{\hat{B}_{ols}}) + (\boldsymbol{B} - \boldsymbol{\hat{B}_{ols}})^\top \boldsymbol{F}^\top \boldsymbol{F}(\boldsymbol{B} - \boldsymbol{\hat{B}_{ols}}) \\ &= T\boldsymbol{\hat{\Sigma}} + (\boldsymbol{B} - \boldsymbol{\hat{B}_{ols}})^\top \boldsymbol{F}^\top \boldsymbol{F}(\boldsymbol{B} - \boldsymbol{\hat{B}_{ols}}),\end{aligned}$$

where

$$\boldsymbol{\hat{B}_{ols}} = \begin{pmatrix} \boldsymbol{\hat{a}}^\top \\ \boldsymbol{\hat{\beta}}^\top \end{pmatrix} = (\boldsymbol{F}^\top\boldsymbol{F})^{-1}\boldsymbol{F}^\top\boldsymbol{R}, \;\; \boldsymbol{\hat{\Sigma}_{ols}} = \frac{1}{T}(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{\hat{B}_{ols}})^\top(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{\hat{B}_{ols}}).$$

Therefore, the posterior distribution in the first step is

$$\begin{aligned}p(\boldsymbol{B}, \boldsymbol{\Sigma}|data) &\propto (2\pi)^{-\frac{NT}{2}} |\boldsymbol{\Sigma}|^{-\frac{T+N+1}{2}} \exp\left\{-\frac{1}{2} tr[\boldsymbol{\Sigma}^{-1}(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{B})^\top(\boldsymbol{R} - \boldsymbol{F}\boldsymbol{B})]\right\} \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{T+N+1}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(T\boldsymbol{\hat{\Sigma}})]} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(\boldsymbol{B} - \boldsymbol{\hat{B}_{ols}})^\top \boldsymbol{F}^\top \boldsymbol{F}(\boldsymbol{B} - \boldsymbol{\hat{B}_{ols}})]}.\end{aligned}$$

Hence the posterior distribution of $\boldsymbol{B}$ conditional on data and $\boldsymbol{\Sigma}$ is

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

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

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

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

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

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

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

where data in the second step include $(\boldsymbol{a}, \boldsymbol{\beta_f})$ drawn from the first step. Assuming the diffuse

Jeffreys' prior $\pi(\boldsymbol{\lambda},\sigma^2) \propto \frac{1}{\sigma^2}$ the posterior distribution of $(\boldsymbol{\lambda},\sigma^2)$ is

$$\begin{aligned} p(\boldsymbol{\lambda},\sigma^2|data,\boldsymbol{B},\boldsymbol{\Sigma}) &\propto (\sigma^2)^{-\frac{N+2}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{a}-\boldsymbol{\beta\lambda})^\top(\boldsymbol{a}-\boldsymbol{\beta\lambda})\right\} \\ &= (\sigma^2)^{-\frac{N+2}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{a}-\boldsymbol{\beta}\hat{\boldsymbol{\lambda}}+\boldsymbol{\beta}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}))^\top(\boldsymbol{a}-\boldsymbol{\beta}\hat{\boldsymbol{\lambda}}+\boldsymbol{\beta}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}))\right\} \\ &= (\sigma^2)^{-\frac{N+2}{2}} \exp\left\{-\frac{N\hat{\sigma}^2}{2\sigma^2}\right\} \exp\left\{-\frac{(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^\top\boldsymbol{\beta}^\top\boldsymbol{\beta}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})}{2\sigma^2}\right\}, \end{aligned}$$

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

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

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

hence obtaining the posterior distribution in equation (13).

A.2 Non-spherical pricing errors

Our framework can also easily accommodate non-spherical cross-sectional pricing errors. To see this note that, under the null of the model, we can rewrite equation (1) as $\boldsymbol{R_t} = \boldsymbol{\beta\lambda} + \boldsymbol{\beta_f f_t} + \boldsymbol{\epsilon_t}$. Consider the cross-sectional regression, and let $\mathbb{E}_T$ define the sample mean operator. Since $\mathbb{E}_T[\boldsymbol{R_t}] = \boldsymbol{\beta\lambda} + \mathbb{E}_T[\boldsymbol{\beta_f f_t}] + \mathbb{E}_T[\boldsymbol{\epsilon_t}] = \boldsymbol{\beta\lambda} + \mathbb{E}_T[\boldsymbol{\epsilon_t}]$, the pricing error $\boldsymbol{\alpha}$ should be equal to $\mathbb{E}_T[\boldsymbol{\epsilon_t}]$. Hence, under the hypothesis that the model is correctly specified, and in the spirit of the central limit theorem, a suitable distributional assumption for the pricing errors $\boldsymbol{\alpha}$ in the second step is

$$\boldsymbol{\alpha}|\boldsymbol{\Sigma} \sim N\left(\boldsymbol{0_N}, \frac{1}{T}\boldsymbol{\Sigma}\right).$$

Implying the distribution of $\bar{\boldsymbol{R}} \equiv \mathbb{E}_T[\boldsymbol{R_t}] \sim \mathcal{N}(\boldsymbol{\beta\lambda}, \frac{1}{T}\boldsymbol{\Sigma})$. The likelihood function in the second step is then

$$\begin{aligned} p(data|\boldsymbol{\lambda}) &= (2\pi)^{-\frac{N}{2}}\left|\frac{1}{T}\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp\left\{-\frac{T}{2}(\bar{\boldsymbol{R}}-\boldsymbol{\beta\lambda})^\top\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{R}}-\boldsymbol{\beta\lambda})\right\} \\ &\propto \exp\left\{-\frac{T}{2}(\boldsymbol{\lambda}^\top\boldsymbol{\beta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta\lambda} - 2\boldsymbol{\lambda}^\top\boldsymbol{\beta}^\top\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{R}})\right\} \end{aligned}$$

Hence, we can now define the following estimator.

Definition 3 (Bayesian Fama-MacBeth GLS2 (BFM-GLS2)) *The BFM-GLS2 posterior distribution of* $\boldsymbol{\lambda}$ *is*

$$\boldsymbol{\lambda}|data,\boldsymbol{B},\boldsymbol{\Sigma} \sim \mathcal{N}(\hat{\boldsymbol{\lambda}}, \boldsymbol{\Sigma_\lambda}) \tag{18}$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{R}}$*,* $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} = \frac{1}{T}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}$*, and where* $\boldsymbol{\beta}$ *and* $\boldsymbol{\Sigma}$ *are drawn from the Normal-inverse-Wishart in (8)-(9).*

In the above, the conditional expectation $\hat{\boldsymbol{\lambda}} = (\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{R}}$ is essentially the Fama-MacBeth GLS estimate of $\boldsymbol{\lambda}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}$ is just the standard covariance matrix of the GLS estimates. Different from our OLS version, we incorporate the uncertainty of $\bar{\boldsymbol{R}}$ by drawing $\boldsymbol{\lambda}$ from the normal distribution with variance matrix $\frac{1}{T}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}$.

In this case, two forces are at play to make useless factors detectable in finite sample. First, as for the BFM and BFM-GLS cases, useless factors will generate posterior draws with diverging $\widehat{\boldsymbol{\lambda}}$ and flipping sing. Second, differently from our OLS version, we incorporate the uncertainty of $\bar{\boldsymbol{R}}$ by drawing $\boldsymbol{\lambda}$ from the normal distribution with variance matrix $\frac{1}{T}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}$.

A.3 Formal derivation of the flat prior pitfall for risk premia

Following the derivation in section A.1, the likelihood function in the second step is

$$p(data|\boldsymbol{\gamma}, \boldsymbol{\lambda}, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{a} - \boldsymbol{\beta}_\gamma \boldsymbol{\lambda}_\gamma)^\top (\boldsymbol{a} - \boldsymbol{\beta}_\gamma \boldsymbol{\lambda}_\gamma)\right\} \tag{19}$$

Assigning a Jeffreys' prior to the parameters[21] $(\boldsymbol{\lambda}, \sigma^2)$, the marginal likelihood function conditional on model index $\boldsymbol{\gamma}$ is

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

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

A.4 Proof of Remark 2

Proof. Consider two nested linear factor models, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$. The only difference between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ is γ_p: γ_p equals 1 in model $\boldsymbol{\gamma}$ but 0 in model $\boldsymbol{\gamma}'$. Let $\boldsymbol{\gamma}_{-p}$ denote a $(K-1) \times 1$ vector of model index excluding γ_p: $\boldsymbol{\gamma}^\top = (\boldsymbol{\gamma}_{-p}^\top, 1)$ and $\boldsymbol{\gamma}'^\top = (\boldsymbol{\gamma}_{-p}^\top, 0)$. Suppose that we rearrange the ordering of

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

factors such that factor p is the last one. To begin with, we introduce some matrix notations:

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

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

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

Equipped with the above, we have by direct calculation

$$\frac{p(data|\gamma_j = 1, \boldsymbol{\gamma_{-j}})}{p(data|\gamma_j = 0, \boldsymbol{\gamma_{-j}})} = \frac{\frac{|\boldsymbol{D_\gamma}|^{\frac{1}{2}}}{|\boldsymbol{\beta_\gamma}^\top \boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma}|^{\frac{1}{2}}} \frac{1}{\left(\frac{SSR_\gamma}{2}\right)^{\frac{N}{2}}}}{\frac{|\boldsymbol{D_{\gamma'}}|^{\frac{1}{2}}}{|\boldsymbol{\beta_{\gamma'}}^\top \boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}|^{\frac{1}{2}}} \frac{1}{\left(\frac{SSR_{\gamma'}}{2}\right)^{\frac{N}{2}}}} = \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \left(\frac{|\boldsymbol{D_\gamma}|}{|\boldsymbol{D_{\gamma'}}|}\right)^{\frac{1}{2}} \left(\frac{|\boldsymbol{\beta_{\gamma'}}^\top \boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}|}{|\boldsymbol{\beta_\gamma}^\top \boldsymbol{\beta_\gamma} + \boldsymbol{D_\gamma}|}\right)^{\frac{1}{2}}$$

$$= \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \psi_p^{-\frac{1}{2}} \left|\boldsymbol{\beta_p}^\top \boldsymbol{\beta_p} + \frac{1}{\psi_p} - \boldsymbol{\beta_p}^\top \boldsymbol{\beta_{\gamma'}} \left(\boldsymbol{\beta_{\gamma'}}^\top \boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}\right)^{-1} \boldsymbol{\beta_{\gamma'}}^\top \boldsymbol{\beta_p}\right|^{-\frac{1}{2}}$$

$$= \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \left(1 + \psi_p \boldsymbol{\beta_p}^\top \left[\boldsymbol{I_N} - \boldsymbol{\beta_{\gamma'}} \left(\boldsymbol{\beta_{\gamma'}}^\top \boldsymbol{\beta_{\gamma'}} + \boldsymbol{D_{\gamma'}}\right)^{-1} \boldsymbol{\beta_{\gamma'}}^\top\right] \boldsymbol{\beta_p}\right)^{-\frac{1}{2}}$$

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

A.5 Confidence Intervals for R^2 in Fama-MacBeth Estimation

In the spirit of Stock (1991) and Lewellen, Nagel, and Shanken (2010), for each of the simulations we use the following approach to construct the confidence interval for cross-sectional R^2, produced by the standard Fama-MacBeth estimation.

First, we choose a hypothetical true cross-sectional R^2, denoted by R_h^2. The expected test asset returns, $\mathbb{E}[\boldsymbol{R_t}]$, are assumed to follow $\mathbb{E}[\boldsymbol{R_t}] = h\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}} + \boldsymbol{\alpha}$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\lambda}}$ are sample estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ from historical data, and $\alpha_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2)$. The constants h and σ_e^2 are chosen to match the hypothetical cross-sectional R^2.

Let $\bar{\boldsymbol{R}}$ denote the vector of historical average asset returns and $Var_N(\boldsymbol{x})$ denote the sample cross-sectional variance of vector $\boldsymbol{x}$. The population cross-sectional R^2 is therefore

$$R_h^2 = \frac{h^2 Var_N(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}})}{Var_N(\mathbb{E}[\boldsymbol{R_t}])} = \frac{h^2 Var_N(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}})}{h^2 Var_N(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}}) + \sigma_e^2}.$$

At the same time, we'd like to maintain the historical cross-sectional dispersion of test asset returns, so we further have the following equation:

$$Var_N(\mathbb{E}[\boldsymbol{R_t}]) = h^2 Var_N(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}}) + \sigma_e^2 = Var_N(\mathbb{E}_T[\boldsymbol{R_t}]).$$

$$h^2 = \frac{R_h^2 Var_N(\mathbb{E}_T[\boldsymbol{R_t}])}{Var_N(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}})},$$

$$\sigma_e^2 = (1 - R_h^2) Var_N(\mathbb{E}_T[\boldsymbol{R_t}]).$$

Solving the system for h and σ_e^2, we simulate a vector of pricing errors from a normal distribution, $\alpha_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2)$. Since the sample variance of the draws α_i is generally different from σ_e^2, with a non-zero cross-sectional covariance with $\hat{\boldsymbol{\beta}}$, we adjust the vector $\boldsymbol{\alpha}$ by: (1) subtracting $\frac{Cov_N(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}}, \boldsymbol{\alpha})}{Var_N(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}})}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}}$ from $\boldsymbol{\alpha}$ such as the sample covariance between $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}}$ is zero; (2) multiplying $\boldsymbol{\alpha}$ by $\frac{\sigma_e}{\sigma_N(\boldsymbol{\alpha})}$ in order that sample cross-sectional variance of $\boldsymbol{\alpha}$ equals σ_e^2.

Second, we simulate a random sample of both factors and test asset returns. Factors are drawn with replacement from their empirical sample, while test assets returns $\boldsymbol{R_t}$ are assumed to follow a parametric normal distribution, that is,

$$\boldsymbol{R_t}|\boldsymbol{f_t} \overset{\text{iid}}{\sim} \mathcal{N}(h\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}} + \boldsymbol{\alpha} + \hat{\boldsymbol{\beta}}\boldsymbol{f_t}, \hat{\boldsymbol{\Sigma}})^{22}.$$

We then use Fama-MacBeth two-step approach to estimate R^2 for every simulated sample $\{\boldsymbol{f_t}, \boldsymbol{R_t}\}_{t=1}^T$. The second step is repeated for 1,000 times. For each hypothetical R^2 between 0 and 1, we find the (5%, 95%) confidence interval in the simulation. Finally, build a 90% confidence interval for $\hat{R}^2$ by including those values of hypothetically true R^2, whose 90% confidence interval contains $\hat{R}^2$.

The confidence intervals for GLS R^2 can be found in a similar way, with the only difference of focusing on cross-sectional R^2 for a linear combination of $\boldsymbol{R_t}$, i.e. $\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{R_t}$. Let $\tilde{\boldsymbol{R}}_{\boldsymbol{t}} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{R_t}$, $\tilde{\boldsymbol{\beta}} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\hat{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{t}} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\epsilon_t} \overset{\text{iid}}{\sim} \mathcal{N}(0, \boldsymbol{I_N})$. The simulation then relies on $\tilde{\boldsymbol{R}}_{\boldsymbol{t}}$ rather than $\boldsymbol{R_t}$,

$$\tilde{\boldsymbol{R}}_{\boldsymbol{t}}|\boldsymbol{f_t} \overset{\text{iid}}{\sim} \mathcal{N}(h\tilde{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}} + \boldsymbol{\alpha} + \tilde{\boldsymbol{\beta}}\boldsymbol{f_t}, \boldsymbol{I_N}).$$

A.6 Large N behavior

In this section we investigate the properties of the BFM procedure in estimating the risk premia and R-squared, as well successfully identifying irrelevant factors in the model, when applied to a large cross-section. For the sake of brevity, here we present the overview of the simulation results, with Tables C4 - C9 summarizing the output.

We consider the same simulation design as described at the beginning of Section III, except for the choice of the cross-section of test assets which time series and cross-sectional features we mimic. In our baseline case in the previous subsections we built a cross-section to emulate the 25 Fama-French portfolios, sorted by size and value. Now instead we consider the properties of the following composite cross-sections to simulate returns:

(a) $N = 55$: 25 Fama-French portfolios, sorted by size and value and 30 industry portfolios;

(b) $N = 100$: 25 Fama-French portfolios, sorted by size and value, 30 industry portfolios, 25 profitability and investment portfolios, 10 momentum portfolios, and 10 long-term reversal

[22] $\hat{\boldsymbol{\Sigma}}$ is the sample estimate of covariance matrix of random errors in time-series regression in equation (1).

portfolios.

The rest of the simulation design stays unchanged, i.e. the strong factor mimics the behavior of HML, with its betas and risk premia corresponding to their in-sample values, cross-sectional R^2_{adj}, as well as portfolio average returns, and variance of the residuals.

Tables C4 and C7 focus on the case of including only a strong factor into the model that is inherently misspecified and estimated on a large cross-section ($N = 55$ and $N = 100$, respectively). Risk premia estimates, recovered by BFM, are centered around the pseudo-true values and confidence intervals, produced by the posterior coverage, have the correct size. Confidence intervals for R^2_{adj} are equally centered around the true values, and overall become quite tight for a large sample size (e.g. $T \geq 600$). The GLS version of the cross-sectional fit is slightly lower than than the true value, and this is largely due to the estimation errors in the weight matrix, that are particularly important for a large cross-section, but overall are very close to the true values.

Tables C5 and C8 focus on the model with just a useless factor. As expected, the standard Fama-MacBeth estimation yields confidence intervals for the risk premia with wrong size, rejecting the zero risk premia for a useless factor with increasing frequency as T becomes large. In contrast, the BFM inference remains valid, with its empirical rejection rates being close to the true size of the tests.

Finally, Tables C6 and C9 consider the mixed case of estimating a model that includes both strong and useless factors. Again, BFM correctly identifies the presence of a spurious factor in the specification, and if anything, becomes somewhat more conservative, underrejecting the zero risk premia associated with it. This conservative inference is also shared by the estimates of the strong factor risk premia, with the latter being particularly evident in a large sample estimation ($T = 20,000$, $N = 100$). Again, the reason for this additional estimation uncertainty seems to lie in the large N behavior of the cross-sectional regression (betas and the weight matrix in case of the GLS). The posterior of a cross-sectional measure of fit remains relatively tight around the true value.

B Data

CAPM. Following Sharpe (1964) and Lintner (1965), the only risk factor is the excess return on market portfolio, which is proxied by a value-weighted portfolio of all CRSP firms incorporated in US and listed on the NYSE, AMEX or NASDAQ. We use 1-month Treasury rate as a proxy for risk-free rate. The data comes from Ken French's website.

Fama-French 3 factor model. Fama and French (1992) extend CAPM by introducing two additional factors, SMB and HML, where SMB is the return difference between portfolios of stocks with small and large market capitalizations, and HML is the return difference between portfolios of stocks with high and low book-to-market ratio. Again, the data comes from Ken French's website.

Carhart (1997). This paper extends Fama-French 3 factor model by including a momentum factor, UMD (also available from Ken French website).

q-factor model. Hou, Xue, and Zhang (2015) introduce a four-factor model that includes market excess return, a new size factor (ME), proxied by the return difference between large and small stocks, an investment factor (I/A), proxied by the return difference of stocks with high and low investment-to-asset ratio, and finally the profitability factor (ROE), created by sorting stocks based on their return-on-equity ratio. We receive the data from the authors. An alternative is Fama-French five factor model, but we only show the result of the first one since factors in these two papers are extremely similar.

Liquidity. Pástor and Stambaugh (2003) created a liquidity factor based on the fact that order flows result in larger return reversals when liquidity is lower. We download the monthly tradable liquidity factor from Stambaugh's website.

Quality-minus-junk. Asness, Frazzini, and Pedersen (2019) introduced the QMJ factor, and demonstrated that profitable, growing, well-managed companies, referred to as 'quality' firms, command a higher rate of return. The factor is available from AQR data library.

CCAPM. Real growth rate in nondurable consumption per capita (quarterly data) is computed from the data on consumption levels, available at St Louis FRED.

Scaled CAPM. Lettau and Ludvigson (2001) considered a conditional SDF $m_{t+1} = a_t - b_t MKT_{t+1}$, where a_t and b_t are linear function of conditional information cay_t. We download cay_t from the authors' website.

Scaled CCAPM. Similar to conditional CAPM, but the SDF includes nondurable consumption growth instead of the market factor.

HC-CCAPM. Jagannathan and Wang (1996) added a labor factor to CAPM. Following their paper, we compute the returns on human capital as:

$$R_t^{labor} = \frac{L_{t-1} + L_{t-2}}{L_{t-2} + L_{t-3}} - 1 \tag{20}$$

where L_t is the disposable labor income per capita.

Scaled HC-CCAPM. Lettau and Ludvigson (2001) extend Jagannathan and Wang (1996) by considering a conditional SDF in which cay is the only conditional information.

Durable consumption model. Yogo (2006) emphasized the role of durable consumption goods in explaining high returns of small stocks and value stocks. We consider a three-factor model: market excess return, non-durable consumption growth and durable (real) consumption growth C_d (seasonally adjusted at annual rates). The data is from Yogo's website, 1952Q1 to 2001Q4.

B.1 Factor list

Table B1: List of the factors for cross-sectional asset pricing models

Factor ID	Factor name and description	Reference	Source/construction
MKT	Market excess return	Sharpe (1964), Lintner (1965)	Ken French website
SMB	Size factor, constructed as a long-short portfolio of stocks sorted by their market cap (small-minus-big)	Fama and French (1992)	Ken French website
HML	Value factor, constructed as a long-short portfolio of stocks sorted by their book-to-market ratio (high-minus-low)	Fama and French (1992)	Ken French website
RMW	Profitability factor, constructed as a long-short portfolio of stocks sorted by their profitability (robust-minus-weak)	Fama and French (2015)	Ken French website
CMA	Investment factor, constructed as a long-short portfolio of stocks sorted by their investment activity (conservative-minus-aggressive)	Fama and French (2015)	Ken French website
UMD	Momentum factor, constructed as a long-short portfolio of stocks sorted by their 12-2 cumulative previous return (up-minus-down),	Carhart (1997), Jegadeesh and Titman (1993)	Ken French website
STREV	Short-term reversal factor, constructed as a long-short portfolio of stocks sorted by their previous month return	Jegadeesh and Titman (1993)	Ken French website
LTREV	Long-term reversal factor, constructed as a long-short portfolio of stocks sorted by their cumulative return accrued in the previous 60-13 months	Jegadeesh and Titman (2001)	Ken French website
q_IA	Investment factor, constructed as a long-short portfolio of stocks sorted by their investment-to-capital	Hou, Xue, and Zhang (2015)	Lu Zhang
q_ROE	Profitability factor, constructed as a long-short portfolio of stocks sorted by their return on equity	Hou, Xue, and Zhang (2015)	Lu Zhang
LIQ_NT	Liquidity factor, computed as the average of individual-stock measures estimated with daily data (residual predictability, controlling for the market factor)	Pástor and Stambaugh (2003)	Robert Stambaugh website
LIQ_TR	Liquidity factor, constructed as a long-short portfolio of stocks sorted by their exposure to LIQ_NT	Pástor and Stambaugh (2003)	Robert Stambaugh website
MGMT	Mispricing factor, constructed as a combination of anomalies, related to firm's management practices (stock issue, accruals, asset growth, etc)	Stambaugh and Yuan (2016)	Robert Stambaugh website
PERF	Mispricing factor, constructed as a combination of anomalies, related to firm's performance (profitability, distress, return on assets, etc)	Stambaugh and Yuan (2016)	Robert Stambaugh website
ACCR	Accruals factor, constructed as a long-short portfolio of stocks sorted by changes in operating working capital per split-adjusted share from the fiscal year end t-2 to t-1, divided by book equity per share in t-1	Sloan (1996)	Robert Stambaugh website

DISSTR	Distress factor, constructed as a long-short portfolio of stocks sorted by the predicted failure probability	Campbell, Hilscher, and Szilagyi (2008)	Robert Stambaugh website
ASSET_Growth	Asset growth factor, constructed as a long-short portfolio of stocks sorted by growth rate of total assets in the previous fiscal year	Cooper, Gulen, and Schill (2008)	Robert Stambaugh website
COMP_ISSUE	Composite issue factor, constructed as a long-short portfolio of stocks sorted by the growth in the firm's total market value of equity above that of the stock's rate of return	Daniel and Titman (2006)	Robert Stambaugh website
GR_PROF	Gross profitability factor, constructed as a long-short portfolio of stocks sorted by the ratio of gross profit to assets creates abnormal benchmark-adjusted returns	Novy-Marx (2013)	Robert Stambaugh website
INV_IN_ASSETS	Investment-in-assets factor, constructed as a long-short portfolio of stocks sorted by the annual change in gross property, plant, and equipment, plus the annual change in inventories, scaled by lagged book value of the assets	Titman, Wei, and Xie (2004)	Robert Stambaugh website
NetOA	Net operating assets factor, constructed as a long-short portfolio of stocks sorted by net operating assets	Hirshleifer, Kewei, Teoh, and Zhang (2004)	Robert Stambaugh website
OSCORE	Ohlson O-score factor, constructed as a long-short portfolio of stocks sorted by the predicted value of a distress measure	Ohlson (1980)	Robert Stambaugh website
ROA	Return-on-assets factor, constructed as a long-short portfolio of stocks sorted by their return on assets	Chen, Novy-Marx, and Zhang (2010)	Robert Stambaugh website
STOCK_ISS	Stock issuance factor, constructed as a long-short portfolio of stocks sorted by their annual log change in split-adjusted shares outstanding	Ritter (1991), Fama and French (2008)	Robert Stambaugh website
INTERM_CR	Innovations to the intermediaries' capital (equity) ratio	He, Kelly, and Manela (2017)	Asaf Manela website
BAB	Betting-against-beta factor, constructed as a portfolio that holds low-beta assets, leveraged to a beta of 1, and that shorts high-beta assets, de-leveraged to a beta of 1	Frazzini and Pedersen (2014)	AQR data library
HML_DEVIL	A version of the HML factor that relies on the current price level to sort the stocks into long and short legs	Asness and Frazzini (2013)	AQR data library
QMJ	Auality-minus-junk factor, constructed as a long-short portfolio of stocks sorted by the combination of their safety, profitability, growth, and the quality of management practices	Asness, Frazzini, and Pedersen (2019)	AQR data library
FIN_UNC	A measure of financial uncertainty	Jurado, Ludvigson, and Ng (2015), Ludvigson, Ma, and Ng (2019)	Sydney Ludvigson website
REAL_UNC	A measure of real economic uncertainty	Jurado, Ludvigson, and Ng (2015), Ludvigson, Ma, and Ng (2019)	Sydney Ludvigson website

MACRO_UNC	A measure of macroeconomic uncertainty	Jurado, Ludvigson, and Ng (2015), Ludvigson, Ma, and Ng (2019)	Sydney Ludvigson website
TERM	Term spread, measured as the difference in 10 year Treasury bonds and fed funds rate	Chen, Ross, and Roll (1986), Fama and French (1993)	FRED-MD database
DELTA_SLOPE	Change in the difference between a 10-year Treasury bond yield and a 3-month Treasury bill yield	Ferson and Harvey (1991)	FRED-MD database
CREDIT	Credit spread, measured as the difference between Moody's BAA corporate bond yields and 10-year Treasury bonds	Chen, Ross, and Roll (1986), Fama and French (1993)	FRED-MD database
DIV	Dividend yield	Campbell (1996)	FRED-MD database
PE	Price-earnings ratio	Basu (1977), Ball (1985)	FRED-MD database
BW_INV_SENT	Investor sentiment measure, aggregated from a set of indices, orthogonal to macroeconomic fundamentals	Baker and Wurgler (2006)	Dashan Huang website
HJTZ_INV_SENT	Investor sentiment measure, extracted with PLS from Baker and Wurgler (2006) proxies	Huang, Jiang, Tu, and Zhou (2015)	Dashan Huang website
BEH_PEAD	Short-term behavioral factor, reflecting post-earnings announcement drift	Daniel, Hirshleifer, and Sun (2019)	Kent Daniel website
BEH_FIN	Long-term behavioral factor, predominantly capturing the impact of share issuance and correction	Daniel, Hirshleifer, and Sun (2019)	Kent Daniel website
MKT*	Market factor with a hedged unpriced component	Daniel, Mota, Rottke, and Santos (2018)	Kent Daniel website
SMB*	SMB with a hedged unpriced component	Daniel, Mota, Rottke, and Santos (2018)	Kent Daniel website
HML*	HML with a hedged unpriced component	Daniel, Mota, Rottke, and Santos (2018)	Kent Daniel website
RMW*	RMW with a hedged unpriced component	Daniel, Mota, Rottke, and Santos (2018)	Kent Daniel website
CMA*	CMA with a hedged unpriced component	Daniel, Mota, Rottke, and Santos (2018)	Kent Daniel website
SKEW	Systematic skewness factor, constructed as a long-short portfolio of stocks sorted on the their predicted systematic skewness rank	Langlois (2019)	Hugues Langlois website
NONDUR	Nondurable consumption growth (real, chain-weighted, per capita)	Chen, Ross, and Roll (1986), Breeden, Gibbons, and Litzenberger (1989)	Monthly consumption expenditure, the chain-weighted price index, and population data are from BEA
SERV	Growth rate (real, chain-weighted, per capita) service expenditure	Breeden, Gibbons, and Litzenberger (1989), Hall (1978)	Monthly expenditure for services, the chain-weighted price index, and population data are from BEA
UNRATE	Unemployment rate	Gertler and Grinols (1982)	FRED-MD database
IND_PROD	Growth rate of industrial production	Chan, Chen, and Hsieh (1985), Chen, Ross, and Roll (1986)	Industrial Production Index is from the Board of Governors of the Federal Reserve System

OIL	Monthly growth rate of the Producer Price index for Crude Petroleum (domestic production)	Chen, Ross, and Roll (1986)	PPI is from U.S. Bureau of Labor Statistics

The table presents the list of factors used in Section IV.2. For each of the variables we present their identification index, the nature of the factor, and the source of data for downloading and/or constructing the time series.

C Additional Tables

Table C1: Tests of risk premia in a correctly specified model with a strong factor

		$\lambda_{intercept}$			λ_{HML}			R^2_{adj}	
	T	10%	5%	1%	10%	5%	1%	5th	95th
				Panel A. OLS					
FM	100	0.115	0.049	0.017	0.114	0.050	0.014	-4.17%	44.29%
	200	0.101	0.055	0.012	0.096	0.047	0.010	0.96%	67.73%
	600	0.106	0.053	0.011	0.100	0.048	0.011	30.52%	84.56%
	1000	0.096	0.050	0.011	0.102	0.044	0.010	47.86%	90.18%
	20000	0.102	0.042	0.012	0.100	0.049	0.007	96.20%	99.43%
BFM	100	0.063	0.027	0.006	0.034	0.010	0.002	-3.37%	35.16%
	200	0.107	0.055	0.012	0.080	0.037	0.005	-2.99%	49.06%
	600	0.095	0.050	0.007	0.080	0.035	0.007	4.31%	69.13%
	1000	0.092	0.054	0.008	0.092	0.047	0.012	32.91%	78.18%
	20000	0.108	0.054	0.012	0.110	0.052	0.008	96.54%	98.49%
				Panel B. GLS					
FM	100	0.221	0.156	0.061	0.212	0.137	0.064	10.62%	73.80%
	200	0.153	0.088	0.024	0.144	0.088	0.024	59.23%	85.71%
	600	0.117	0.062	0.017	0.115	0.060	0.014	83.38%	94.46%
	1000	0.106	0.053	0.011	0.112	0.052	0.011	90.03%	96.49%
	20000	0.100	0.058	0.010	0.103	0.047	0.011	99.47%	99.81%
BFM	100	0.151	0.088	0.026	0.149	0.086	0.026	26.42%	65.69%
	200	0.123	0.066	0.016	0.124	0.066	0.015	49.07%	73.52%
	600	0.106	0.055	0.012	0.105	0.056	0.011	76.40%	87.30%
	1000	0.107	0.054	0.011	0.103	0.051	0.011	85.12%	91.54%
	20000	0.098	0.057	0.009	0.100	0.051	0.013	99.15%	99.50%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_i^* in a correctly specified model with an intercept and a strong factor. Hypothetical true value of R^2_{adj} is 100%. Fama-MacBeth estimates are constructed using OLS (GLS) two-step cross-sectional regressions, with standard errors including Shanken correction. Confidence intervals for BFM estimates are constructed using a posterior distribution of Fama-MacBeth estimates of λ. The last two columns report the 5th and 95th percentiles of cross-sectional R^2_{adj} across 1000 simulations, evaluated at the simulation point estimates for FM, and its posterior mode for BFM.

Table C2: Tests of risk premia in a correctly specified model with a useless factor

	T	$\lambda_{intercept}$ 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
Panel A. OLS									
FM	100	0.087	0.033	0.008	0.019	0.003	0.000	-4.19%	48.44%
	200	0.071	0.034	0.006	0.052	0.011	0.000	-4.20%	56.84%
	600	0.077	0.031	0.005	0.131	0.037	0.000	-4.22%	58.50%
	1000	0.071	0.035	0.006	0.205	0.066	0.002	-4.16%	61.28%
	20000	0.133	0.077	0.027	0.709	0.479	0.140	-4.11%	70.07%
BFM	100	0.039	0.011	0.002	0.001	0.001	0.000	-2.29%	-0.12%
	200	0.038	0.013	0.001	0.002	0.000	0.000	-2.32%	0.30%
	600	0.048	0.012	0.002	0.017	0.006	0.001	-2.21%	1.26%
	1000	0.048	0.011	0.000	0.031	0.010	0.000	-2.14%	1.60%
	20000	0.007	0.000	0.000	0.103	0.051	0.014	-1.76%	57.93%
Panel B. GLS									
FM	100	0.215	0.130	0.052	0.211	0.117	0.033	-3.10%	37.27%
	200	0.141	0.080	0.023	0.139	0.072	0.013	-3.73%	22.82%
	600	0.108	0.053	0.014	0.142	0.065	0.010	-3.77%	21.16%
	1000	0.097	0.053	0.009	0.171	0.082	0.012	-3.71%	20.92%
	20000	0.108	0.047	0.010	0.621	0.559	0.372	-2.59%	16.97%
BFM	100	0.136	0.074	0.022	0.029	0.011	0.001	-1.94%	10.60%
	200	0.112	0.060	0.014	0.012	0.004	0.000	-2.39%	8.37%
	600	0.097	0.047	0.010	0.010	0.002	0.000	-2.65%	7.49%
	1000	0.096	0.047	0.009	0.010	0.002	0.000	-2.76%	8.32%
	20000	0.071	0.034	0.004	0.077	0.033	0.004	-3.26%	7.66%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true value of λ_c and $\lambda^*_{useless} = 0$ in a correctly specified model with an intercept and a useless factor. The true value of R^2 is 0%. Fama-MacBeth estimates are constructed using OLS (GLS) two-step cross-sectional regressions, with standard errors including Shanken correction. Confidence intervals for BFM estimates are constructed using a posterior distribution of Fama-MacBeth estimates of λ. The last two columns report the 5th and 95th percentiles of cross-sectional R^2_{adj} across 1000 simulations, evaluated at the simulation point estimates for FM, and its posterior mode for BFM.

Table C3: Tests of risk premia in a correctly specified model with useless and strong factors

	T	$\lambda_{intercept}$ 10%	5%	1%	λ_{HML} 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
Panel A. OLS												
FM	100	0.073	0.038	0.006	0.071	0.033	0.006	0.043	0.012	0.000	-4.45%	59.30%
	200	0.073	0.035	0.006	0.090	0.045	0.008	0.028	0.006	0.000	10.83%	74.84%
	600	0.076	0.033	0.005	0.100	0.046	0.009	0.027	0.005	0.000	38.73%	85.72%
	1000	0.073	0.034	0.006	0.090	0.046	0.009	0.026	0.005	0.000	54.15%	91.13%
	20000	0.087	0.032	0.006	0.061	0.026	0.005	0.025	0.008	0.000	96.87%	99.47%
BFM	100	0.039	0.013	0.001	0.023	0.007	0.001	0.002	0.001	0.000	-1.97%	41.74%
	200	0.048	0.023	0.000	0.046	0.015	0.000	0.002	0.001	0.000	0.03%	53.72%
	600	0.054	0.025	0.003	0.062	0.026	0.006	0.002	0.000	0.000	40.56%	72.82%
	1000	0.084	0.034	0.006	0.064	0.021	0.002	0.002	0.000	0.000	57.63%	80.89%
	20000	0.071	0.033	0.005	0.069	0.032	0.004	0.004	0.000	0.000	97.04%	98.60%
Panel B. GLS												
FM	100	0.193	0.129	0.043	0.205	0.133	0.043	0.180	0.121	0.031	14.05%	74.80%
	200	0.135	0.080	0.021	0.141	0.075	0.024	0.129	0.062	0.010	60.15%	86.83%
	600	0.101	0.054	0.010	0.109	0.052	0.009	0.091	0.037	0.004	83.31%	94.23%
	1000	0.104	0.055	0.010	0.105	0.048	0.011	0.085	0.039	0.003	89.95%	96.32%
	20000	0.095	0.047	0.006	0.101	0.052	0.005	0.088	0.035	0.004	99.44%	99.81%
BFM	100	0.133	0.074	0.023	0.132	0.074	0.023	0.026	0.009	0.001	27.96%	66.80%
	200	0.106	0.054	0.012	0.106	0.053	0.012	0.013	0.004	0.000	48.99%	73.62%
	600	0.094	0.047	0.009	0.093	0.046	0.009	0.007	0.001	0.000	76.54%	87.35%
	1000	0.090	0.045	0.010	0.091	0.044	0.007	0.005	0.001	0.000	85.30%	91.46%
	20000	0.092	0.043	0.008	0.092	0.046	0.008	0.004	0.001	0.000	99.14%	99.51%

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 100%. Fama-MacBeth estimates are constructed using OLS (GLS) two-step cross-sectional regressions, with standard errors including Shanken correction. Confidence intervals for BFM estimates are constructed using a posterior distribution of Fama-MacBeth estimates of λ. The last two columns report the 5th and 95th percentiles of cross-sectional R^2_{adj} across 1000 simulations, evaluated at the simulation point estimates for FM, and its posterior mode for BFM.

Table C4: Tests of risk premia in a misspecified model with a strong factor ($N = 55$)

	T	λ_c 10%	λ_c 5%	λ_c 1%	λ_{strong} 10%	λ_{strong} 5%	λ_{strong} 1%	R^2_{adj} 5th	R^2_{adj} 95th
					Panel A: OLS				
FM	100	0.107	0.058	0.014	0.115	0.059	0.012	-1.86%	13.05%
	200	0.091	0.046	0.009	0.102	0.052	0.011	-1.84%	14.85%
	600	0.104	0.052	0.013	0.109	0.060	0.014	-1.66%	14.95%
	1,000	0.100	0.056	0.009	0.123	0.061	0.013	-1.23%	13.51%
	20,000	0.104	0.049	0.009	0.109	0.048	0.009	3.53%	8.03%
BFM	100	0.017	0.004	0.000	0.002	0.000	0.000	-1.55%	-0.67%
	200	0.046	0.017	0.004	0.023	0.006	0.000	-1.68%	2.73%
	600	0.089	0.042	0.012	0.071	0.036	0.005	-1.74%	8.44%
	1,000	0.093	0.047	0.007	0.089	0.040	0.007	-1.71%	9.59%
	20,000	0.101	0.048	0.011	0.099	0.042	0.008	3.29%	7.77%
					Panel B: GLS				
FM	100	0.509	0.428	0.305	0.513	0.431	0.305	20.93%	64.71%
	200	0.295	0.206	0.102	0.274	0.187	0.091	39.99%	66.57%
	600	0.170	0.109	0.042	0.176	0.113	0.034	55.85%	71.32%
	1,000	0.170	0.106	0.034	0.176	0.105	0.031	60.10%	72.32%
	20,000	0.145	0.082	0.020	0.141	0.073	0.022	69.17%	71.82%
BFM	100	0.265	0.181	0.086	0.262	0.186	0.079	18.72%	59.19%
	200	0.163	0.100	0.032	0.147	0.086	0.024	34.01%	58.42%
	600	0.116	0.060	0.017	0.118	0.058	0.014	51.25%	65.95%
	1,000	0.120	0.064	0.016	0.121	0.063	0.014	56.89%	68.65%
	20,000	0.106	0.053	0.009	0.095	0.048	0.013	68.97%	71.63%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_i^* in a misspecified model with an intercept and a strong factor, estimates of a cross-section of 55 test assets. The true value of the cross-sectional R^2_{adj} is 5.72% (70.71%) in OLS (GLS) estimation. Fama-MacBeth estimates are constructed using OLS (GLS) two-step cross-sectional regressions, with standard errors including Shanken correction. Confidence intervals and their size for BFM estimates are constructed using posterior coverage of Fama-MacBeth estimates of λ. The last two columns report the 5th and 95th percentiles of cross-sectional R^2_{adj} across 1000 simulations, evaluated at the simulation point estimates for FM, and its posterior mode for BFM. The test assets mimic the time series and cross-sectional properties of 25 Fama-French size-value portfolios, and 30 industry portfolios, while the strong factor proxies the HML factor (all the data available from Ken French website).

Table C5: Tests of risk premia in a misspecified model with a useless factor ($N = 55$)

		λ_c			$\lambda_{useless}$			R^2_{adj}	
	T	10%	5%	1%	10%	5%	1%	5th	95th
					Panel A: OLS				
FM	100	0.095	0.050	0.013	0.084	0.031	0.003	-1.86%	20.80%
	200	0.084	0.042	0.007	0.093	0.038	0.004	-1.86%	18.01%
	600	0.103	0.051	0.011	0.155	0.077	0.011	-1.87%	13.62%
	1000	0.101	0.051	0.009	0.226	0.131	0.033	-1.87%	14.17%
	20000	0.191	0.126	0.050	0.716	0.650	0.460	-1.87%	9.88%
BFM	100	0.013	0.002	0.000	0.000	0.000	0.000	-1.05%	-0.47%
	200	0.039	0.015	0.001	0.002	0.000	0.000	-1.28%	-0.43%
	600	0.076	0.040	0.006	0.004	0.000	0.000	-1.44%	-0.49%
	1000	0.082	0.035	0.004	0.022	0.009	0.000	-1.50%	-0.26%
	20000	0.129	0.059	0.005	0.078	0.037	0.008	-1.68%	-0.20%
					Panel B: GLS				
FM	100	0.492	0.411	0.287	0.540	0.468	0.302	-1.34%	23.18%
	200	0.267	0.196	0.088	0.364	0.274	0.153	-1.59%	13.59%
	600	0.166	0.101	0.035	0.408	0.323	0.198	-1.62%	10.09%
	1000	0.154	0.089	0.028	0.475	0.401	0.266	-1.55%	8.68%
	20000	0.155	0.100	0.044	0.806	0.772	0.705	-0.68%	6.68%
BFM	100	0.259	0.180	0.085	0.1695	0.105	0.041	-0.14%	12.85%
	200	0.162	0.094	0.030	0.065	0.027	0.005	-0.89%	6.15%
	600	0.108	0.058	0.013	0.056	0.022	0.003	-1.27%	5.31%
	1000	0.112	0.057	0.014	0.064	0.021	0.004	-1.38%	4.79%
	20000	0.051	0.020	0.001	0.093	0.049	0.011	-0.72%	1.72%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true value of λ_c and $\lambda_{useless}^* = 0$ in a misspecified model with an intercept and a useless factor, estimated on a cross-section of 55 portfolios. The true value of the cross-sectional R^2 is zero. The test assets mimic the time series and cross-sectional properties of 25 Fama-French size-value portfolios, and 30 industry portfolios.

Table C6: Tests of risk premia in a misspecified model with useless and strong factors ($N = 55$)

		λ_c			λ_{strong}			$\lambda_{useless}$			R^2_{adj}	
	T	10%	5%	1%	10%	5%	1%	10%	5%	1%	5th	95th
						Panel A: OLS						
FM	100	0.097	0.052	0.014	0.099	0.045	0.009	0.108	0.041	0.006	-3.18%	24.77%
	200	0.085	0.041	0.006	0.090	0.049	0.011	0.117	0.051	0.008	-2.98%	23.50%
	600	0.095	0.045	0.013	0.108	0.060	0.012	0.191	0.100	0.015	-2.12%	21.21%
	1000	0.083	0.043	0.006	0.125	0.072	0.016	0.258	0.171	0.047	-1.55%	21.37%
	20000	0.109	0.055	0.012	0.146	0.086	0.038	0.766	0.704	0.535	2.67%	17.78%
BFM	100	0.014	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-1.41%	1.61%
	200	0.037	0.014	0.001	0.017	0.005	0.000	0.002	0.000	0.000	-2.03%	7.26%
	600	0.069	0.031	0.005	0.058	0.027	0.002	0.007	0.000	0.000	-2.27%	10.96%
	1000	0.064	0.024	0.003	0.069	0.027	0.005	0.026	0.008	0.001	-2.09%	11.73%
	20000	0.028	0.006	0.000	0.068	0.033	0.004	0.080	0.042	0.009	2.62%	8.73%
						Panel B: GLS						
FM	100	0.488	0.406	0.282	0.495	0.411	0.281	0.537	0.465	0.306	22.01%	66.13%
	200	0.263	0.194	0.088	0.252	0.167	0.077	0.359	0.279	0.151	40.47%	67.07%
	600	0.157	0.094	0.032	0.161	0.093	0.027	0.398	0.313	0.199	55.81%	71.59%
	1000	0.146	0.087	0.029	0.144	0.090	0.026	0.475	0.393	0.251	59.94%	72.50%
	20000	0.140	0.094	0.035	0.134	0.089	0.041	0.789	0.747	0.667	68.88%	72.37%
BFM	100	0.253	0.182	0.081	0.260	0.176	0.077	0.168	0.104	0.039	20.36%	60.07%
	200	0.156	0.092	0.028	0.140	0.082	0.021	0.063	0.029	0.004	34.51%	58.44%
	600	0.105	0.058	0.013	0.108	0.049	0.011	0.054	0.024	0.002	51.25%	66.14%
	1000	0.105	0.054	0.015	0.111	0.056	0.009	0.057	0.024	0.003	56.78%	68.73%
	20000	0.051	0.019	0.001	0.045	0.018	0.001	0.093	0.045	0.013	68.73%	71.57%

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}, and $\lambda^*_{useless} \equiv 0$ in a misspecified model with an intercept, a strong, and a useless factor, estimated on a cross-section of 55 portfolios. The true value of the cross-sectional R^2_{adj} is 5.72% (70.71%) in OLS (GLS) estimation. The test assets mimic the time series and cross-sectional properties of 25 Fama-French size-value portfolios, and 30 industry portfolios, while the strong factor proxies the HML factor.

Table C7: Tests of risk premia in a misspecified model with a strong factor ($N = 100$)

		λ_c			λ_{strong}			R^2_{adj}	
	T	10%	5%	1%	10%	5%	1%	5th	95th
					Panel A: OLS				
FM	200	0.091	0.044	0.010	0.109	0.055	0.012	-0.98%	13.10%
	600	0.102	0.052	0.013	0.116	0.057	0.019	-0.42%	12.31%
	1000	0.099	0.056	0.011	0.117	0.060	0.014	0.31%	11.45%
	20000	0.099	0.044	0.010	0.116	0.068	0.017	4.31%	7.48%
BFM	200	0.016	0.004	0.000	0.006	0.001	0.000	-0.82%	0.74%
	600	0.076	0.034	0.006	0.060	0.028	0.005	-0.86%	7.90%
	1000	0.081	0.046	0.007	0.076	0.037	0.007	-0.72%	8.59%
	20000	0.097	0.044	0.011	0.106	0.057	0.014	4.18%	7.42%
					Panel B: GLS				
FM	200	0.495	0.411	0.287	0.481	0.403	0.268	29.38%	58.85%
	600	0.255	0.169	0.077	0.257	0.178	0.073	47.36%	62.09%
	1000	0.234	0.149	0.059	0.205	0.129	0.049	51.98%	63.52%
	20000	0.166	0.100	0.035	0.170	0.097	0.036	61.42%	63.98%
BFM	200	0.249	0.163	0.070	0.233	0.158	0.073	25.67%	52.96%
	600	0.132	0.082	0.023	0.137	0.077	0.020	42.87%	56.77%
	1000	0.125	0.073	0.021	0.112	0.060	0.014	48.49%	59.70%
	20000	0.101	0.056	0.012	0.098	0.053	0.013	61.14%	63.75%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for the pseudo-true values of λ_i^* in a misspecified model with an intercept and a strong factor, estimated on a cross-section of 100 portfolios. The true value of R^2_{adj} is 5.85% (62.97%) for the OLS (GLS) estimation. Fama-MacBeth estimates are constructed using OLS (GLS) two-step cross-sectional regressions, with standard errors including Shanken correction. Confidence intervals and their size for BFM estimates are constructed using a posterior distribution of Fama-MacBeth estimates of λ. The last two columns report the 5th and 95th percentiles of cross-sectional R^2_{adj} across 1000 simulations, evaluated at the simulation point estimates for FM, and its posterior mode for BFM. The simulations design follows the methodology described in Section III, with the test assets mimicking the composite cross-section of 25 Fama-French size-B/M portfolios, 30 industry portfolios, 25 profitability and investment portfolios, 10 momentum portfolios, as well as 10 long-term reversal portfolios (all available from Ken French website). A strong factor mimics the behavior of HML.

Table C8: Tests of risk premia in a misspecified model with a useless factor ($N = 100$)

	T	λ_c 10%	λ_c 5%	λ_c 1%	$\lambda_{useless}$ 10%	$\lambda_{useless}$ 5%	$\lambda_{useless}$ 1%	R^2_{adj} 5th	R^2_{adj} 95th
				Panel A: OLS					
FM	200	0.091	0.041	0.009	0.147	0.079	0.015	-1.01%	14.81%
	600	0.110	0.062	0.010	0.280	0.180	0.064	-1.00%	13.55%
	1000	0.096	0.053	0.007	0.341	0.248	0.101	-1.00%	12.22%
	20000	0.221	0.151	0.061	0.805	0.759	0.643	-1.00%	12.20%
BFM	200	0.014	0.003	0.000	0.000	0.000	0.000	-0.50%	0.02%
	600	0.070	0.030	0.003	0.014	0.002	0.000	-0.66%	0.26%
	1000	0.073	0.034	0.005	0.026	0.010	0.001	-0.71%	0.27%
	20000	0.097	0.035	0.003	0.091	0.045	0.008	-0.81%	1.09%
				Panel B: GLS					
FM	200	0.472	0.397	0.272	0.530	0.454	0.320	-0.72%	14.95%
	600	0.230	0.149	0.064	0.458	0.363	0.235	-0.52%	9.69%
	1000	0.196	0.122	0.048	0.487	0.407	0.275	-0.37%	8.61%
	20000	0.106	0.063	0.021	0.760	0.717	0.640	1.71%	6.15%
BFM	200	0.244	0.161	0.066	0.150	0.092	0.031	-0.16%	7.69%
	600	0.129	0.073	0.021	0.078	0.035	0.008	-0.55%	6.76%
	1000	0.118	0.066	0.019	0.070	0.033	0.006	-0.60%	6.24%
	20000	0.076	0.034	0.005	0.093	0.049	0.009	1.72%	3.99%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true value of λ_c and $\lambda^*_{useless} = 0$ in a misspecified model with an intercept and a useless factor, estimated on a cross-section of 100 portfolios. The true value of R^2 is zero. The test assets mimic the time series and cross-sectional properties of composite cross-section of 25 Fama-French size-B/M portfolios, 30 industry portfolios, 25 profitability and investment portfolios, 10 momentum portfolios, as well as 10 long-term reversal portfolios, while the strong factor mimics the behavior of HML (all the data is available from Ken French website).

Table C9: Tests of risk premia in a misspecified model with useless and strong factors ($N = 100$)

	T	λ_c 10%	λ_c 5%	λ_c 1%	λ_{strong} 10%	λ_{strong} 5%	λ_{strong} 1%	$\lambda_{useless}$ 10%	$\lambda_{useless}$ 5%	$\lambda_{useless}$ 1%	R^2_{adj} 5th	R^2_{adj} 95th
						Panel A: OLS						
FM	200	0.088	0.043	0.009	0.098	0.045	0.008	0.187	0.104	0.026	-1.24%	21.18%
	600	0.103	0.056	0.011	0.104	0.064	0.015	0.319	0.217	0.081	-0.20%	19.57%
	1000	0.105	0.049	0.009	0.115	0.058	0.013	0.389	0.284	0.134	0.70%	18.26%
	20000	0.102	0.057	0.014	0.140	0.075	0.028	0.823	0.782	0.684	4.00%	17.66%
BFM	200	0.012	0.003	0.000	0.005	0.000	0.000	0.000	0.000	0.000	-0.51%	5.41%
	600	0.062	0.025	0.003	0.046	0.021	0.002	0.016	0.004	0.000	-0.63%	10.77%
	1000	0.062	0.029	0.005	0.057	0.025	0.003	0.029	0.008	0.001	-0.05%	10.78%
	20000	0.026	0.008	0.001	0.042	0.015	0.000	0.089	0.039	0.011	3.99%	8.18%
						Panel B: GLS						
FM	200	0.467	0.392	0.268	0.471	0.383	0.254	0.523	0.454	0.315	29.71%	59.43%
	600	0.227	0.149	0.059	0.228	0.154	0.060	0.455	0.363	0.227	47.45%	62.22%
	1000	0.191	0.118	0.046	0.178	0.114	0.036	0.480	0.401	0.262	52.07%	63.68%
	20000	0.098	0.054	0.020	0.095	0.062	0.024	0.749	0.707	0.621	61.21%	64.16%
BFM	200	0.243	0.160	0.066	0.230	0.155	0.072	0.152	0.087	0.031	26.13%	53.50%
	600	0.126	0.072	0.022	0.132	0.071	0.017	0.074	0.033	0.007	42.68%	56.92%
	1000	0.117	0.067	0.017	0.109	0.057	0.013	0.068	0.034	0.006	48.64%	59.79%
	20000	0.068	0.032	0.002	0.066	0.034	0.006	0.087	0.047	0.009	61.02%	63.71%

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 on the cross-section of 100 portfolios. The true value of R^2_{adj} is 5.85% (62.97%) in OLS (GLS) estimation. The test assets mimic the time series and cross-sectional properties of composite cross-section of 25 Fama-French size-B/M portfolios, 30 industry portfolios, 25 profitability and investment portfolios, 10 momentum portfolios, as well as 10 long-term reversal portfolios, while the strong factor mimics the behavior of HML (all the data is available from Ken French website).

Table C10: The probability of retaining risk factors using BF

		T	55%	57%	59%	61%	63%	65%
			Panel A: strong factors					
Jeffreys Prior	f_{strong}	200	0.860	0.845	0.830	0.812	0.792	0.771
		600	0.987	0.985	0.985	0.983	0.981	0.979
		1000	0.998	0.998	0.997	0.996	0.996	0.995
Spike-and-Slab Prior	f_{strong}	200	0.749	0.718	0.685	0.654	0.618	0.586
		600	0.982	0.979	0.978	0.972	0.970	0.964
		1000	0.996	0.996	0.996	0.995	0.994	0.992
			Panel B: useless factors					
Jeffreys Prior	$f_{useless}$	200	1.000	0.995	0.982	0.940	0.862	0.726
		600	1.000	0.999	0.998	0.988	0.971	0.920
		1000	1.000	1.000	1.000	0.999	0.990	0.971
Spike-and-Slab Prior	$f_{useless}$	200	0.419	0.248	0.149	0.083	0.040	0.022
		600	0.119	0.062	0.028	0.012	0.007	0.004
		1000	0.083	0.028	0.011	0.001	0.000	0.000
			Panel C: strong and useless factors					
Jeffreys Prior	f_{strong}	200	0.928	0.912	0.891	0.878	0.860	0.838
		600	0.994	0.994	0.992	0.991	0.991	0.989
		1000	0.999	0.999	0.999	0.999	0.999	0.999
	$f_{useless}$	200	0.955	0.894	0.788	0.642	0.489	0.360
		600	0.957	0.895	0.764	0.618	0.461	0.354
		1000	0.957	0.893	0.787	0.645	0.483	0.357
Spike-and-Slab Prior	f_{strong}	200	0.753	0.725	0.697	0.666	0.629	0.592
		600	0.982	0.980	0.979	0.972	0.970	0.961
		1000	0.996	0.996	0.996	0.995	0.994	0.994
	$f_{useless}$	200	0.283	0.154	0.085	0.044	0.023	0.011
		600	0.077	0.031	0.006	0.002	0.000	0.000
		1000	0.053	0.008	0.000	0.000	0.000	0.000

The table shows the frequency of retaining risk factors for different choice sets across 1,000 simulations of different size (T=200, 600, and 1,000). In Panel A, the candidate risk factor is truly cross-sectionally priced and strongly identified, while in Panel B they are not. Panel C summarizes the case of using both strong and useless candidate factors in the model. A candidate factor is retained in the model, if its marginal posterior probability, $p(\gamma_i = 1|data)$, is greater than a certain threshold, i.e. 55%, 57%, 59%, 61%, 63% and 65%.

Table C11: Two-pass regressions with tradable factors: 25 Fama-French portfolios

Model	Factors	FM $\hat{\lambda}_j$	FM R^2_{adj}	BFM $\bar{\lambda}_j$	BFM $R^2_{adj,mode}$	BFM $R^2_{adj,median}$
Panel A: OLS						
CAPM	Intercept	1.421***	17.69	1.404***	-0.91	14.61
		[0.627, 2.215]	[-4.35, 66.61]	[0.595, 2.256]	[-4.03, 52.11]	
	MKT	-0.647		-0.631		
		[-1.429, 0.135]		[-1.458, 0.163]		
Fama and French (1992)	Intercept	1.273***	60.71	1.249***	56.04	55.66
		[0.730, 1.816]	[20.00, 85.14]	[0.682, 1.834]	[42.75, 69.46]	
	MKT	-0.686**		-0.664**		
		[-1.227, -0.145]		[-1.242, -0.102]		
	SMB	0.140***		0.140***		
		[0.075, 0.205]		[0.073, 0.205]		
	HML	0.380***		0.379***		
		[0.322, 0.438]		[0.319, 0.439]		
Quality-minus-junk	Intercept	0.626*	74.42	0.648*	69.47	68.79
Asness, Frazzini and Pedersen (2014)		[-0.048, 1.300]	[44.80, 95.20]	[-0.055, 1.308]	[55.48, 79.46]	
	QMJ	0.369***		0.358***		
		[0.148, 0.589]		[0.130, 0.591]		
	MKT	-0.097		-0.116		
		[-0.763, 0.568]		[-0.772, 0.580]		
	SMB	0.209***		0.206***		
		[0.157, 0.260]		[0.151, 0.262]		
	HML	0.338***		0.340***		
		[0.288, 0.388]		[0.289, 0.392]		
Panel B: GLS						
CAPM	Intercept	1.362***	48.05	1.323***	47.06	42.65
		[0.941, 1.783]	[26.96, 100.00]	[0.846, 1.802]	[14.11, 65.30]	
	MKT	-0.788***		-0.749***		
		[-1.206, -0.369]		[-1.227, -0.287]		
Fama and French (1992)	Intercept	1.397***	88.49	1.353***	85.43	85.43
		[0.924, 1.870]	[82.86, 97.71]	[0.846, 1.878]	[80.03, 89.89]	
	MKT	-0.827***		-0.783***		
		[-1.298, -0.356]		[-1.311, -0.283]		
	SMB	0.179***		0.178***		
		[0.145, 0.213]		[0.142, 0.215]		
	HML	0.348***		0.350***		
		[0.312, 0.385]		[0.310, 0.391]		
Quality-minus-junk	Intercept	0.966***	89.48	0.982***	87.36	86.42
Asness, Frazzini and Pedersen (2014)		[0.395, 1.537]	[83.20, 96.40]	[0.383, 1.616]	[81.20, 90.90]	
	QMJ	0.378***		0.359***		
		[0.204, 0.552]		[0.172, 0.539]		
	MKT	-0.425		-0.438		
		[-0.984, 0.133]		[-1.052, 0.157]		
	SMB	0.197***		0.196***		
		[0.160, 0.234]		[0.156, 0.235]		
	HML	0.359***		0.359***		
		[0.321, 0.398]		[0.320, 0.399]		

The table summarises risk premia estimates and cross-sectional fit for a selection of models with tradable risk factors on a cross-section of monthly excess returns for 25 Fama-French size/value portfolios. Each model is estimated via OLS and GLS. We report point estimates and 5% confidence intervals for risk premia, which are constructed based on the asymptotic normal distribution, and cross-sectional R^2 and its (5%, 95%) confidence level constructed as in Lewellen, Nagel, and Shanken (2010) for FM estimation. In Bayesian Fama-MacBeth estimation, we provide the posterior mean of λ, denoted by $\bar{\lambda}_j$, its (2.5%, 97.5%) credible intervals, the posterior mode and median of the cross-sectional R^2, as well as its (5%, 95%) credible intervals. *, ** and *** denote significance at the 90%, 95% and 99% level, respectively.

Table C12: Two-pass regressions with tradable factors: 25 Fama-French portfolios

Model	Factors	FM $\hat{\lambda}_j$	FM R^2_{adj}	BFM $\bar{\lambda}_j$	BFM $R^2_{adj,mode}$	BFM $R^2_{adj,median}$
			Panel A: OLS			
CCAPM	Intercept	1.444**	6.62	1.814**	-2.02	5.45
		[0.155, 2.732]	[-4.35, 93.74]	[0.106, 3.664]	[-4.23, 41.75]	
	ΔC_{nd}	0.383		0.254		
		[-0.221, 0.988]		[-0.462, 0.956]		
Scaled CAPM	Intercept	4.489***	16.17	3.731*	17.39	25.65
		[1.479, 7.499]	[-14.29, 61.14]	[-0.268, 7.514]	[-3.91, 57.83]	
	cay	1.141*		0.563		
		[-0.198, 2.480]		[-2.262, 3.085]		
	MKT	-2.119		-1.444		
		[-4.890, 0.653]		[-5.241, 2.643]		
	$MKT \times cay$	-8.554		-5.188		
		[-19.227, 2.119]		[-17.911, 9.415]		
Scaled HC-CAPM	Intercept	4.405***	13.86	3.343*	38.95	36.59
		[1.182, 7.629]	[-26.32, 54.53]	[-0.500, 7.182]	[-0.06, 66.30]	
	cay	1.038		0.552		
		[-0.444, 2.520]		[-1.957, 2.848]		
	ΔY	0.437		0.034		
		[-0.421, 1.295]		[-0.995, 0.928]		
	MKT	-2.049		-1.148		
		[-5.031, 0.933]		[-4.946, 2.772]		
	$\Delta Y \times cay$	1.428		0.724		
		[-1.047, 3.903]		[-3.458, 4.645]		
	$MKT \times cay$	-7.782		-4.127		
		[-18.579, 3.015]		[-17.926, 10.532]		
			Panel B: GLS			
CCAPM	Intercept	2.274***	-2.51	2.336***	-3.03	-0.56
		[1.386, 3.162]	[-4.35, 98.96]	[1.360, 3.266]	[-4.03, 11.09]	
	ΔC_{nd}	0.139		0.088		
		[-0.114, 0.391]		[-0.222, 0.383]		
Scaled CAPM	Intercept	2.623***	49.49	2.668***	56.26	45.67
		[0.794, 4.451]	[16.57, 78.29]	[0.859, 4.376]	[4.39, 71.46]	
	cay	0.362		0.205		
		[-0.759, 1.483]		[-0.853, 1.277]		
	MKT	-0.596		-0.642		
		[-2.412, 1.220]		[-2.325, 1.149]		
	$MKT \times cay$	0.644		0.310		
		[-5.695, 6.984]		[-5.988, 6.529]		
Scaled HC-CAPM	Intercept	2.486**	50.14	2.604***	58.32	46.98
		[0.510, 4.463]	[11.58, 77.26]	[0.801, 4.372]	[5.58, 73.71]	
	cay	0.361		0.199		
		[-0.902, 1.623]		[-0.879, 1.269]		
	ΔY	-0.415*		-0.242		
		[-0.878, 0.048]		[-0.672, 0.157]		
	MKT	-0.479		-0.588		
		[-2.441, 1.482]		[-2.357, 1.241]		
	$\Delta Y \times cay$	0.395		0.150		
		[-1.761, 2.552]		[-1.718, 2.036]		
	$MKT \times cay$	1.158		0.477		
		[-5.888, 8.205]		[-5.890, 6.968]		

The table summarises risk premia estimates and cross-sectional fit for a selection of models with nontradable risk factors on a cross-section of quarterly excess returns for 25 Fama-French size/value portfolios. Each model is estimated via OLS and GLS. We report point estimates and 5% confidence intervals for risk premia, which are constructed based on the asymptotic normal distribution, and cross-sectional R^2 and its (5%, 95%) confidence level constructed as in Lewellen, Nagel, and Shanken (2010) for FM estimation. In Bayesian Fama-MacBeth estimation, we provide the posterior mean of λ, denoted by $\bar{\lambda}_j$, its (2.5%, 97.5%) credible intervals, the posterior mode and median of the cross-sectional R^2, as well as its (5%, 95%) credible intervals. *, ** and *** denote significance at the 90%, 95% and 99% level, respectively.

Table C13: Two-pass regressions with tradable factors: 25 Fama-French + 17 industry portfolios

		FM		BFM		
Model	Factors	$\hat{\lambda}_j$	R^2_{adj}	$\bar{\lambda}_j$	$R^2_{adj,mode}$	$R^2_{adj,median}$
		Panel A: OLS				
Carhart (1997)	Intercept	0.755**	47.20	0.780**	38.35	37.69
		[0.167, 1.342]	[8.03, 85.59]	[0.158, 1.395]	[22.22, 54.27]	
	MKT	-0.162		-0.188		
		[-0.759, 0.435]		[-0.808, 0.432]		
	SMB	0.144***		0.142***		
		[0.082, 0.206]		[0.078, 0.207]		
	HML	0.320***		0.316***		
		[0.251, 0.388]		[0.245, 0.388]		
	UMD	0.759		0.673		
		[-0.360, 1.878]		[-0.476, 1.802]		
q-factor model	Intercept	0.744***	45.05	0.761***	36.47	36.00
Hou, Xue, and Zhang (2015)		[0.244, 1.243]	[1.38, 83.38]	[0.239, 1.287]	[17.11, 55.08]	
	ROE	0.173		0.156		
		[-0.139, 0.485]		[-0.154, 0.481]		
	IA	0.287***		0.279***		
		[0.117, 0.456]		[0.117, 0.455]		
	ME	0.230***		0.221***		
		[0.135, 0.325]		[0.121, 0.320]		
	MKT	-0.199		-0.211		
		[-0.703, 0.304]		[-0.741, 0.311]		
Liquidity Factor	Intercept	0.958***	2.77	0.963***	-1.24	6.20
Pástor and Stambaugh (2003)		[0.417, 1.499]	[-5.13, 41.13]	[0.377, 1.527]	[-4.35, 33.40]	
	LIQ	0.210		0.153		
		[-1.001, 1.421]		[-1.077, 1.445]		
	MKT	-0.274		-0.281		
		[-0.822, 0.274]		[-0.851, 0.340]		
		Panel B: GLS				
Carhart (1997)	Intercept	0.839***	88.26	0.909***	84.86	83.97
		[0.467, 1.210]	[75.62, 95.57]	[0.487, 1.339]	[78.95, 88.12]	
	MKT	-0.235		-0.309		
		[-0.610, 0.140]		[-0.737, 0.120]		
	SMB	0.162***		0.161***		
		[0.134, 0.190]		[0.130, 0.193]		
	HML	0.351***		0.350***		
		[0.316, 0.386]		[0.312, 0.390]		
	UMD	1.426***		1.184***		
		[0.832, 2.020]		[0.541, 1.861]		
q-factor model	Intercept	1.107***	42.89	1.103***	37.42	36.31
Hou, Xue, and Zhang (2015)		[0.761, 1.454]	[0.27, 73.41]	[0.714, 1.486]	[20.22, 52.16]	
	ROE	0.270**		0.247**		
		[0.057, 0.482]		[0.008, 0.502]		
	IA	0.277***		0.263***		
		[0.134, 0.420]		[0.107, 0.428]		
	ME	0.221***		0.216***		
		[0.156, 0.287]		[0.144, 0.288]		
	MKT	-0.524***		-0.520***		
		[-0.869, -0.179]		[-0.900, -0.138]		
Liquidity Factor	Intercept	1.186***	28.97	1.159***	18.11	23.73
Pástor and Stambaugh (2003)		[0.874, 1.497]	[-1.97, 76.87]	[0.803, 1.519]	[4.38, 52.53]	
	LIQ	0.089		0.065		
		[-0.567, 0.744]		[-0.696, 0.871]		
	MKT	-0.598***		-0.571***		
		[-0.909, -0.288]		[-0.936, -0.212]		

Model	Factors	FM $\hat{\lambda}_j$	FM R^2_{adj}	BFM $\bar{\lambda}_j$	BFM $R^2_{adj,mode}$	BFM $R^2_{adj,median}$
Panel A: OLS						
CAPM	Intercept	0.966***	4.67	0.957***	-1.13	3.06
		[0.427, 1.506]	[-2.50, 54.90]	[0.382, 1.522]	[-2.46, 28.63]	
	MKT	-0.277		-0.269		
		[-0.823, 0.269]		[-0.838, 0.311]		
Fama-French (1992)	Intercept	1.016***	43.30	1.002***	34.25	33.70
		[0.540, 1.491]	[1.82, 81.66]	[0.517, 1.500]	[21.94, 46.14]	
	MKT	-0.445*		-0.430*		
		[-0.916, 0.026]		[-0.912, 0.068]		
	SMB	0.147***		0.144***		
		[0.086, 0.208]		[0.077, 0.208]		
	HML	0.316***		0.313***		
		[0.248, 0.383]		[0.242, 0.383]		
Quality-minus-junk	Intercept	0.836***	49.31	0.836***	38.61	39.34
Asness et all (2019)		[0.323, 1.350]	[9.14, 84.49]	[0.314, 1.365]	[24.59, 55.77]	
	QMJ	0.153		0.147		
		[-0.065, 0.372]		[-0.066, 0.373]		
	MKT	-0.293		-0.289		
		[-0.796, 0.210]		[-0.814, 0.233]		
	SMB	0.179***		0.175***		
		[0.114, 0.243]		[0.105, 0.240]		
	HML	0.302***		0.300***		
		[0.234, 0.369]		[0.230, 0.373]		
Panel B: GLS						
CAPM	Intercept	1.192***	30.60	1.170***	26.02	25.73
		[0.884, 1.500]	[0.58, 78.48]	[0.815, 1.517]	[6.00, 51.18]	
	MKT	-0.604***		-0.583***		
		[-0.911, -0.298]		[-0.923, -0.227]		
Fama-French (1992)	Intercept	1.182***	86.16	1.150***	82.72	82.34
		[0.853, 1.510]	[73.03, 93.53]	[0.788, 1.519]	[77.36, 86.62]	
	MKT	-0.596***		-0.564***		
		[-0.923, -0.268]		[-0.937, -0.198]		
	SMB	0.160***		0.160***		
		[0.132, 0.187]		[0.129, 0.191]		
	HML	0.349***		0.348***		
		[0.314, 0.384]		[0.310, 0.386]		
Quality-minus-junk	Intercept	0.970***	86.74	0.977***	82.27	82.76
Asness et all (2019)		[0.606, 1.334]	[74.51, 93.35]	[0.561, 1.369]	[77.59, 86.93]	
	QMJ	0.293***		0.278***		
		[0.159, 0.427]		[0.125, 0.439]		
	MKT	-0.393**		-0.399*		
		[-0.754, -0.033]		[-0.791, 0.012]		
	SMB	0.166***		0.165***		
		[0.138, 0.194]		[0.134, 0.196]		
	HML	0.353***		0.352***		
		[0.318, 0.388]		[0.315, 0.392]		

The table summarises risk premia estimates and cross-sectional fit for a selection of models with tradable risk factors on a cross-section of monthly excess returns for 25 Fama-French and 17 industry portfolios. Each model is estimated via OLS and GLS. We report point estimates and 5% confidence intervals for risk premia, which are constructed based on the asymptotic normal distribution, and cross-sectional R^2 and its (5%, 95%) confidence level constructed as in Lewellen, Nagel, and Shanken (2010) for FM estimation. In Bayesian Fama-MacBeth estimation, we provide the posterior mean of λ, denoted by $\bar{\lambda}_j$, its (2.5%, 97.5%) credible intervals, the posterior mode and median of the cross-sectional R^2, as well as its (5%, 95%) credible intervals. *, ** and *** denote significance at the 90%, 95% and 99% level, respectively.

Table C14: Two-pass regressions with nontradable factors: 25 Fama-French + 17 industry portfolios

Model	Factors	Fama-MacBeth $\hat{\lambda}_j$	Fama-MacBeth R^2_{adj}	Bayesian Estimation $\bar{\lambda}_j$	Bayesian Estimation $R^2_{adj,mode}$	Bayesian Estimation $R^2_{adj,median}$
Panel A: OLS						
CCAPM	Intercept	1.817***	3.56	1.975***	-1.24	2.09
		[0.905, 2.729]	[-2.50, 62.08]	[0.795, 3.135]	[-2.45, 26.80]	
	ΔC_{nd}	0.189		0.123		
		[-0.216, 0.594]		[-0.285, 0.502]		
Scaled CAPM	Intercept	2.141***	15.28	2.344***	2.98	13.06
		[0.529, 3.754]	[-7.89, 95.68]	[0.692, 3.944]	[-4.51, 40.69]	
	cay	1.408*		0.574		
		[-0.182, 2.999]		[-0.935, 1.939]		
	MKT	0.108		-0.142		
		[-1.471, 1.686]		[-1.747, 1.499]		
	$cay \times MKT$	-2.554		-2.159		
		[-8.811, 3.702]		[-8.813, 4.620]		
Scaled CCAPM	Intercept	1.607***	17.09	1.983***	9.5	14.68
		[0.394, 2.820]	[-7.89, 100.00]	[0.761, 3.180]	[-3.72, 41.89]	
	cay	1.137		0.511		
		[-0.394, 2.669]		[-0.871, 1.865]		
	ΔC_{nd}	0.452		0.175		
		[-0.103, 1.007]		[-0.261, 0.598]		
	$cay \times \Delta C_{nd}$	-0.102		0.030		
		[-1.752, 1.547]		[-1.175, 1.292]		
HC-CAPM	Intercept	2.185***	6.11	2.221***	-1.47	6.44
		[0.720, 3.650]	[-5.13, 60.05]	[0.667, 3.735]	[-4.29, 30.93]	
	ΔY	0.398*		0.201		
		[-0.011, 0.808]		[-0.290, 0.667]		
	MKT	0.123		0.050		
		[-1.392, 1.638]		[-1.513, 1.650]		
Panel B: GLS						
CCAPM	Intercept	2.351***	2.64	2.442***	-0.57	2.18
		[1.700, 3.003]	[-2.50, 97.95]	[1.612, 3.258]	[-2.04, 12.96]	
	ΔC_{nd}	0.211**		0.132		
		[0.034, 0.387]		[-0.081, 0.351]		
Scaled CAPM	Intercept	2.516***	39.51	2.653***	43.32	34.70
		[1.467, 3.566]	[10.45, 74.11]	[1.616, 3.705]	[3.60, 65.39]	
	cay	0.783**		0.424		
		[0.056, 1.511]		[-0.311, 1.119]		
	MKT	-0.504		-0.639		
		[-1.548, 0.541]		[-1.674, 0.406]		
	$cay \times MKT$	3.785*		2.327		
		[-0.412, 7.981]		[-2.053, 6.791]		
Scaled CCAPM	Intercept	2.244***	3.31	2.378***	2.96	3.56
		[1.378, 3.109]	[-7.89, 81.66]	[1.539, 3.237]	[-4.76, 16.90]	
	cay	0.742*		0.425		
		[-0.006, 1.489]		[-0.316, 1.104]		
	ΔC_{nd}	0.160		0.119		
		[-0.078, 0.398]		[-0.116, 0.342]		
	$cay \times \Delta C_{nd}$	0.355		0.190		
		[-0.343, 1.053]		[-0.455, 0.840]		
HC-CAPM	Intercept	2.908***	38.10	2.808***	44.54	33.56
		[2.053, 3.762]	[8.54, 73.72]	[1.809, 3.826]	[1.94, 65.91]	
	ΔY	-0.155		-0.089		
		[-0.381, 0.072]		[-0.357, 0.174]		
	MKT	-0.892**		-0.793		
		[-1.742, -0.043]		[-1.803, 0.208]		

Model	Factors	FM		BFM		
		$\hat{\lambda}_j$	R^2_{adj}	$\bar{\lambda}_j$	$R^2_{adj,mode}$	$R^2_{adj,median}$
Panel A: OLS						
Scaled HC-CAPM	Intercept	2.099***	13.72	2.243**	15.99	21.86
		[0.549, 3.649]	[-13.89, 48.75]	[0.491, 3.955]	[-1.21, 47.71]	
	cay	1.124		0.529		
		[-0.326, 2.574]		[-1.066, 1.981]		
	ΔY	0.088		0.106		
		[-0.314, 0.489]		[-0.399, 0.599]		
	MKT	0.169		-0.068		
		[-1.340, 1.679]		[-1.805, 1.744]		
	$cay \times \Delta Y$	1.277		0.579		
		[-1.361, 3.914]		[-1.930, 2.919]		
	$cay \times MKT$	-2.125		-1.605		
		[-8.058, 3.808]		[-8.349, 5.354]		
Durable CCAPM	Intercept	2.281**	23.16	2.220**	14.23	14.59
		[0.025, 4.537]	[-7.89, 100.00]	[0.403, 4.028]	[-3.59, 41.78]	
	ΔC_{nd}	0.544**		0.194		
		[0.054, 1.034]		[-0.163, 0.525]		
	ΔC_d	0.481		0.104		
		[-0.101, 1.062]		[-0.353, 0.531]		
	MKT	-0.102		-0.063		
		[-2.466, 2.262]		[-1.948, 1.863]		
Panel B: GLS						
Scaled HC-CAPM	Intercept	2.662***	38.48	2.698***	43.12	35.02
		[1.626, 3.698]	[4.33, 72.67]	[1.555, 3.844]	[2.31, 64.66]	
	cay	0.726*		0.416		
		[-0.029, 1.481]		[-0.324, 1.146]		
	ΔY	-0.165		-0.088		
		[-0.440, 0.110]		[-0.372, 0.198]		
	MKT	-0.652		-0.685		
		[-1.684, 0.379]		[-1.816, 0.438]		
	$cay \times \Delta Y$	0.681		0.333		
		[-0.648, 2.009]		[-1.017, 1.616]		
	$cay \times MKT$	3.266		2.123		
		[-0.950, 7.482]		[-2.173, 6.502]		
Durable CCAPM	Intercept	1.958***	29.26	1.994***	4.12	25.58
		[0.634, 3.281]	[-7.89, 67.63]	[0.831, 3.125]	[-2.69, 63.36]	
	ΔC_{nd}	0.164		0.065		
		[-0.065, 0.394]		[-0.126, 0.260]		
	ΔC_d	0.289*		0.123		
		[-0.011, 0.589]		[-0.117, 0.377]		
	MKT	0.002		-0.026		
		[-1.322, 1.326]		[-1.165, 1.125]		

The table summarises risk premia estimates and cross-sectional fit for a selection of models with nontradable risk factors on a cross-section of quarterly excess returns for 25 Fama-French and 17 industry portfolios. Each model is estimated via OLS and GLS. We report point estimates and 5% confidence intervals for risk premia, which are constructed based on the asymptotic normal distribution, and cross-sectional R^2 and its (5%, 95%) confidence level constructed as in Lewellen, Nagel, and Shanken (2010) for FM estimation. In Bayesian Fama-MacBeth estimation, we provide the posterior mean of λ, denoted by $\bar{\lambda}_j$, its (2.5%, 97.5%) credible intervals, the posterior mode and median of the cross-sectional R^2, as well as its (5%, 95%) credible intervals.*, ** and *** denote significance at the 90%, 95% and 99% level, respectively.

Table C15: Posterior factor probabilities and risk premia of 2.6 million sparse models

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

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

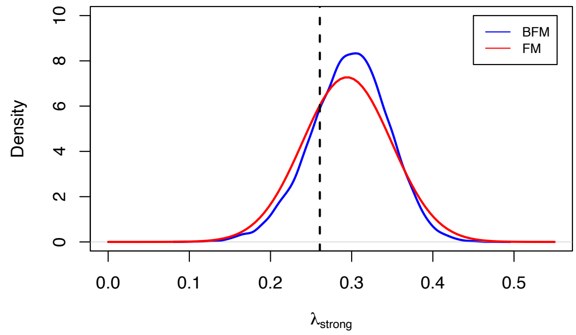
Table C16: Factor Models with highest posterior probability (continuous spike-and-slab, $\psi = 10$)

	model:									
factor:	1	2	3	4	5	6	7	8	9	10
HML	✓	✓	✓	✓	✓		✓	✓	✓	✓
MKT*	✓	✓	✓	✓	✓	✓	✓	✓	✓	
SMB*		✓			✓	✓				✓
STRev	✓		✓		✓	✓	✓		✓	✓
IPGrowth	✓	✓	✓		✓		✓		✓	
BEH_PEAD	✓		✓		✓	✓			✓	✓
PE			✓					✓		
CMA*		✓	✓	✓	✓					
TERM		✓						✓	✓	✓
UMD		✓	✓		✓	✓		✓		
DIV		✓	✓	✓	✓	✓				
BW_ISENT	✓		✓			✓	✓	✓		✓
NONDUR		✓			✓	✓		✓	✓	
DeltaSLOPE		✓			✓	✓				
SERV	✓	✓	✓	✓				✓		✓
REAL_UNC		✓		✓	✓		✓	✓	✓	✓
Oil		✓			✓		✓	✓	✓	
LIQ_TR			✓					✓	✓	
STOCK_ISS		✓			✓		✓		✓	✓
LIQ_NT	✓				✓		✓	✓		
FIN_UNC		✓			✓	✓	✓	✓		
UNRATE	✓	✓				✓		✓		
DEFAULT				✓				✓	✓	
HJTZ_ISENT			✓	✓			✓			
NetOA	✓			✓	✓			✓	✓	
INV_IN_ASSETS	✓		✓				✓	✓	✓	
CMA	✓		✓	✓	✓	✓	✓	✓		✓
COMP_ISSUE	✓			✓			✓	✓		✓
RMW*								✓	✓	
HML*		✓	✓	✓			✓	✓	✓	✓
MKT			✓		✓		✓	✓		✓
ACCR				✓			✓	✓		
MACRO_UNC		✓	✓	✓		✓			✓	
ROA		✓	✓	✓	✓			✓	✓	
LTRev			✓		✓	✓				✓
SMB			✓			✓	✓	✓		✓
ASSET_Growth		✓	✓					✓	✓	✓
INTERM_CAP_RATIO			✓	✓	✓		✓			
DISSTR	✓	✓		✓	✓	✓	✓		✓	
BAB	✓		✓	✓				✓	✓	
IA	✓							✓		✓
PERF					✓		✓		✓	✓
MGMT										
O_SCORE				✓			✓			✓
ROE		✓	✓			✓	✓			
BEH_FIN	✓			✓		✓	✓	✓	✓	
GR_PROF	✓			✓			✓			
QMJ	✓			✓			✓			
RMW		✓		✓						✓
SKEW									✓	
HML_DEVIL	✓		✓				✓	✓		
Probability (%)	0.0122	0.0111	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0089

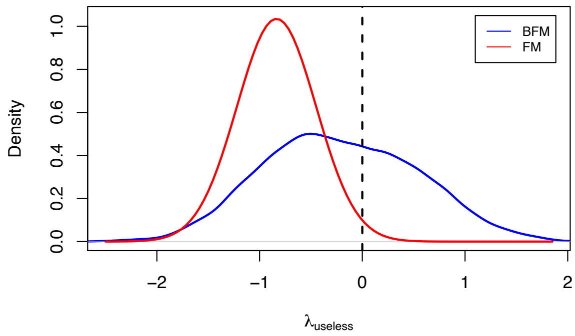
Factors and posterior model probabilities of ten most likely specifications computed using the continuous spike and slab approach of section II.2.3, $\psi = 10$, 51 factors, and all possible models with up to 5 factors, yielding about 2.25 quadrillion models and a model prior probability of the order of 10^{-16}. Specifications organised by columns with the symbol ✓ indicating that the factor in the corresponding row is included. The data is monthly, 1973:10 to 2016:12. Test assets: cross-section of 25 Fama-French size and book-to-market and 30 Industry portfolios. The 51 factors considered are described in Table B.1 of Appendix B.

D Additional Figures

Panel A: OLS

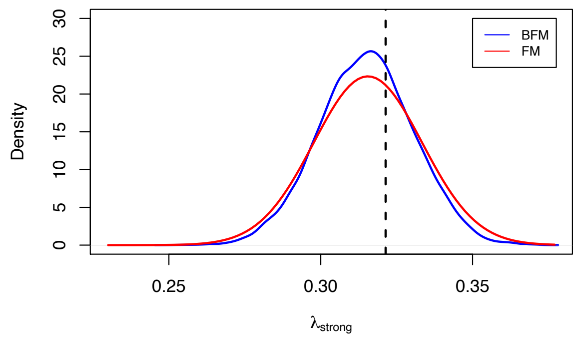


(a) Strong factor

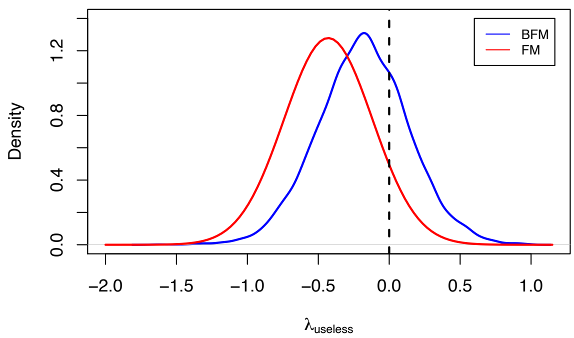


(b) Useless factor

Panel B: GLS



(c) Strong factor



(d) Useless factor

Figure D1: Posterior distribution of the risk premia estimates in a misspecified model that includes both strong and irrelevant factors.

The graph presents posterior distribution of risk premia estimates for a misspecified model with both strong and useless factors in one representative simulation. Panels (a) and (b) display posterior distribution of the BFM-OLS estimates of risk premia, along with the frequentist distribution implied by the point estimates and standard errors. Panels (c) and (d) report the same objects for GLS. T =1000.

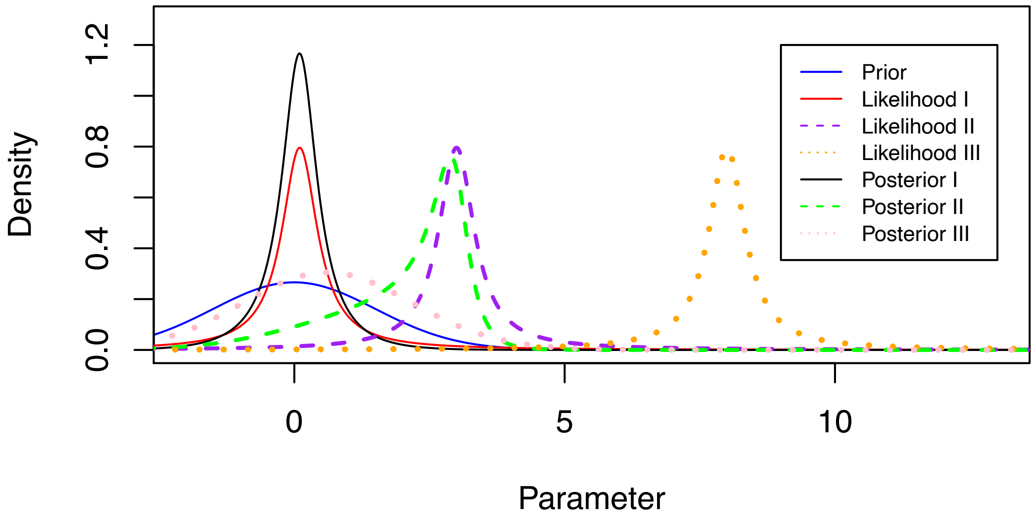


Figure D2: Posterior distributions with heavy-tailed likelihood and thin-tailed prior

Posterior shapes generated by a thin-tails (Gaussian) prior and a heavy-tails (Cauchy) likelihood, as the likelihood peak moves away from the prior mean. Posteriors I, II and III are generated, respectively, by the likelihood I, II and III.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

SRC

Systemic Risk Centre

The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Tel: +44 (0)20 7405 7686
systemicrisk.ac.uk
src@lse.ac.uk