Marginal Likelihood based Factor Selection in the International Setting

Stefan van de Vorst (434864)
Supervisor: dr. Xiao Xiao
Co-Reader: Prof. dr. M. van der Wel

Erasmus School of Economics
Erasmus University Rotterdam

Master Thesis Econometrics and Management Science

Specialization: Quantitative Finance

September 20, 2019

Abstract

The relevance for the identification of priced risk factors on the international level has increased tremendously in the last couple of decades. We revisit the recent work of Barillas and Shanken (2018) and Chib et al. (2018) who introduce a marginal likelihood based factor selection methodology. We argue that the specification of the Barillas and Shanken (2018) priors of the alpha's (across the candidate models) implies a prior bias towards sparse factor models, and find simulation results indicate that the factor selection methodology tends to favour sparser factor models, as opposed to the factor model implied by the simulated DGP, excessively. We find we can drastically increase the precision of the factor selection methodology by increasing the spreads of the priors of the alpha's, and find the precision of the methodology to be robust in a setting with student-t distributed factors. We apply the factor selection methodology, using priors for the alpha's with increased spreads, to select priced risk factors out of a set of prominent global factors as proposed in the literature, and find our selected factor model outperforms several prominent factor models proposed in the literature in terms of relative pricing performance.

Contents

1 Introduction

The relevance for the identification of priced risk factors on the international level has increased tremendously in the last couple of decades, along with the share of investors with foreign equity holdings. Brusa et al. (2014) write that aggregate foreign equity holdings as a percentage of global gross domestic product have increased steadily from roughly 3% in the 1980's to 30% in 2011. Following the traditional CAPM of Sharpe (1964) and Lintner (1965), the excess-return of the market portfolio is the only risk factor that carries a price of risk, and is able to fully explain the cross-section of expected excess-returns of all assets. Brusa et al. (2014) argue that, under the assumption of purchasing power parity (PPP), the traditional CAPM can easily be extended to global markets, the World CAPM. Currently, there exists an extensive body of literature that discusses anomalies of the (World) CAPM, and as such, the literature provides us with numerous (global) candidate (excess-return) factors that help explain the anomalies of the (World) CAPM (Brusa et al. (2014), Fama and French (2012), Asness et al. (2013)).

The relevance for the identification of priced risk factors increases with the number of proposed (global) candidate (excess-return) factors, as the identification of priced risk factors helps us to identify which of the candidate factors help explain expected excess-returns of (global) assets, and which of the candidate factors are in fact non-risk factors with expected excess-returns that can be fully explained by other, priced, risk factors. The quest to identify priced (excess-return) risk factors that fully explain the cross-section of asset expected excess-returns corresponds to the quest to find the mean-variance efficient portfolio (Huberman and Kandel (1987), Back (2015)). Thus, identifying priced (excess-return) risk factors on the international level is relevant from an explanatory viewpoint as well as from the viewpoint of an investor who aims to find the mean-variance efficient portfolio on the international level.

The recent research of Barillas and Shanken (2018) introduces a Bayesian, marginal likelihood based, factor identification (or, selection) methodology that allows us to filter out priced risk factors from a set of (global) candidate excess-return factors. Barillas and Shanken (2018) compare the relative pricing performance of candidate factor models simultaneously by the means of the marginal likelihoods of the respective models. The marginal likelihood of a model is defined as the likelihood that a model will generate the observed data, given a prior on the model parameters. The marginal likelihood based factor model comparison methodology may also be interpreted as a factor selection methodology. Out of a set of candidate factors, one could simply select the combination of factors that form the factor model with the highest marginal likelihood, as compared to other factor combinations, as priced risk factors.

For each candidate factor model, Barillas and Shanken (2018) specify the prior of the unconditional mean of the proposed priced risk factors, or, alternatively, the alpha, as a proper distribution and specify the prior of the remaining "nuisance" parameters as an improper Jeffreys (1961) prior. Chib et al. (2018) revisit the framework of Barillas and Shanken (2018) and point out that the marginal likelihood based factor selection methodology is unsound as the improper Jeffreys (1961) priors of the "nuisance" parameters across the candidate models depend on arbitrary constants that may vary across the priors. The marginal likelihoods are therefore non-comparable across models and cannot be used to locate the true priced risk factors. Chib et al. (2018) show that the "nuisance" parameters across the models are connected by invertible maps and, using the change-of-variable technique, Chib et al. (2018) derive new improper priors for the "nuisance" parameters across the models that all depend on a single arbitrary constant. Using the improper priors of Chib et al. (2018), the marginal likelihoods of the candidate factor models can be properly compared as the single arbitrary constant, common across all improper priors, always cancels out when constructing Bayes factors. Conveniently, the methodology allows the marginal likelihoods to be derived as closed form expressions.

A Bayesian, marginal likelihood based, factor selection methodology, similar to Barillas and Shanken (2018) and Chib et al. (2018), has been developed by Chib and Zeng (2018). Chib and Zeng (2018) base their methodology on the assumption that excess-return factors are student-t distributed, while the methodology of Barillas and Shanken (2018) and Chib et al. (2018) is based on the assumption that excess-return factors are normally distributed. While the methodology of Chib and Zeng (2018) accounts for fat tails exhibited by empirical (stock) return data (Fama (1965)), marginal likelihoods are not available as closed form solutions and must be estimated using MCMC methods, which may require substantial computing power. Chib et al. (2018) consider the advances in their paper, as well as the advances of Chib and Zeng (2018), as complementary, and state the advances open doors to an exciting new wave of reliable Bayesian work on the comparison of factor models.

The marginal likelihood based factor selection methodology presented by Barillas and Shanken (2018) and Chib et al. (2018) is related to a GMM based factor selection approach as discussed in Cochrane (2005). Both methodologies attempt to partition a set of candidate excess-return factors into a set of priced risk factors that, ceteris paribus, affect the SDF, and a set of non-risk factor that, ceteris paribus, do not affect the SDF. Only the former set of factors consists of priced risk factors, as investors only demand a risk premium on an asset's expected return when it is exposed to a factor that, ceteris paribus, affects the SDF. The marginal likelihood based approach of Barillas and Shanken (2018) and Chib et al. (2018) evaluates which particular partition of the candidate factors is most supported by the data. The GMM approach splits the candidate factors by separating factors with estimated direct effects on the SDF that are significant from factors with estimated direct effects on the SDF that are insignificant.

Unlike the GMM based factor selection methodology of Cochrane (2005), the marginal likelihood based factor selection methodology of Barillas and Shanken (2018) and Chib et al. (2018), does not require the use of test-asset data, as the information contained in test-asset data cancels out when constructing Bayes factors. Indeed, in their research, Barillas and Shanken (2017) find that, although test-asset data provides valuable information when assessing the pricing performance of a factor model in an absolute setting, test-asset data provides no information when simultaneously comparing the relative pricing performance of candidate factor models. Many papers in the empirical literature, for example, Brusa et al. (2014), Hou et al. (2015) and Hou et al. (2011), frame the comparison of the relative pricing performance of factor models in terms of success in pricing, solely, test-assets, which Barillas and Shanken (2017) find may lead to a false inference about factor model comparison. Furthermore, Barillas and Shanken (2017) provide a thorough discussion on their finding that, when the comparison of the relative pricing performance of factor models is framed appropriately in terms of success in pricing both test-assets as well as excluded factors, the extent to which each factor model is able to price excluded factors, not test-assets, is what matters for factor model comparison.

The fact that the marginal likelihood based factor selection methodology of Barillas and Shanken (2018) and Chib et al. (2018) is not based on test-asset data gives it an advantage over test-asset based factor selection methodologies (such as Cochrane (2005), or Pukthuanthong et al. (2019)) in certain scenarios. In a scenario where only a few test-assets are of interest, test-asset based factor selection methods might fail to select factors that are actually priced risk factors, but not sufficiently related to the particular set of test-assets in question, even when a large number of observations are available. When a large set of test-assets is of interest, and all priced risk factors are assumed to be sufficiently related to the test-assets, accuracy of test-asset based factor selection methods may suffer from a large N small T issue when, given a number of available observations T, the number of test-assets N is sufficiently large. The marginal likelihood based methodology, conveniently, does not suffer from these issues.

We extend upon the the research of Chib et al. (2018) and Barillas and Shanken (2018), and further investigate the specification of the priors of the unconditional means of the proposed priced risk factors, or alpha's, across the candidate models. Barillas and Shanken (2018) specify the priors of the alpha's (across the candidate models) as proper normal distributions with means of zero. Barillas and Shanken (2018) derive a theoretical restriction on the potential magnitude of hyper-parameter k, governing the spreads of the priors, and set it to equal the squared maximum (attainable) Sharpe ratio (over the portfolio) of the candidate factors, divided by the number of candidate factors, as the unconditional mean of the candidate factors is directly related to the squared maximum Sharpe ratio of the candidate factors. We argue that the Barillas and Shanken (2018) priors of the alpha's, with means of zero, imply a prior bias towards sparse factor models, as economic intuition suggests that the (absolute values of the) unconditional means of excess-returns of a set of tradeable factors are positive, instead of zero, such that investors are compensated for bearing risk. The prior bias towards sparse models reflects our preference of a sparse over a less-sparse factor model, on the condition that the sparse model is statistically valid. We argue that spreads of the priors of the alpha's that are excessively narrow (or strict) may imply an excessive prior bias towards sparse factor models, in the sense that less-sparse models will only be preferred over sparse models if posterior evidence against the statistical validity of the sparse models is excessively strong.

In a simulation study, we we find the precision of the marginal likelihood based factor selection methodology of Chib et al. (2018) and Barillas and Shanken (2018) to be wanting, when the priors of the alpha's (across the candidate models) are specified as suggested by Barillas and Shanken (2018). Using the Barillas and Shanken (2018) priors of the alpha's, the marginal likelihood based factor selection methodology tends to excessively favour sparser factor models, as opposed to the true factor model as implied by the simulated DGP, in turn suggesting that the priors imply an excessive prior bias towards sparse factor models. Indeed, we find we can substantially improve upon the precision of the marginal likelihood based factor selection methodology by setting hyper-parameter k, though conflicting with theoretical restrictions on the potential magnitude of k, equal to a *multiple* of the squared maximum Sharpe ratio of the candidate factors, divided by the number of candidate factors, effectively increasing the spreads of the priors of the alpha's and decreasing the prior bias towards sparse factor models. Using our specification of the priors of the alpha's, we find the precision of the marginal likelihood based factor selection methodology to be robust in a setting with student-t, as opposed to normally, distributed factors and to be much more satisfactory than the precision of the GMM based factor selection methodology of Cochrane (2005).

In an empirical study, we use the marginal likelihood based factor selection methodology of Chib et al. (2018) and Barillas and Shanken (2018), using priors for the alpha's with increased spreads, to select priced risk factors out of a set of prominent global (excess-return) factors proposed in the literature. Our selected factor model outperforms several prominent factor models proposed in the literature in terms of pricing performance w.r.t. excluded candidate factors, and in terms of the overall ability to explain differences across the cross-section of expected excess-returns of global stock portfolios. It remains a challenge to fully explain the cross-section of expected excess-returns of global stocks however, as we find none of our considered factor models are likely able to price all of our global stock portfolios without pricing error.

The remainder of our paper will be organised as follows. In section 2, we revisit the research of Barillas and Shanken (2017) and discuss the marginal likelihood based and GMM based factor selection methodologies of, respectively, Chib et al. (2018) and Barillas and Shanken (2018), and Cochrane (2005). We introduce (global) candidate factor and test-asset data in section 3. Section 3 also provides a brief discussion on the backgrounds of our candidate factors. Sections 4 and 5 respectively present our simulation study and our empirical study. Section 6 concludes.

2 Methodology

2.1 Preliminaries

We consider an international investor, situated in home country j, who invests in assets across various countries. Following the fundamental asset valuation equation of Cochrane (2005), the expected discounted excess-return of country i's asset, from the perspective of a country j investor equals 0:

$$E_t(M_{t+1}^j r_{j,t+1}^i) = E_t(M_{t+1}^j (R_{i,t+1}^i S_{j,t}^i / S_{j,t+1}^i - R_{j,t}^f)) = 0, \quad (1)$$

where M_t^j denotes the (nominal) stochastic discount factor (SDF) of country j, $R_{j,t+1}^i$ ($r_{j,t+1}^i$) denotes the gross (excess) return of country i's asset in terms of country j's currency, and $S_{j,t}^i$ denotes the exchange rate between foreign currency i and domestic currency j. The exchange rate is defined in units of foreign currency per unit of domestic currency. The gross risk-free rate of country j is denoted by $R_{j,t}^f$. We assume the perspective of an US investor in our research, and suppress j when we consider the US as the home country.

Following Hansen and Jagannathan (1991), we specify the SDF as an affine function of risk factors f_t (a $K \times 1$ vector). In turn, Eq. (1) implies a beta factor model for excess-returns. So, given

$$M_{t+1} = 1 - b_t'[f_{t+1} - E_t(f_{t+1})], \quad E_t(M_{t+1} r_{t+1}^i) = 0, \quad (2)$$

it holds (see Cochrane (2005))

$$E_t(r_{t+1}^i) = \beta_{i,t}' \lambda_t, \quad \beta_{i,t} = var_t(f_{t+1})^{-1} cov_t(f_{t+1}, r_{t+1}^i), \quad \lambda_t = var_t(f_{t+1}) b_t.$$

The vector $\beta_{i,t}$ contains the conditional sensitivities of excess-return r_{t+1}^i to the K risk factors, and the vector λ_t contains the conditional prices of the risks the risk factors carry. When f_{t+1} exclusively contains excess-returns (in our research we consider, exclusively, excess-return factors), we end up with the (asset pricing) factor model

$$r_{t+1} = \boldsymbol{\beta}_t f_{t+1} + \epsilon_{t+1}, \quad \boldsymbol{\beta}_t = [\beta_{1,t}, ..., \beta_{N,t}]', \quad E_t(\epsilon_{t+1}) = 0, \quad (3)$$

with r_{t+1} denoting a $N \times 1$ vector of N test-asset excess-returns, and $\lambda_t = E_t(f_{t+1})$.

2.2 Absolute Evaluation of Asset Pricing Factor Models

Gibbons, Ross and Shanken (Gibbons et al. (1989)) develop a test to evaluate the factor model as given by Eq. (3) (the GRS-test) in an absolute sense. Gibbons et al. (1989) assume an unconditional setting with constant factor sensitivities $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ and prices of risk $E_t(f_t) = E(f_t)$, and test the null $H_0 : \alpha = 0$ in the factor regression model:

$$r_{t+1} = \alpha + \boldsymbol{\beta} f_{t+1} + \epsilon_{t+1}, \quad \epsilon_t \sim N_N(0, \boldsymbol{\Sigma}), \quad \boldsymbol{\beta} = [\beta_1, ..., \beta_N]', \tag{4}$$

disturbances ϵ_t are assumed to be normally distributed. In case $\alpha \neq 0$, the factor regression model does not reduce to the (unconditional form of) the factor model given by Eq. (3), meaning that the cross-section of expected excess-returns is not fully explained by the factors f_t.

We define $\boldsymbol{R} = (r_1, ..., r_T)'$ and $\boldsymbol{F} = (f_1, ..., f_T)'$. The GRS test is an F-test with N restrictions and test statistic

$$z = \frac{T - N - K}{N} \frac{\hat{\alpha}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\alpha}}{1 + Sh(\boldsymbol{F})^2} \sim F(N, T - N - K), \quad \hat{\alpha}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\alpha} = (Sh(\boldsymbol{F}, \boldsymbol{R})^2 - Sh(\boldsymbol{F})^2), \tag{5}$$

with $F(a, b)$ denoting the F-distribution with a and b degrees of freedom. $Sh(\boldsymbol{F})$ and $Sh(\boldsymbol{F}, \boldsymbol{R})$ denote maximum sample Sharpe ratios over, respectively, a portfolio of the K factors and a portfolio of the K factors and N test-assets. Estimate $\hat{\alpha}$ denotes the OLS estimate of α, while $\hat{\boldsymbol{\Sigma}}$ denotes the ML (biased) estimate of $\boldsymbol{\Sigma}$. The null is rejected in case a significant increase in maximum sample Sharpe ratio can be attained by constructing a Sharpe ratio maximizing portfolio consisting of test-assets and factors, as opposed to factors only. Testing $H_0 : \alpha = 0$ by the means of the GRS test is thus equivalent to testing whether the mean-variance efficient portfolio can be constructed by the K risk factors, exclusively.

2.3 Relative Evaluation of Asset Pricing Factor Models

While an (excess-return) factor model can be evaluated in an absolute sense by testing whether it adequately prices a set of test-assets by the means of a GRS test, the GRS test results are not informative for the relative pricing performance of the factor model, as compared to other competing models. Barillas and Shanken (2017) show that, when the comparison of the relative pricing performance of factor models is of interest, the comparison should be framed in terms of success in pricing excluded factors, as opposed to the pricing of test-assets. Barillas and Shanken (2017) find that framing the comparison of the relative pricing performance of factor models in terms of success in pricing, solely, test-assets, may lead to a false inference about model comparison.

2.3.1 Comparing Nested Models

Let us consider a set of factors f_t that can be partitioned $f_t = (f_{1,t}, f_{2,t})'$. We label the (asset pricing) factor model consisting of factors f_t, and the factor model consisting of factors $f_{1,t}$ as model $\mathcal{M}$ and model $\mathcal{M}_1$, respectively. Model $\mathcal{M}_1$ is nested in the model $\mathcal{M}$ in the sense that model $\mathcal{M}_1$ is a restricted version of model $\mathcal{M}$. We write the factor regression model as specified in Eq. (4) consisting of factors f_t as

$$r_t = \alpha_r + \boldsymbol{\beta}_1 f_{1,t} + \boldsymbol{\beta}_2 f_{2,t} + \epsilon_t, \tag{6}$$

after partitioning $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ conform the factor partition $f_t = (f_{1,t}, f_{2,t})'$. The relationship between the parameters of the factor regression model consisting of factors $f_{1,t}$

$$r_t = \alpha_{r1} + \boldsymbol{b} f_{1,t} + e_t, \tag{7}$$

and parameters of regression model (6) depends on the parameters of the regression model

$$f_{2,t} = \alpha_{21} + \boldsymbol{d} f_{1,t} + u_t, \tag{8}$$

where factors excluded from model $\mathcal{M}_1$ are regressed on the factors included in $\mathcal{M}_1$. Substituting regression equation (8) in regression equation (6) gives

$$r_t = (\alpha_r + \boldsymbol{\beta}_2 \alpha_{21}) + (\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \boldsymbol{d}) f_{1,t} + (\boldsymbol{\beta}_2 u_t + \epsilon_t).$$

The relationship between the parameters of the regression model (6) and the regression model (7) is thus given as

$$\alpha_{r1} = \alpha_r + \boldsymbol{\beta}_2 \alpha_{21}, \quad \boldsymbol{b} = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \boldsymbol{d}, \tag{9}$$

under orthogonality conditions that u_t and ϵ_t have means of 0 and are uncorrelated with $f_{1,t}$.

Relationship (9) implies that nested model $\mathcal{M}_1$ is valid, in the sense that the factors in $\mathcal{M}_1$ price all test-assets as well as excluded factors $f_{2,t}$ ($\alpha_{r1} = 0$ and $\alpha_{21} = 0$), if and only if the excluded factors $f_{2,t}$ are priced by the factors in the nested model $\mathcal{M}_1$ ($\alpha_{21} = 0$) and test-assets are priced by the factors in the larger model $\mathcal{M}$ ($\alpha_r = 0$). Furthermore, the relationship implies that, in case excluded factors $f_{2,t}$ are priced by the factors in model $\mathcal{M}_1$ ($\alpha_{21} = 0$), model predictions for test-asset expected excess-returns are identical under both $\mathcal{M}$ and $\mathcal{M}_1$, with identical pricing errors $\alpha_r = \alpha_{r1}$ (following from (9)). Under $\mathcal{M}_1$ and $\mathcal{M}$, predicted expected test-asset excess-returns are $\boldsymbol{b}E(f_{1,t})$ and $\boldsymbol{\beta}_1 E(f_{1,t}) + \boldsymbol{\beta}_2 E(f_{2,t})$, respectively (assuming no pricing errors). In case $\alpha_{21} = 0$, Eq. (8) implies $E(f_{2,t}) = \boldsymbol{d}E(f_{1,t})$, ans so $\mathcal{M}$ predicts $\boldsymbol{\beta}_1 E(f_{1,t}) + \boldsymbol{\beta}_2 \boldsymbol{d} E(f_{1,t}) = (\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \boldsymbol{d}) E(f_{1,t}) = \boldsymbol{b}E(f_{1,t})$, which equals the prediction of $\mathcal{M}_1$.

Thus, in case the excluded factors $f_{2,t}$ are priced by model $\mathcal{M}_1$ ($\alpha_{21} = 0$), $\mathcal{M}_1$ is the superior model (in terms of sparsity), as compared to $\mathcal{M}$, regardless of the pricing performance of model

$\mathcal{M}_1$ with respect to the test-assets. This, because pricing performances of models $\mathcal{M}_1$ and $\mathcal{M}$, with respect to the test-assets, are identical in case $\alpha_{21} = 0$, but we favour a sparse model over a less sparse model. In case model $\mathcal{M}_1$ fails to price factors $f_{2,t}$ ($\alpha_{21} \neq 0$), model $\mathcal{M}_1$ is inferior to model $\mathcal{M}$ in the statistical sense that an asset pricing model solely comprising the factors $f_{1,t}$ wrongly implies $\alpha_{21} = 0$, and in the sense that factors $f_{1,t}$ can not possibly form the mean-variance efficient portfolio, as the mean-variance efficient portfolio comprising both factors $f_{1,t}$ and $f_{2,t}$ will attain a higher Sharpe ratio than the mean-variance efficient portfolio solely comprising factors $f_{1,t}$.

Although it might seem that, in case $\alpha_{21} \neq 0$, the general model $\mathcal{M}$ can only improve upon the pricing of the test-assets, as compared to model $\mathcal{M}_1$, this is not necessarily the case. For example, in case α_{21}, α_r, α_{r1} and $\boldsymbol{\beta}_2$ are scalars, and α_{21} and $\boldsymbol{\beta}_2$ have an opposite sign, it holds that $\alpha_{r1} < \alpha_r$. Thus, in case both α_{r1} and α_r are positive and non-zero, the pricing performance of the larger model $\mathcal{M}$ will be worse than the pricing performance of model $\mathcal{M}_1$, as judged by the magnitude of test-asset alphas, even though $\mathcal{M}$ is the better model.

2.3.2 Comparing Non-Nested Models

When comparing relative pricing performance of non-nested factor models, we can use the fact that each of the factor models is a nested version of the factor model that includes all the factors. Thus, our discussion in section 2.3.1 also has implications in a setting where we compare non-nested models. Let us denote two factor models that, respectively, include factors $f_{1,t} = (\mathrm{I}_t, \mathrm{II}_t)'$ and $f_{2,t} = (\mathrm{I}_t, \mathrm{III}_t)'$ by $\mathcal{M}_1$ and $\mathcal{M}_2$. Both models are nested in the factor model including all factors $f_t = (\mathrm{I}_t, \mathrm{II}_t, \mathrm{III}_t)'$, denoted by $\mathcal{M}$. Factors I, II and III denote arbitrary factors.

Let α_{21}, α_{12} and α_r denote regression constants of, respectively, the regression of III on the factors $f_{1,t}$, the regression of II on the factors $f_{2,t}$ and the regression of test-assets on the factors f_t. Following our discussion in section 2.3.1, model $\mathcal{M}_1$ is valid, in the sense that the factors in the model price all test-assets as well as excluded factor III, if and only if $\alpha_{21} = 0$ and $\alpha_r = 0$. Model $\mathcal{M}_2$ is valid if and only if $\alpha_{12} = 0$ and $\alpha_r = 0$. We can only distinguish between relative pricing performance of models $\mathcal{M}_1$ and $\mathcal{M}_2$ by focusing on the extent of deviations from the excluded-factor restrictions $\alpha_{21} = 0$ and $\alpha_{12} = 0$, regardless of the validity of the restriction $\alpha_r = 0$, which we will illustrate with an example.

Suppose that $\alpha_{21} = 0$, and $\alpha_{12} \neq 0$. In this case, following our discussion in section 2.3.1 model $\mathcal{M}_1$ is superior to model $\mathcal{M}$ in terms of sparsity. On the other hand, model $\mathcal{M}_2$ is inferior to model $\mathcal{M}_1$ in the statistical sense that an asset pricing model solely comprising factors $f_{2,t}$ wrongly implies $\alpha_{12} \neq 0$ and in the sense that the maximum attainable Sharpe ratio of a portfolio of factors $f_{2,t}$ is lower than the maximum attainable Sharpe ratio of a portfolio of factors $f_{1,t}$.

Let α_{r1} and α_{r2}, respectively, denote the constant of the regression of test-assets on the factors in $\mathcal{M}_1$ and $\mathcal{M}_2$. In case $\alpha_{21} = 0$, the relation between the parameters of the factor regression models (including constants) corresponding to factor models $\mathcal{M}$ and $\mathcal{M}_1$ implies $\alpha_r = \alpha_{r1}$. Thus, when $\alpha_r = 0$ model $\mathcal{M}_1$ is not only valid in the sense that the factors $f_{1,t}$ price the factors $f_{2,t}$, but also valid in the sense that the factors $f_{1,t}$ fully explain the cross-section of test-asset expected excess-returns.

Although, $\mathcal{M}_1$ clearly is the superior model as compared to $\mathcal{M}_2$ in case $\alpha_{21} = 0$ and $\alpha_{12} \neq 0$, pricing performance of model $\mathcal{M}_1$ w.r.t. test-assets, as judged by test-asset alphas, can actually be worse as compared to model $\mathcal{M}_2$, in case $\alpha_r \neq 0$. The relation between parameters of the factor regression models (including constants) corresponding to factor models $\mathcal{M}$ and $\mathcal{M}_2$ is given as $\alpha_{r2} = \alpha_r + \boldsymbol{\beta}_1 \alpha_{12}$, with $\boldsymbol{\beta}_1$ denoting the regression coefficient of $f_{1,t}$ in the factor regression model corresponding to the full factor model $\mathcal{M}$. So in case $\alpha_{21} = 0$, $\alpha_{r2} = \alpha_{r1} + \boldsymbol{\beta}_1 \alpha_{12}$. Suppose α_{r2}, α_{r1}, α_{21} and $\boldsymbol{\beta}_1$ are scalars. When $\alpha_{12} \neq 0$ has the opposite sign as $\boldsymbol{\beta}_1$, it holds $\alpha_{r2} < \alpha_{r1}$. In case both α_{r1} and α_{r2} are positive and non-zero, the pricing performance of model $\mathcal{M}_1$ will be worse than the pricing performance of model $\mathcal{M}_2$, as judged by the magnitude of test-asset alphas, even though $\mathcal{M}_1$ is the better model.

The examples discussed in this section and section 2.3.1 thus serve to illustrate that, by focusing on the pricing of test-assets, in isolation of factors, when comparing the relative pricing performance of factor models (nested or non-nested), a false inference about model comparison can be obtained. Test-assets should be solely used to evaluate whether the factors in an factor model fully explain the cross-section of expected excess-returns of the test-assets, but provide no information that is relevant for the comparison of the relative pricing performance of competing factor models. When the comparison of the relative pricing performance of competing factor models is of interest, the extent to which each model is able to price excluded factors is what matters for model comparison.

2.4 Factor Selection: a GMM Approach

A factor is a priced risk factor if and only if the factor has a direct effect on the SDF. A factor is a non-risk factor that is priced by other, priced, risk factors if and only if the factor has no direct effect on the SDF. Factor (model) selection thus essentially boils down to determining which factors have a direct effect on the SDF, and which factors do not. In the current section, we present the classical GMM based factor (model) selection approach (as discussed in Cochrane (2005)). In section 2.5, we discuss the Bayesian marginal likelihood based factor (model) selection approach as introduced by Barillas and Shanken (2018) and Chib et al. (2018).

Let f_t^* denote the vector of H candidate excess-return risk factors. We define the SDF as an affine function of the candidate risk factors (we assume an unconditional setting):

$$M_{t+1} = 1 - b' f^*_{t+1},$$

In case we specify $M_{t+1} = 1 - b'[f^*_{t+1} - E(f^*_{t+1})]$ we end up with the unconditional version of the SDF specification as displayed in Eq. (2). We opt not to use that specification in this setting to avoid having to estimate $E(f^*_{t+1})$. The GMM approach allows us to estimate the direct effects of the various candidate factors on the SDF, b, and thus to isolate priced risk factors from a set of candidate factors. A priced risk factor likely has an estimated direct effect that is significant, a factor with an estimated direct effect that is insignificant is likely to be non-risk factor.

The fundamental asset valuation equation provides us with a set of N moment conditions:

$$E(M_t r_t) = 0, \quad E(u_t(b)) = 0, \quad u_t(b) = r_t - b' f^*_t r_t,$$

with r_t denoting a vector of excess-returns of N test-assets. Given b, $u_t(b)$ captures the pricing error at time t, which is expected to be 0 as implied by the fundamental asset valuation equation. With GMM, we estimate b such that the distance between sample moments and implied population moments is minimized:

$$\hat{b} = \arg\min_b \, g_T(b)' \boldsymbol{W} g_T(b), \quad g_T(b) = \frac{1}{T}\sum_{t=1}^{T} u_t(b) = E_T(u_t(b)), \quad E_T(\cdot) = \frac{1}{T}\sum_{t=1}^{T}(\cdot),$$

with $E_T(\cdot)$ denoting the sample mean. The matrix $\boldsymbol{W}$ is a weighting matrix. Typically, the weighting matrix is set $\boldsymbol{W} = \boldsymbol{I}$ or $\boldsymbol{W} = \boldsymbol{S} = E(u_t(b)u_t(b)')$. In the former case, pricing errors of all test-assets are given equal weights. Pricing errors of test-assets are given higher weights when their respective (co-)variances are smaller and vice versa, in the latter case. The matrix $\boldsymbol{S}$ may be estimated in a first stage estimate of b, where weighting matrix $\boldsymbol{W} = \boldsymbol{I}$ is used. We solve analytically for the GMM estimate $\hat{b}$ (see Cochrane (2005)):

$$\hat{b} = (\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{-1}E_T(r_t), \quad \boldsymbol{X}' = E_T(f^*_t r'_t).$$

The asymptotic distribution of the estimator $\hat{b}$, using weighting matrix $\boldsymbol{W} = \boldsymbol{I}$, is given as:

$$\sqrt{T}(\hat{b} - b) \xrightarrow{d} N(0, \boldsymbol{V}), \quad \boldsymbol{V} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{S}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{10}$$

When we use weighting matrix $\boldsymbol{W} = \boldsymbol{S}$, the expression $\boldsymbol{V}$ collapses to $\boldsymbol{V} = (\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}$.

We are now equipped with the tools to evaluate whether a candidate factor $f^*_{2,t}$ has a direct effect on the SDF. We estimate $b = [b'_1 \; b_2]'$ (where b_2 is a scalar) of the model

$$M_{t+1} = 1 - b' f^*_{t+1} = 1 - (b'_1 f^*_{1,t+1} + b_2 f^*_{2,t+1}), \quad f^*_t = [(f^*_{1,t})' \; f^*_{2,t}]',$$

and test whether the candidate factor $f^*_{2,t}$ has a significant direct effect on the SDF by the means of an asymptotic z-test. We thus test $H_0 : b_2 = 0$, where the asymptotic distribution of the test statistic follows directly from Eq. (10)

$$Z = \frac{\hat{b}_2}{\sqrt{var(\hat{b}_2)}} \xrightarrow{d} N(0, 1).$$

To split the set of candidate factors into a set of priced risk factors and a set of non-risk factors, we employ a hybrid backward elimination / forward selection strategy, see Algorithm 1. After applying our hybrid factor selection strategy, we end up with a set of factors in f_t that all have a statistically significant direct effect on the SDF, and are therefore likely to be priced risk factors. None of factors in $\tilde{f}_t$ have a statistically significant direct effect on the SDF, all factors in $\tilde{f}_t$ are therefore likely to be non-risk factors.

Algorithm 1 Factor Selection by hybrid backward elimination / forward selection

1: Let f_t and $\tilde{f}_t$ respectively denote potential risk factors included in the model for the SDF, and excluded in the model for the SDF. Start of with including all H candidate risk factors in f_t, so that $\tilde{f}_t$ is empty.

2: Estimate b of the model $M_{t+1} = 1 - b' f_{t+1}$.

3: Test for significance of each individual factor in f_{t+1} (we use a 5% significance level). If one or more factors are found to have an insignificant effect, transfer the factor with the weakest significance from f_t to $\tilde{f}_t$, and go back to step 2. If all factors are found to have a significant effect, proceed to step 4.

4: Let E denote the number of factors in $\tilde{f}_t$. For all $e = 1, ..., E$, estimate b^e_1 and b^e_2 of the model $M_{t+1} = 1 - (b^e_1)' f_{t+1} - b^e_2 \tilde{f}_{e,t+1}$, with $\tilde{f}_{e,t+1}$ denoting element e of $\tilde{f}_{t+1}$. If none of the factors in $\tilde{f}_{t+1}$ are found to have a significant effect, or $E = 0$, terminate the procedure. If one or more factors have a significant effect, transfer the factor with the strongest significance from $\tilde{f}_t$ to f_t, and return back to step 2.

2.5 Factor Selection: a Bayesian Marginal Likelihood based Approach

Given H candidate excess-return risk factors (collected in vector f^*_t), a total of $J = 2^H - 1$ candidate factor models can be constructed. Each candidate factor might either be a priced risk factor or a non-risk factor with an expected excess-return that is fully explained by other, priced, risk factors (we assume at least one of the factors in our set of candidate factors is a priced risk factor). Following Barillas and Shanken (2018), Chib et al. (2018) compare the (relative) pricing performance of all candidate factor models simultaneously by the means of marginal likelihoods.

The marginal likelihood of a model is defined as the likelihood that a model will generate the observed data, given a prior on the model parameters. The procedure of Chib et al. (2018) can be used as a factor selection procedure. Given H candidate factors and J candidate factor models, one could simply select the factors that make up the model with the highest marginal likelihood as priced risk factors. Let $\mathcal{M}_j$, $j = 1, ..., J$, denote any of the candidate factor models. Each model $\mathcal{M}_j$, $j = 1, ..., J$, partitions the vector of H candidate factors, f_t^*, into a vector of K_j proposed priced risk factors $f_{j,t}$ and a vector of $M_j = (H - K_j)$ implied non-risk factors $\tilde{f}_{j,t}$: $f_t^* = (f_{j,t}',\ \tilde{f}_{j,t}')'$. The partition of the vector of candidate factors f_t^* into a vector of (proposed) priced risk factors $f_{j,t}$ and a vector of (implied) non-risk factors $\tilde{f}_{j,t}$ is unique for each model $\mathcal{M}_j$, $j = 1, ..., J$. Let r_t denote a vector of excess-returns of N test-assets. For each model $\mathcal{M}_j$, $j = 1, ..., J$, we collect test-assets r_t and non-risk factors $\tilde{f}_{j,t}$ in the vector $y_{j,t}$.

Assuming a setting with student-t distributed (excess-returns of) test-assets and candidate factors, we write the joint distribution of test-assets r_t and the partition of $f_t^* = (f_{j,t}',\ \tilde{f}_{j,t}')'$ implied by model $\mathcal{M}_j$, $j = 1, ..., J$, as (we assume an unconditional setting)

$$\begin{pmatrix} f_{j,t} \\ y_{j,t} \end{pmatrix} \sim t_{H+N}\left(\begin{pmatrix} \mu_j \\ \tilde{\mu}_j \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_j & \boldsymbol{\Omega}_j \\ \boldsymbol{\Omega}_j' & \tilde{\boldsymbol{\Omega}}_j \end{pmatrix}, \nu \right), \quad \boldsymbol{\Omega}_j = cov(f_{j,t}, y_{j,t}), \quad y_{j,t} = \begin{pmatrix} \tilde{f}_{j,t} \\ r_t \end{pmatrix},$$

with $t_d(\mu, \boldsymbol{\Sigma}, \nu)$ denoting the d-dimensional multivariate student-t distribution with location parameter μ, scale parameter $\boldsymbol{\Sigma}$ and ν d.o.f.. In case $\nu \to \infty$, the student-t distribution collapses to a normal distribution. We can write model $\mathcal{M}_j$, $j = 1, ..., J$, as a marginal distribution of priced risk factors $f_{j,t}$ and a (conditional) joint distribution of test-assets r_t and non-risk factors $\tilde{f}_{j,t}$ (collected jointly in $y_{j,t}$), conditional on priced risk factors $f_{j,t}$:

$$\begin{aligned} f_{j,t} &= \mu_j + \epsilon_{j,t}, \\ y_{j,t} &= \tilde{\mu}_j + \boldsymbol{\beta}_j(f_{j,t} - \mu_j) + \varepsilon_{j,t}, \end{aligned}$$

where

$$\begin{pmatrix} \epsilon_{j,t} \\ \varepsilon_{j,t} \end{pmatrix} \sim t_{H+N}\left(0, \begin{pmatrix} \boldsymbol{\Sigma}_j & 0 \\ 0 & \tilde{\boldsymbol{\Sigma}}_{y_j} \end{pmatrix}, \nu \right), \quad \tilde{\boldsymbol{\Sigma}}_{y_j} = \tilde{\boldsymbol{\Omega}}_j - \boldsymbol{\Omega}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\Omega}_j, \quad \boldsymbol{\beta}_j = \boldsymbol{\Omega_j}' \boldsymbol{\Sigma}_j^{-1},$$

with $\boldsymbol{\beta}_j$ denoting the matrix of regression coefficients in the regression of the test-assets and non-risk factors (collected jointly in $y_{j,t}$) on the priced risk factors $f_{j,t}$.

Model $\mathcal{M}_j$, $j = 1, ..., J$, proposes the factors in $f_{j,t}$ to be priced risk factors, and therefore proposes that each single factor in $f_{j,t}$ has a direct effect on the SDF:

$$M_{t+1} = 1 - b_j'[f_{j,t+1} - E(f_{j,t+1})], \quad b_j = var(f_{j,t+1})^{-1}\lambda_j,$$

As model $\mathcal{M}_j$, $j = 1, ..., J$, implies that the factors in $\tilde{f}_{j,t}$ are not priced risk factors, the model implies that the factors in $\tilde{f}_{j,t}$ do not affect the SDF (ceteris paribus). The fundamental asset valuation equation

$$E\left(M_{t+1}\begin{pmatrix}f_{j,t}\\ y_{j,t}\end{pmatrix}\right)=0$$

implies $\mu_j=\lambda_j$ and $\tilde{\mu}_j=\boldsymbol{\beta}_j\lambda_j$. Each model $\mathcal{M}_j$, $j=1,...,J$, can thus be written as

$$\begin{aligned}
&f_{j,t}=\Big(\frac{\nu}{\nu-2}\Big)\boldsymbol{\Sigma}_j b_j+\epsilon_{j,t},\quad \epsilon_{j,t}\sim t_{K_j}(0,\boldsymbol{\Sigma}_j,\nu),\\
&\tilde{f}_{j,t}=\boldsymbol{\beta}_{j,f}f_{j,t}+\nu_{j,t},\quad \nu_{j,t}\sim t_{M_j}(0,\tilde{\boldsymbol{\Sigma}}_j,\nu),\quad cov(\epsilon_{j,t},\nu_{j,t})=0,\\
&r_t=\boldsymbol{\beta}_{j,r}f_{j,t}+u_{j,t},\quad u_{j,t}\sim t_N(0,\tilde{\boldsymbol{\Sigma}}_{r_j},\nu),\quad cov(\epsilon_{j,t},u_{j,t})=0,\quad cov(\nu_{j,t},u_{j,t})=\boldsymbol{C}_j,
\end{aligned}$$

Assuming the model is correctly specified, the expected excess-returns of both test-assets r_t and non-risk factors $\tilde{f}_{j,t}$ are fully explained by the risk factors $f_{j,t}$. Following our discussion in section 2.3.1, in case the proposed priced risk factors in a factor model price the non-risk factors as well as test-assets, the test-assets are also priced by the joint set of proposed priced risk factors and non-risk factors. We are able to derive a closed form expression for the marginal likelihood of model $\mathcal{M}_j$, $j=1,...,J$, under the assumption that (excess-returns of) factors and test-assets are normally distributed. Setting $\nu\to\infty$, and following our discussion in section 2.3.1, we re-write model $\mathcal{M}_j$, $j=1,...,J$, as

$$\begin{aligned}
&f_{j,t}=\alpha_j+\epsilon_{j,t},\quad \alpha_j=\boldsymbol{\Sigma}_j b_j,\quad \epsilon_{j,t}\sim N_{K_j}(0,\boldsymbol{\Sigma}_j),\\
&\tilde{f}_{j,t}=\boldsymbol{\beta}_{j,f}f_{j,t}+\nu_{j,t},\quad \nu_{j,t}\sim N_{M_j}(0,\tilde{\boldsymbol{\Sigma}}_j),\\
&r_t=\boldsymbol{\beta}_r f_t^*+u_t,\quad u_t\sim N_N(0,\tilde{\boldsymbol{\Sigma}}_r),
\end{aligned}\tag{11}$$

with shocks $\epsilon_{j,t}$, $\nu_{j,t}$ and u_t being mutually independent, and with the sub-model of the test-assets being identical across all models $\mathcal{M}_j$, $j=1,...,J$.

Let $\beta_{j,f}=\text{vec}(\boldsymbol{\beta}_{j,f})$ and $\beta_r=\text{vec}(\boldsymbol{\beta}_r)$ respectively denote the vectorizations of $\boldsymbol{\beta}_{j,f}$ and $\boldsymbol{\beta}_r$. Let $\sigma_r=\text{vech}(\tilde{\boldsymbol{\Sigma}}_r)$, $\sigma_j=\text{vech}(\boldsymbol{\Sigma}_j)$ and $\tilde{\sigma}_j=\text{vech}(\tilde{\boldsymbol{\Sigma}}_j)$ denote the half vectorizations of the three covariance matrices. The parameters of model $\mathcal{M}_j$, $j=1,...,J$, are then

$$\theta_j=(\alpha_j',\eta_j',\beta_r',\sigma_r')'\in(\Theta_{\alpha_j}',\ \Theta_{\eta_j}',\ \Theta_{\beta_r}',\ \Theta_{\sigma_r}')',\quad \eta_j=(\beta_{j,f}',\sigma_j',\tilde{\sigma}_j')'\in\Theta_{\eta_j},$$

where Θ_{α_j}, Θ_{η_j}, Θ_{β_r} and Θ_{σ_r} respectively denote the parameter spaces of α_j, η_j, β_r and σ_r.

We specify a prior density of parameter θ_j:

$$p(\theta_j|\mathcal{M}_j)=\pi(\alpha_j|\mathcal{M}_j,\eta_j)\psi(\eta_j|\mathcal{M}_j)\psi_r(\beta_r,\sigma_r).$$

As parameters β_r and σ_r are identical across all candidate models, the prior density $\psi_r(\beta_r,\sigma_r)$ is identical across all candidate models as well. Following Barillas and Shanken (2018), Chib et al. (2018) specify the conditional prior of α_j, $\pi(\alpha_j|\mathcal{M}_j,\eta_j)$, as a proper density:

$$\pi(\alpha_j|\mathcal{M}_j,\eta_j)=\phi_{K_j}(\alpha_j|0,k\boldsymbol{\Sigma}_j),\tag{12}$$

with $\phi_d(\cdot|\mu, \boldsymbol{\Sigma})$ denoting the pdf of the d-dimensional multivariate normal distribution with mean μ and covariance $\boldsymbol{\Sigma}$. Barillas and Shanken (2018) derive a theoretical restriction on the potential magnitude of hyper-parameter k (Appendix A.1), controlling the spread of the prior. As the unconditional mean of the candidate factors is directly related to the squared maximum (attainable) Sharpe ratio (over the portfolio) of the candidate factors, Barillas and Shanken (2018) set k to equal the maximum Sharpe ratio of the candidate factors, divided by H:

$$k = \text{ShMax}^2/H, \quad \text{ShMax} = \tau\text{ShMkt},$$

with ShMax and ShMkt respectively denoting the maximum Sharpe ratio of the candidate factors and the Sharpe ratio of the market portfolio. Assuming the candidate factors span the mean-variance efficient portfolio, ShMax equals the Sharpe ratio of the mean-variance efficient portfolio. Under the hypothesis that the market portfolio is not mean-variance efficient, ShMax is specified to be a multiple, governed by τ, of ShMkt. Economic intuition provides limits on the magnitude of τ. Barillas and Shanken (2018) suggests using τ in the range $\tau \in [1.5\ 3]$. We discuss the specification of the proper prior of parameter α_j in further detail in section 2.5.1.

Let $\mathcal{M}_1$ stand for the model in which all of the H candidate factors are (proposed) priced risk-factors (omitting the pricing equation for the test-assets, for simplicity):

$$f_{1,t} = \alpha_1 + \epsilon_{1,t}, \quad \epsilon_{1,t} \sim N_H(0, \boldsymbol{\Sigma}_1), \quad \eta_1 = \sigma_1 = \text{vech}(\boldsymbol{\Sigma}_1).$$

Chib et al. (2018) specify a Jeffreys (1961) improper prior for η_1 (with c an arbitrary constant):

$$\psi(\eta_1|\mathcal{M}_1) = c\tilde{\psi}(\eta_1|\mathcal{M}_1) = c|\boldsymbol{\Sigma}_1|^{-\frac{H+1}{2}}, \quad \tilde{\psi}(\eta_1|\mathcal{M}_1) = |\boldsymbol{\Sigma}_1|^{-\frac{H+1}{2}}. \tag{13}$$

To derive the improper priors of $\eta_j, j = 2, ..., J$, Chib et al. (2018) make use of the fact that the "nuisance" parameters $\{\eta_j\}_{j=1}^J$ are all connected by invertible maps. Thus, the parameter η_1 of model $\mathcal{M}_1$ and parameter η_j of model $\mathcal{M}_j$ $(j > 1)$ are connected by the invertible map

$$\eta_j = g_j(\eta_1), \text{ such that } \eta_1 = g_j^{-1}(\eta_j).$$

We derive the inverse map $\eta_1 = g_j^{-1}(\eta_j)$ in Appendix A.2. The invertible maps can be used to derive the improper priors of $\{\eta_j\}_{j=2}^J$ by applying the change-of-variable technique to η_1 with a prior as specified in Eq. (13):

$$\psi(\eta_j|\mathcal{M}_j) = c\tilde{\psi}(g_j^{-1}(\eta_j)|\mathcal{M}_1)\left|\det\left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'}\right)\right|, \ j = 2, ..., J, \tag{14}$$

the last term being the absolute value of the Jacobian of the transformation. When the prior of η_1 is specified according to Eq. (13), the change of variable technique implies the following improper priors of $\{\eta_j\}_{j=2}^J$ (see Chib et al. (2018) for the derivation):

$$\psi(\eta_j|\mathcal{M}_j) = c|\boldsymbol{\Sigma}_j|^{-\frac{2K_j - H+1}{2}}|\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{H+1}{2}}, \ j = 2, ..., J. \tag{15}$$

In general, improper priors invalidate Bayesian model comparison by marginal likelihoods. When we multiply an improper prior by an arbitrary constant we end up with exactly the same improper prior, as an improper prior is a distribution whose integral over the parameter space is infinite. Thus, when prior $\pi(\theta)$ is improper, $c_i\pi(\theta)$ is exactly the same prior for any $c_i > 0$. The use of improper priors thus renders marginal likelihoods incomparable (in general), as marginal likelihoods will depend on arbitrary constants. Fixing the arbitrary constants at some fixed value does not solve the problem (in general) as Bayes factors depend on that choice. In our setting, though, the use of the Chib et al. (2018) improper priors for "nuisance" parameters $\{\eta_j\}_{j=1}^J$ does not render the marginal likelihoods incomparable. The invertible maps that connect all parameters $\{\eta_j\}_{j=1}^J$ and the change of variable formula (Eq. (14)) force our priors to all depend on a single arbitrary constant c. When we multiply one of our priors for $\{\eta_j\}_{j=1}^J$ with an arbitrary constant, the invertible maps that connect the parameters $\{\eta_j\}_{j=1}^J$ and the change of variable formula force us to multiply all other priors of $\{\eta_j\}_{j=1}^J$ with exactly the same constant as well. Thus, although marginal likelihoods will still depend on a single arbitrary constant c, this arbitrary constant will always cancel out when we construct Bayes factors, rendering our marginal likelihoods comparable.

We collect all observed excess-returns of candidate risk factors and test-assets in the observation matrix $\boldsymbol{Y} = (y_1, ..., y_T)'$, where $y_t = ((f_t^*)', r_t')'$. The marginal likelihood of model $\mathcal{M}_j$, $j = 1, ..., J$, is then given as

$$m(\boldsymbol{Y}|\mathcal{M}_j) = \int_{\Theta_{\sigma_r}} \int_{\Theta_{\beta_r}} \int_{\Theta_{\eta_j}} \int_{\Theta_{\alpha_j}} p(\boldsymbol{Y}|\mathcal{M}_j, \theta_j)\pi(\alpha_j|\mathcal{M}_j, \eta_j)\psi(\eta_j|\mathcal{M}_j)\psi_r(\beta_r, \sigma_r)\mathrm{d}\theta_j.$$

The density function $p(\boldsymbol{Y}|\mathcal{M}_j, \theta_j)$ is the likelihood function implied by model $\mathcal{M}_j$, and can be directly derived from the formulation of model $\mathcal{M}_j$ as defined in Eq. (11). Let observation matrices $\boldsymbol{F} = (f_{j,1}, ..., f_{j,T})'$, $\tilde{\boldsymbol{F}} = (\tilde{f}_{j,1}, ..., \tilde{f}_{j,T})'$ and $\boldsymbol{R} = (r_1, ..., r_T)'$ denote observation matrices of excess-returns of, respectively, (proposed) priced risk factors, (implied) non-risk factors and test-assets. The marginal likelihood of model $\mathcal{M}_j$, $j = 1, ..., J$, can be split up

$$m(\boldsymbol{Y}|\mathcal{M}_j) = m(\boldsymbol{F}|\mathcal{M}_j)m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)m(\boldsymbol{R}),$$

with $m(\boldsymbol{F}|\mathcal{M}_j)$, $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ and $m(\boldsymbol{R})$ denoting the marginal likelihoods of the sub-models, as implied by model $\mathcal{M}_j$ (as defined in Eq. (11)), of, respectively, the priced risk factors, the non-risk factors and the test assets. A derivation is given in Appendix A.3. As the sub-model of the test-assets is identical across all models $\mathcal{M}_j$, $j = 1, ..., J$, the contribution of the information contained in the test-assets to the marginal likelihood of model $\mathcal{M}_j$, $m(\boldsymbol{R})$, is identical across all candidate models $\mathcal{M}_j$, $j = 1, ..., J$. When constructing ratio's of marginal likelihoods (Bayes factors) of candidate models, the information contained in the test-assets thus always cancels out (see A.3). Thus, when simultaneously comparing the (relative) pricing performance of the J candidate factor models by the means of marginal likelihoods, the inclusion of test-assets in the factor models may be omitted (resulting in an empty test-asset vector r_t and $N = 0$), as the test-assets provide no relevant information for the comparison of the marginal likelihoods. This key insight is in line with our discussion in section 2.3, namely that test-assets provide no information that is relevant for the comparison of (relative) pricing performance of competing factor models. When we omit the inclusion of test-assets in our candidate factor models, the marginal likelihood of model $\mathcal{M}_j$, $j = 1, ..., J$ simplifies to

$$m(\boldsymbol{Y}|\mathcal{M}_j) = m(\boldsymbol{F}|\mathcal{M}_j)m(\tilde{\boldsymbol{F}}|\mathcal{M}_j).$$

Under the assumption of normally distributed factors we are, conveniently, able to derive closed form expressions for the marginal likelihoods $m(\boldsymbol{F}|\mathcal{M}_j)$ and $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$, in turn resulting in a closed form expression for marginal likelihood $m(\boldsymbol{Y}|\mathcal{M}_j)$. Derivations are given in Appendix A.3. The closed form expressions of $m(\boldsymbol{F}|\mathcal{M}_j)$ and $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ are given as:

$$\begin{aligned}
m(\boldsymbol{F}|\mathcal{M}_j) &= k^{-\frac{K_j}{2}} \left(\frac{1}{2}\right)^{\frac{K_j M_j}{2}} \left(\frac{1}{\pi}\right)^{\frac{K_j T}{2}} \mathit{\Gamma}_{K_j}\left(\frac{T-M_j}{2}\right) \left|\boldsymbol{S}_j\right|^{-\frac{T-M_j}{2}} (T+k^{-1})^{-\frac{K_j}{2}}, \\
m(\tilde{\boldsymbol{F}}|\mathcal{M}_j) &= \left(\frac{1}{2}\right)^{-\frac{K_j M_j}{2}} \left(\frac{1}{\pi}\right)^{\frac{M_j(T-K_j)}{2}} \mathit{\Gamma}_{M_j}\left(\frac{T}{2}\right) \left|\tilde{\boldsymbol{S}}_j\right|^{-\frac{T}{2}} \left|\boldsymbol{F}'\boldsymbol{F}\right|^{-\frac{M_j}{2}}, \qquad (16) \\
\boldsymbol{S}_j &= \sum_{t=1}^{T} (f_{j,t} - \hat{\alpha}_j)(f_{j,t} - \hat{\alpha}_j)' + \frac{k^{-1}T}{T+k^{-1}} \hat{\alpha}_j \hat{\alpha}_j', \quad \tilde{\boldsymbol{S}}_j = \sum_{t=1}^{T} (\tilde{f}_{j,t} - \hat{\boldsymbol{\beta}}_{j,f} f_{j,t})(\tilde{f}_{j,t} - \hat{\boldsymbol{\beta}}_{j,f} f_{j,t})',
\end{aligned}$$

with $\mathit{\Gamma}_d(\cdot)$ denoting the d dimensional multivariate gamma function, and $\hat{\alpha}_j$ and $\hat{\boldsymbol{\beta}}_{j,f}$ denoting OLS estimates of α_j and $\boldsymbol{\beta}_{j,f}$.

Model $\mathcal{M}_j$, $j = 1, ..., J$, implies the restriction that all supposed non-risk factors $\tilde{f}_{j,t}$ are priced by the proposed priced risk factors $f_{j,t}$ with a (restricted) pricing error of zero. The marginal likelihood of model $\mathcal{M}_j$, $j = 1, ..., J$, is directly tied to the (negative) impact of the model's zero pricing error restriction on the sample fit of the model via $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ (via the $\tilde{\boldsymbol{S}}_j$ term in $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$). Ceteris paribus, marginal likelihood $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ will reach its maximum when the expected excess-returns of the supposed non-risk factors $\tilde{f}_{j,t}$ are perfectly explained by the proposed priced risk factors $f_{j,t}$ in-sample, i.e. model $\mathcal{M}_j$'s zero pricing error implication that $E(\tilde{f}_{j,t}) - E(\boldsymbol{\beta}_{j,f} f_{j,t}) = 0$ is perfectly supported by the sample data. Ceteris paribus, the less model $\mathcal{M}_j$'s implication that $E(\tilde{f}_{j,t}) - E(\boldsymbol{\beta}_{j,f} f_{j,t}) = 0$ is supported by the sample data, the more marginal likelihood $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ will be negatively affected. In case model $\mathcal{M}_j$'s zero pricing error restriction is poorly supported by the sample data, the supposed non-risk factors $\tilde{f}_{j,t}$ may contain priced risk factors with expected excess-returns that can not be fully explained by the proposed priced risk factors $f_{j,t}$, while the proposed priced risk factors $f_{j,t}$ may contain non-risk factors that have no explanatory power for explaining expected excess-returns. Thus, ceteris paribus, the stronger the (negative) impact of model $\mathcal{M}_j$'s zero pricing error restriction on the model's sample fit, the lower $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ and, in turn, the lower $m(\boldsymbol{Y}|\mathcal{M}_j)$.

Using the marginal likelihoods, we can compute the posterior probability that model $\mathcal{M}_j$, $j = 1, ..., J$, has generated the observed data:

$$P(\mathcal{M}_j|\boldsymbol{Y}) = \frac{m(\boldsymbol{Y}|\mathcal{M}_j)P(\mathcal{M}_j)}{\sum\limits_{i=1}^{J} m(\boldsymbol{Y}|\mathcal{M}_i)P(\mathcal{M}_i)},$$

where $P(\mathcal{M}_i)$ denotes the prior probability that model $\mathcal{M}_i$ has generated the data.

2.5.1 Alpha Prior

As previously discussed, the prior distribution of parameter α_j is, for each factor model $\mathcal{M}_j$, $j = 1, ..., J$, specified as

$$\pi(\alpha_j|\mathcal{M}_j, \eta_j) = \phi_{K_j}(\alpha_j|0, k\boldsymbol{\Sigma}_j).$$

The prior distribution implies a prior belief that $E(\alpha_j) = 0$, which in fact conflicts with the economic intuition that (absolute values of) expected excess-returns of risky traded portfolios are positive such that investors are compensated for bearing risk. In the subsequent discussion, we will show the prior distribution implies a prior bias towards sparse factor models, meaning models with few proposed priced risk factors, as opposed to less-sparse factor models. As previously discussed, each factor model implies the restriction that all supposed non-risk factors are priced by the proposed priced risk factors with zero pricing error. The prior distribution ensures sparse factor models will be favoured over less-sparse models, unless posterior evidence against the validity of the zero-pricing error restrictions of sparse factor models is strong enough. Thus, although the prior distribution may conflict with economic intuition, it reflects our preference towards sparse factor models over less-sparse models, on the condition that the zero pricing error restrictions of the sparse factor models are valid.

To investigate the implications of the prior distribution of parameter α_j in closer detail, we continue our discussion with an example setting. We consider a setting with two normally distributed candidate factors, $f_{1,t}$ and $f_{2,t}$, and suppose that $f_{1,t}$ is known to be a priced risk-factor while $f_{2,t}$ is a candidate factor that may either be a priced risk factor, or a non-risk factor. In the latter case, factor $f_{2,t}$ will be priced by $f_{1,t}$. This leaves us with two candidate factor models. In the first model, denoted by $\mathcal{M}_1$, $f_{2,t}$ is a priced risk factor that is not priced by $f_{1,t}$. Factor $f_{2,t}$ is priced by $f_{1,t}$ in the second model, denoted by $\mathcal{M}_2$. Model $\mathcal{M}_1$ is written as

$$f_t = \alpha + \epsilon_t, \quad \epsilon_t \sim N_2(0, \boldsymbol{\Sigma}), \quad f_t = \begin{pmatrix} f_{1,t} \\ f_{2,t} \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}.$$

For illustrative purposes, we treat α_1 and $\boldsymbol{\Sigma}$ as given. We specify an informative prior distribution for α_2

$$\pi(\alpha_2) = \phi_1(\alpha_2|m, ks^2), \quad s^2 = \sigma_2^2 - \rho^2\sigma_1^{-2},$$

with hyper-parameters m and k.

Model $\mathcal{M}_1$ can be re-written as a marginal distribution of $f_{1,t}$ and a conditional distribution of $f_{2,t}$, conditional on $f_{1,t}$ such that

$$f_{1,t} = \alpha_1 + u_t, \quad u_t \sim N_1(0, \sigma_1^2),$$
$$f_{2,t} = \alpha_e + \beta f_{1,t} + v_t, \quad v_t \sim N_1(0, s^2),$$

where

$$\alpha_e = \alpha_2 - \beta\alpha_1, \quad \beta = \rho\sigma_1^{-2}, \quad s^2 = \sigma_2^2 - \rho^2\sigma_1^{-2}, \quad cov(u_t, v_t) = 0.$$

Factor model $\mathcal{M}_2$ implies $f_{2,t}$ to be priced by $f_{1,t}$, and is thus a restricted version of (the conditional version of) model $\mathcal{M}_1$ with restriction $\alpha_e = 0$. Assuming factors f_t are normally distributed, model $\mathcal{M}_1$ is statistically valid. As model $\mathcal{M}_2$ is a restricted version of model $\mathcal{M}_1$, it is statistically valid if and only if pricing error α_e indeed equals 0.

Let $\boldsymbol{F}_1 = (f_{1,1}, ..., f_{1,T})'$ and $\boldsymbol{F}_2 = (f_{2,1}, ..., f_{2,T})'$ respectively denote observation vectors of $f_{1,t}$ and $f_{2,t}$. Given α_1, β and $\boldsymbol{F}_1$, observing $\boldsymbol{F}_2$ implies observing $\boldsymbol{E}_2 = \boldsymbol{F}_2 + \beta\alpha_1\iota_T - \beta\boldsymbol{F}_1$ with $\boldsymbol{E}_2 \sim N_T(\alpha_2\iota_T, s^2 I_T)$, where ι_T and I_T respectively denote a $(T \times 1)$ vector of ones and the $(T \times T)$ identity matrix. Let $\hat{\alpha}_2 = \bar{\boldsymbol{E}}_2$ denote the sample mean of $\boldsymbol{E}_2$, and let $\hat{\alpha}_e = \hat{\alpha}_2 - \beta\alpha_1$. Given an observed $\hat{\alpha}_2$, $\hat{\alpha}_e$ varies as we vary $\beta\alpha_1$. If, for example, we decrease $\beta\alpha_1$, $\hat{\alpha}_e$ increases.

We compare our models $\mathcal{M}_1$ and $\mathcal{M}_2$ by the means of their marginal likelihoods, respectively denoted by $m(\boldsymbol{F}_1, \boldsymbol{F}_2|\mathcal{M}_1)$ and $m(\boldsymbol{F}_1, \boldsymbol{F}_2|\mathcal{M}_2)$. The marginal likelihoods can be split up:

$$m(\boldsymbol{F}_1, \boldsymbol{F}_2|\mathcal{M}_1) = m(\boldsymbol{F}_1|\mathcal{M}_1)m(\boldsymbol{F}_2|\boldsymbol{F}_1, \mathcal{M}_1), \quad m(\boldsymbol{F}_1, \boldsymbol{F}_2|\mathcal{M}_2) = m(\boldsymbol{F}_1|\mathcal{M}_1)m(\boldsymbol{F}_2|\boldsymbol{F}_1, \mathcal{M}_2),$$

with $m(\boldsymbol{F}_1|\mathcal{M}_j)$ and $m(\boldsymbol{F}_2|\boldsymbol{F}_1, \mathcal{M}_j)$ respectively denoting the marginal likelihoods of the sub-models of $f_{1,t}$ and $f_{2,t}$ as implied by model $\mathcal{M}_j$, $j = 1, 2$. As $m(\boldsymbol{F}_1|\mathcal{M}_1) = m(\boldsymbol{F}_1|\mathcal{M}_2)$, only marginal likelihoods $m(\boldsymbol{F}_2|\boldsymbol{F}_1, \mathcal{M}_j)$, $j = 1, 2$ matter for model comparison. Let us abbreviate $m(\boldsymbol{F}_2|\boldsymbol{F}_1, \mathcal{M}_1)$ and $m(\boldsymbol{F}_2|\boldsymbol{F}_1, \mathcal{M}_2)$ with m_1 and m_2 respectively, for ease of reference.

Marginal likelihoods m_1 and m_2 can be written in closed form:

$$m_2 = \phi_T(\boldsymbol{F}_2|\beta\boldsymbol{F}_1, s^2 I_T),$$
$$m_1 = (T + k^{-1})^{-\frac{1}{2}}k^{-\frac{1}{2}}(2\pi s^2)^{-\frac{T}{2}}\exp(-\frac{1}{2}s^{-2}d), \quad c = (T + k^{-1})^{-\frac{1}{2}}k^{-\frac{1}{2}}, \tag{17}$$
$$d = (\boldsymbol{E}_2 - \tilde{\alpha}_2\iota_T)'(\boldsymbol{E}_2 - \tilde{\alpha}_2\iota_T) + k^{-1}(\tilde{\alpha}_2 - m)^2, \quad \tilde{\alpha}_2 = (T + k^{-1})^{-1}(\boldsymbol{E}_2\iota_T + k^{-1}m),$$

we suppress the derivation of m_1 as the derivation is similar to the marginal likelihood derivations presented in Appendix A.3. In terms of marginal likelihood, model $\mathcal{M}_1$ is preferred over model

$\mathcal{M}_2$ in case $m_1 > m_2$, or $m_1/m_2 > 1$. Marginal likelihood m_2 reaches maximum value when $\hat{\alpha}_e = 0$, in this case the restriction $\alpha_e = 0$ does not negatively impact the sample fit of restricted model $\mathcal{M}_2$ and the sample fits of both models $\mathcal{M}_1$ and $\mathcal{M}_2$ are identical. Holding T fixed, as $|\hat{\alpha}_e|$ increases, m_2 decreases, and when $|\hat{\alpha}_e| \to \infty$, $m_2 \to 0$. Fixing $|\hat{\alpha}_e| \neq 0$, as $T \to \infty$, $m_2 \to 0$. Thus, the more the restriction $\alpha_e = 0$ hurts the sample fit of restricted model $\mathcal{M}_2$, or the more evidence is available (due to more observations being available) that in fact the restriction $\alpha_e = 0$ is invalid, the lower marginal likelihood m_2 will be.

For fixed k, marginal likelihood m_1 reaches maximum value when we set $m = \hat{\alpha}_2$, the sample mean of $\boldsymbol{E}_2$. Setting $m = \hat{\alpha}_2$ gives $\tilde{\alpha}_2 = \hat{\alpha}_2$, and as $\hat{\alpha}_2 = \bar{\boldsymbol{E}}_2$ is the OLS estimate of α_2, setting $m = \hat{\alpha}_2$ will thus result in d reaching its minimal value, in turn resulting in m_2 reaching maximal value. As c is a decreasing function of k, m_1 is a decreasing function of k as well, given $m = \hat{\alpha}_2$. Setting $m = \hat{\alpha}_2$ and $k = 0$ thus results in m_1 attaining maximum value

$$m_1 = \phi_T(\boldsymbol{E}_2|\hat{\alpha}_2\iota_T, s^2 I_T) = \phi_T(\boldsymbol{F}_2|(\hat{\alpha}_2 - \beta\alpha_1)\iota_T + \beta\boldsymbol{F}_1, s^2 I_T) = \phi_T(\boldsymbol{F}_2|\hat{\alpha}_e\iota_T + \beta\boldsymbol{F}_1, s^2 I_T).$$

When setting $m = \hat{\alpha}_2$ and $k = 0$, it will always hold that $m_1 \geq m_2$, with $m_1 = m_2$ if and only if $\hat{\alpha}_e = 0$. Thus, when setting $m = \hat{\alpha}_2$ and $k = 0$, model $\mathcal{M}_1$ will always be preferred over model $\mathcal{M}_2$ unless the restriction $\alpha_e = 0$ has absolutely no impact on the sample fit of model $\mathcal{M}_2$.

Given $\hat{\alpha}_2$ and k, the larger the distance between m and $\hat{\alpha}_2$, $|m - \hat{\alpha}_2|$, the smaller m_1, due to d being an increasing function of $|m - \hat{\alpha}_2|$. Consider setting, in contrast to the economic intuition that $|\alpha_2| > 0$ and $|E(\hat{\alpha}_2)| > 0$, $m = 0$ with $k = 0$. In this case m_1 collapses to (with 0_T a $(T \times 1)$ zero vector):

$$m_1 = \phi_T(\boldsymbol{E}_2|0_T, s^2 I_T) = \phi_T(\boldsymbol{F}_2| - (\beta\alpha_1)\iota_T + \beta\boldsymbol{F}_1, s^2 I_T).$$

We know from our previous discussion that, setting $m = 0$ with $k = 0$, $m_1 = m_2$ in case $\hat{\alpha}_2 = 0 = m$ and $\hat{\alpha}_e = 0$. So in case $|m - \hat{\alpha}_2| > 0$, and $\hat{\alpha}_e = 0$, model $\mathcal{M}_2$ will be preferred over $\mathcal{M}_1$. This illustrates that, given $\hat{\alpha}_2$ and k, as $|m - \hat{\alpha}_2|$ increases, either $|\hat{\alpha}_e|$ will have to increase, or T will have to increase with fixed $|\hat{\alpha}_e| > 0$, in order for m_1/m_2 to remain constant. Summarizing, given $|\hat{\alpha}_2| > 0$ (and for fixed k), when setting $m = 0$ as opposed to $m = \hat{\alpha}_2$, the sample fit of model $\mathcal{M}_2$ (measured by the magnitude of $|\hat{\alpha}_e|$) will have to be poorer or more evidence (more observations) will have to be available against the statistical validity of model $\mathcal{M}_2$, in order for model $\mathcal{M}_1$ to maintain the same degree of favourability, relative to model $\mathcal{M}_2$.

The questions remains as to what motivates Barillas and Shanken (2018) and Chib et al. (2018) to set $m = 0$ as opposed to, for example, $\hat{\alpha}_2$. First of all, hyper-parameter m should be chosen based on prior information, a rule that is violated when setting $m = \hat{\alpha}_2$, as $\hat{\alpha}_2$ contains posterior information. Second, setting $m = 0$, albeit conflicting with economic intuition, actually has a clear motivation in our factor model comparison framework. As setting $m = 0$ implies $E(\alpha_2) = 0$, setting $m = 0$ in turn implies a prior belief that the candidate factor $f_{2,t}$ does not carry a risk premium and, therefore, that the factor is a non-risk factor with no direct effect on the SDF (a factor with no risk premium must be uncorrelated with the SDF). The prior belief that $f_{2,t}$ is a non-risk factor in turn implies a prior belief that the sparse model $\mathcal{M}_2$, sparse in the sense that, as opposed to model $\mathcal{M}_1$, only one out of the two factors is regarded as a priced risk factor, is statistically valid. Furthermore, as economic intuition suggests $|E(\hat{\alpha}_2)| > 0$, the prior bias towards sparse model $\mathcal{M}_2$ is expected to be stronger when setting $m = 0$, as opposed to $m = \hat{\alpha}_2$, due to the negative impact of $|m - \hat{\alpha}_2|$ on m_1. The prior bias towards the sparse factor model is perfectly in line with the general preference of a sparse factor model over a less-sparse factor model, under the (prior) assumption that the sparse factor model is statistically valid. Without a proper prior bias towards sparse model $\mathcal{M}_2$, less sparse model $\mathcal{M}_1$ would be favoured over model $\mathcal{M}_2$ too easily a posteriori, as the zero-pricing error restriction imposed on model $\mathcal{M}_2$ can only negatively impact the sample fit of model $\mathcal{M}_2$, while it does not impact the sample fit of model $\mathcal{M}_1$. Summarizing, assuming $|\hat{\alpha}_2| > 0$, setting $m = 0$ implies a prior bias towards sparse model $\mathcal{M}_2$ and ensures model $\mathcal{M}_2$ will only be rejected in favour of the less sparse model $\mathcal{M}_1$ when posterior evidence against the prior belief that sparse model $\mathcal{M}_2$ is statistically valid, and that $f_{2,t}$ is a non-risk factor, is sufficiently strong.

After setting $m = 0$, we are left with the issue of choosing k. Assuming $|\hat{\alpha}_2| > 0$, k has two effects on marginal likelihood m_1 (17). First, k has a negative effect on m_1 via its negative effect on c, and as $k \to \infty$, $c \to 0$ and thus $m_1 \to 0$. Second, k has a positive effect on m_1 via its negative effect on d. As $k \to \infty$, $d \to (\boldsymbol{E}_2 - \hat{\alpha}_2 \iota_T)'(\boldsymbol{E}_2 - \hat{\alpha}_2 \iota_T)$, which is the minimum value d can attain due to the fact that $\hat{\alpha}_2$ is the OLS estimate of α_2. As m_1 is an exponential function of $-d$ but a lineair function of c, the positive effect of k on m_1 will be stronger than the negative effect on the condition that k is sufficiently small. As we keep increasing k, d will start to converge towards its minimum value meaning that the positive effect of k on m_1 will start to fade as we keep increasing k. The negative effect of k on m_1 via c will not fade as we keep increasing k though, and as such the negative effect of k on m_1 will start to dominate the positive effect when k is sufficiently large. Summarizing, if k is set either excessively small or excessively large, m_1 may be penalized excessively in the sense that model $\mathcal{M}_1$, although statistically valid but less sparse, will only be favoured in case the posterior evidence against the statistical validity of sparse model $\mathcal{M}_2$ is exceptionally strong. Setting k either excessively small or excessively large may thus lead towards an excessively strong prior bias towards the sparse, but potentially statistically invalid, model $\mathcal{M}_2$.

3 Data

In this section, we present the empirical data we use in our analysis. Our set of candidate excess-return factors will be discussed first, after which we will turn our discussion to our set(s) of test-assets. Our sample period runs from February 1995 until April 2018, the sample consists of $T = 278$ observations in total. We assume the perspective of an US investor, and denominate all (excess-)returns in US dollars. All monthly (excess-)returns are expressed in percentages.

3.1 Factors

Our set of candidate (excess-return) factors consists of two different aggregate global market factors, WMKT and LWMKT, where WMKT and LWMKT are excess-returns (in excess of the US risk-free rate) of a global market portfolio denominated in, respectively, US dollars and local currencies. Although LWMKT is not an excess-return denominated in US dollars, Brusa et al. (2014) find that the factor mimicking portfolio (assuming an US investor) of LWMKT is highly correlated with LWMKT and delivers virtually identical returns. In addition to the global market factors, our set of candidate factors consists of global versions of the SMB, HML, CMA and RMW factors of Fama and French (2015), a global version of the MOM factor of Carhart (1997), a global version of the BAB factor of Frazzini and Pedersen (2014), a global version of the QMJ factor of Asness et al. (2019), a global version of the DHML factor of Asness and Frazzini (2013), the currency factor Global Tail of Fan et al. (2019) and the currency factors Carry and Dollar of Brusa et al. (2014). Before describing our raw factor data in section 3.1.2, we briefly discuss the backgrounds of our candidate factors.

3.1.1 Factors Background

Brusa et al. (2014) argue that the traditional CAPM can easily be extended to global markets, the World CAPM, under the assumption of (purchasing power parity) PPP. Thus, following the World CAPM (and assuming the perspective of an US investor), the WMKT factor is the only risk factor that carries a price of risk. Following Dumas and Solnik (1995), Brusa et al. (2014) question the PPP assumption of the World CAPM and investigate whether international equity investors are compensated for bearing exchange rate risk. Brusa et al. (2014) argue the assumption on PPP to be unrealistic, as investors who invest abroad like to consume at home, even when deviations from PPP are present. Brusa et al. (2014) present the three-factor International CAPM Redux model, which, in addition to the LWMKT factor, includes two currency factors, Dollar and Carry, which effectively summarise variation in a broad cross-section of bilateral exchange rates and account for exchange rate risk. The Dollar factor is the average excess-return earned by an U.S. investor who invests in a broad portfolio of foreign currencies. The Carry factor is the average excess-return earned by a U.S. investor that goes short (long) in a portfolio of low (high) interest rates currencies.

Asness et al. (2013) and Fama and French (2012) investigate the role of the size, value and momentum effects, as originally discovered by Fama and French (1992) and Jegadeesh and Titman (1993) to be present across the cross-section of expected excess-returns of US stocks, on the international level. Fama and French (2012) construct global versions of the original size (SMB), value (HML) and momentum (MOM) factors of Fama and French (1993) and Carhart (1997) and find the global factors help explain size, value and momentum anomalies of the World CAPM. Following Fama and French (1993), Fama and French (2012) construct the HML value factor by sorting on book-to-price (B/P) ratios that are constructed with book value and price data that is lagged six months to make sure the data would actually have been available at the time of portfolio construction. Asness and Frazzini (2013) recommend to construct B/P ratios with current, as opposed to lagged, price data, as price data would have been known at the time of portfolio construction with certainty, and construct the DHML value factor.

Fama and French (2015) show, using the dividend discount model, that, ceteris paribus, expected earnings (directly linked to expected profitability) are positively related to the expected return of a stock, while, ceteris paribus, expected future investment is negatively related to the expected return of a stock. Indeed, Fama and French (2015) find the presence of profitability and investment effects in the cross-section of US stock returns and find these effects to be unexplained by the original three-factor asset pricing model of Fama and French (1993). Fama and French (2015) propose a new five-factor model that includes the RMW and CMA factors, respectively constructed by sorting stocks on profitability and investment characteristics. In our research, we use global versions of the RMW and CMA factors.

Asness et al. (2019) investigate whether quality stocks command higher prices than low-quality, or junk, stocks. Asness et al. (2019) define quality as characteristics that investors should be willing to pay for. In the research of Asness et al. (2019), profitability, growth and safety characteristics form the basis for the definition of quality. Asness et al. (2019) show that investors indeed pay more for firms with higher quality characteristics, but also find the explanatory power of quality for asset prices to be limited. Consistent with the limited pricing performance of quality, high quality stocks have delivered high risk-adjusted returns while low quality stocks have delivered low risk-adjusted returns. Thus a quality minus junk portfolio (the QMJ factor) that invests long in quality stocks and shorts junk stocks produces a high risk-adjusted return in the US and globally across 24 countries.

Frazzini and Pedersen (2014) argue the basic premise of the CAPM that all agents invest in the market portfolio and leverage or de-leverage the portfolio to suit risk-preferences to be unrealistic, as many investors face leverage constraints. Frazzini and Pedersen (2014) argue that agents with leverage constraints overweight high (market) beta assets to suit risk preferences, causing these assets to offer lower risk adjusted returns. Unconstrained investors can exploit this effect by shorting high beta assets and leveraging up low beta assets. Frazzini and Pedersen (2014) construct a betting against beta (BAB) factor by longing a portfolio of low beta assets, leveraged to a beta of one, and shorting a portfolio of high beta assets, de-leveraged to a beta of one, with offsetting positions in the risk-free asset to make it self-financing. Indeed, the BAB factor produces a high risk-adjusted return in the US and globally across 19 developed countries.

Fan et al. (2019) find that, from the perspective of an investor of any country, US tail risk carries a negative price of risk in the cross-section of currency returns. In their paper, Fan et al. (2019) argue that currencies which offer high returns when US tail risk spikes, receive a negative risk premium as they essentially provide a hedge against US tail risk, and vice versa. Fan et al. (2019) sort currencies according to their sensitivities to the US tail risk (their US tail beta) and show a long-short US tail beta sorted portfolio can identify the global component of the US tail risk factor. Fan et al. (2019) find that their two-factor asset pricing model containing their Global Tail factor explains a large portion of the cross-section of expected returns of carry and momentum currency portfolios (assuming the perspective of an US investor), and outperforms a foreign exchange market CAPM-equivalent single factor model containing only the Dollar factor.

3.1.2 Factors Data

We use MSCI World Total Return Indices denominated in US dollars and local currencies respectively to construct the WMKT and LWMKT factors, the MSCI indices are from the Bloomberg database. Global versions of the SMB, HML, CMA, RMW and MOM factors are extracted from the Kenneth French Data Library[1]. Data of the US risk-free rate is extracted from the Kenneth French Data Library as well (KFDL). Global versions of the BAB, DHML and QMJ factors are obtained from the AQR Data Library[2]. The Carry and Dollar factors are available on Adrien Verdelhan's website[3] (we use the "all currencies" dataset). We construct the Global Tail factor (which we abbreviate with GT) of Fan et al. (2019) using the methodology of Fan et al. (2019).

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

[2] https://www.aqr.com/Insights/Datasets

[3] http://web.mit.edu/adrienv/www/Data.html

Table 1: Summary Statistics Potential Factors

	Mean	Std. Dev	Skewness	Kurtosis	$\hat{\alpha}_{WMKT}$	$\hat{\alpha}_{LWMKT}$
WMKT	0.52	3.95	-0.85	4.55		
LWMKT	0.52	4.25	-0.75	4.71		
SMB	0.11	1.99	-0.22	5.15	0.15*	0.15*
HML	0.30	2.44	0.57	8.33	0.36***	0.35***
DHML	0.24	3.07	0.68	13.1	0.19*	0.19*
RMW	0.38	1.54	-0.05	5.14	0.46***	0.46***
CMA	0.21	1.96	0.78	7.07	0.32***	0.31***
MOM	0.63	4.13	-1.00	9.10	0.77***	0.76***
QMJ	0.49	2.22	0.09	4.42	0.70***	0.69***
BAB	0.94	2.97	-0.30	5.62	0.99***	1.04***
Dollar	0.08	1.79	-0.25	4.02	0.02	-0.04
Carry	0.62	2.41	-0.19	3.76	0.47***	0.49***
GT	-0.30	1.91	0.29	4.89	-0.26**	-0.26**

Summary statistics of our candidate (excess-return) factors. Estimates $\hat{\alpha}_{WMKT}$ and $\hat{\alpha}_{LWMKT}$ denote estimated constants (alpha's) of test regressions, in each regression returns of one of our potential non-market factors are regressed on, respectively, market factor WMKT or market factor LWMKT. Test regressions are performed using full sample data (we assume an unconditional setting). Estimated alphas that differ significantly from 0 are marked with a *, ** and ***, where significance levels of, respectively 10%, 5% and 1% are used. GT is the abbreviation of Global Tail.

Summary statistics of our candidate factors are displayed in Table 1. Estimates $\hat{\alpha}_{WMKT}$ and $\hat{\alpha}_{LWMKT}$ denote estimated alpha's of test regressions, in each regression returns of one of our candidate non-market factors are regressed on a constant (alpha) and on, respectively, market factor WMKT and market factor LWMKT. Test regressions are performed using full sample data (we assume an unconditional setting). The test regression results suggest expected excess-returns of none of the factors, with the exception of the Dollar factor, are fully explained by either WMKT or LWMKT. The results therefore suggest we can improve upon the efficiency of the global market portfolios in terms of mean-variance trade-off. All factors display excess kurtosis and most factors display skewness, contradicting the normality assumption of the Bayesian factor selection methodology of Barillas and Shanken (2018) and Chib et al. (2018). Although Chib and Zeng (2018) relax the assumption of normality and allow for excess kurtosis of the factors, the methodology is still based on the assumption of symmetric distributions. Chib et al. (2018) consider the advances in their paper, as well as those described in Chib and Zeng (2018), as complementary, and state the advances open doors to an exciting new wave of reliable Bayesian work on the comparison of factor models. We will evaluate the robustness of the normality assumption of the Bayesian factor selection methodology of Barillas and Shanken (2018) and Chib et al. (2018) against factors displaying excess kurtosis in a simulation study in section 4.

The sample correlation matrix of our candidate factors is displayed in Figure 11, in Appendix B. We observe a substantial correlation of 0.97 between the WMKT and LWMKT factors, indicating these factors to be almost identical. As both these factors capture the performance of the global market portfolio, this comes at no surprise. We do not observe other correlations between factor pairs of the same magnitude, all other correlations are smaller than 0.5, with the exception of correlations between the pairs HML and DHML (0.68), CMA and HML (0.73), and QMJ and RMW (0.75). The relatively high correlation between HML and DHML can be explained by the fact that both factors are constructed to capture the value effect.

3.2 Test-assets

We consider four sets of test-assets. Our first set consist of Country Market (equity) indices of 20 developed countries: Austria, Australia, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Italy, Japan, the Netherlands, New Zealand, Norway, Singapore, Spain, Sweden, Switzerland, the UK and the USA. The second set includes, for each of the 20 respective countries, 3 different indices: a Country Market (equity) index, a Country Value (equity) index and a Country Growth (equity) index. To construct the Country Growth and Country Value indices of a country, that country's stocks are sorted using univariate 30-40-30 sorts on BE/ME (book-to-market). The third set of test-assets contains 25 global portfolios formed by bi-variate 5×5 sorts on ME (size) and BE/ME, 25 global portfolios formed by bi-variate 5×5 sorts on ME and OP (operating profitability), 25 global portfolios formed by bi-variate 5×5 sorts on ME and INV (investment) and 25 global portfolios formed by bi-variate 5×5 sorts on ME and MOM (momentum). We form our fourth, and aggregate, set of test-assets by combining our second set of test-assets with our third set of test-assets. Our first, second, third and fourth sets thus, respectively, contain $N_1 = 20$, $N_2 = 60$, $N_3 = 100$ and $N_4 = 160$ test-assets. All test-asset returns are US dollar denominated and are extracted from the KFDL[4].

4 Simulation Study

We perform a simulation study to evaluate the performance of our factor model selection procedures discussed in sections 2.4 and 2.5 in a finite sample setting. We will discuss the simulation procedure first, where-after we will discuss our results.

4.1 Simulation Procedure

Let us consider a setting with H candidate excess-return factors (collected in vector f_t^*) and N test-assets (collected in vector r_t). To simulate a financial market in this setting, we may use a total of $J = 2^H - 1$ DGP's. Each of our candidate factors might either be a priced risk factor

[4]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

or a non-risk factor with an expected excess-return that is fully explained by other, priced, risk factors (we assume at least one of our candidate factors to be a priced risk factor). In each DGP, the expected excess-returns of all N test-assets are fully explained by the respective set of priced risk factors. Let $\mathcal{D}_j$, $j = 1, ..., J$, denote any of the possible DGP's. Each DGP $\mathcal{D}_j$, $j = 1, ..., J$, has a unique partition of the set of H candidate factors, f_t^*, into a set of K_j implied priced risk factors $f_{j,t}$ and a set of implied $M_j = (H - K_j)$ non-risk factors $\tilde{f}_{j,t}$: $f_t^* = (f_{j,t}', \ \tilde{f}_{j,t}')'$.

We assume a setting with student-t distributed factors and test-assets. When simulating from student-t distributions, simulated factor and test-asset data exhibits excess kurtosis, reflecting real-world factor and test-asset data. Following the fundamental asset valuation equation and our derivations in section 2.5, each DGP $\mathcal{D}_j$, $j = 1, ..., J$, can be written as

$$\begin{aligned} f_{j,t} &= \Big(\frac{\nu}{\nu - 2}\Big)\boldsymbol{\Sigma}_j b_j + \epsilon_{j,t}, \quad \epsilon_{j,t} \sim t_{K_j}(0, \boldsymbol{\Sigma}_j, \nu), \\ \tilde{f}_{j,t} &= \boldsymbol{\beta}_{j,f} f_{j,t} + \nu_{j,t}, \quad \nu_{j,t} \sim t_{M_j}(0, \tilde{\boldsymbol{\Sigma}}_j, \nu), \\ r_t &= \boldsymbol{\beta}_r f_t^* + u_t, \quad u_t \sim t_N(0, \tilde{\boldsymbol{\Sigma}}_r, \nu), \end{aligned} \tag{18}$$

with shocks $\epsilon_{j,t}$, $\nu_{j,t}$ and u_t being mutually independent. To fix the parameters of a DGP, we first estimate the parameters of the model implied by the DGP by the means of maximum likelihood (ML), and then fix the DGP parameters at the ML estimates to ensure that generated data resembles real-world data. We consider parameter ν given, and consider values of interest $\nu = 5$ (excess kurtosis) and $\nu = \infty$ (normal distribution). Following Chib et al. (2018), who perform a, albeit less extensive, similar simulation study, we only simulate a DGP $\mathcal{D}_j$ in case all the estimates of the parameters in parameter vector $b_j = [b_j^1, ..., b_j^{K_j}]'$ of the model implied by DGP $\mathcal{D}_j$ differ significantly from 0 (we use a significance level of 1%). In case one of the elements of parameter b_j equals 0, one of the factors in $f_{j,t}$ has no direct effect on the SDF, contradicting the implication of respective DGP $\mathcal{D}_j$ that all priced risk factors $f_{j,t}$ have a direct effect on the SDF. Thus a DGP $\mathcal{D}_j$ is valid if and only if none of the elements of parameter b_j equal 0.

For each DGP $\mathcal{D}_j$, parameter $\boldsymbol{\beta}_{j,r} = [\beta_{j,r}^1, ..., \beta_{j,r}^{K_j}]$ is a matrix of $N \times K_j$ regression coefficients, with $\beta_{j,r}^i$, $i = 1, ..., K_j$, denoting the vector of N regression coefficients that correspond with the same factor $f_{j,t}^i$, $f_{j,t} = [f_{j,t}^1, ..., f_{j,t}^{K_j}]'$. When estimating $\beta_{j,r}^i$, we may test the null $H_0 : \beta_{j,r}^i = 0$ with an (asymptotic) F-test. Details of this test are discussed in Appendix A.4. In case the estimate of the parameter $\beta_{j,r}^i$ of the model implied by DGP $\mathcal{D}_j$ does not differ significantly from 0 (we use a significance level of 1%), we set $\beta_{j,r}^i = 0$. In that case, DGP $\mathcal{D}_j$ implies the factor $f_{j,t}^i$ to be a priced risk factor, that does not influence any of the N test-assets. We expect test-asset based factor selection methodologies to be unable to accurately identify $f_{j,t}^i$ as a priced risk factor in such a scenario, as the factor $f_{j,t}^i$ will be unrelated to the relevant test-assets.

4.2 Results

We proceed discussing our simulation results. Table 2 displays our candidate factors considered in the simulation study, each factor is assigned to a roman numeral for ease of reference. Using the $H = 12$ candidate factors, a total of $J = 2^H - 1 = 4095$ potential factor model DGP's may be constructed (see Eq. 18). Each DGP we simulate is simulated with four distinct sets of test-assets. The first, second, third and fourth sets of test-assets respectively contain $N_1 = 20$, $N_2 = 60$, $N_3 = 100$ and $N_4 = 160$ test-assets. We use data of the factors and the four sets of test-assets, as described in sections 3.1 and 3.2 respectively, to estimate the parameters of the models implied by the DGP's we simulate. Each simulated DGP is simulated $Z = 100$ times in total. To keep the disucussion of the results manageable, while ensuring that results are not contingent on a particular DGP, we simulate 13 random eligible DGP's, Chib et al. (2018) follow a similar approach in their simulation study. Table 4 displays the DGP's we simulate. Each simulated DGP is represented by a set of implied priced risk factors, as implied by the respective DGP. Note that each of the DGP's we simulate implies a multi-factor model.

Table 2: Candidate Factors Simulation Study

WMKT	SMB	HML	DHML	RMW	CMA	MOM	QMJ	BAB	Dllr	Crry	GT
I	II	III	IV	V	VI	VII	IIX	IX	X	XI	XII

Roman Numerals Corresponding To Factors. The Dollar and Carry Factors are respectively abbreviated with Dllr and Crry.

We start of by discussing the simulation study results of the Bayesian (marginal likelihood based) factor selection methodology of Barillas and Shanken (2018) and Chib et al. (2018). As discussed in section 2.5, Barillas and Shanken (2018) set hyper-parameter k, governing the spreads of the priors of the alpha's of the candidate models (see Eq. (11) and Eq. (12)) as

$$k = \text{ShMax}^2/H, \quad \text{ShMax} = \tau\text{ShMkt}, \quad \text{ShMkt} = Sh(\text{WMKT}), \tag{19}$$

with ShMax and ShMkt respectively denoting the maximum attainable Sharpe ratio and the Sharpe ratio of the market portfolio, and with Sh(WMKT) denoting the sample Sharpe ratio over the WMKT portfolio, calculated each simulation iteration with simulated data. Economic intuition provides limits on the magnitude of τ, as we do not expect ShMax to deviate too much from ShMkt. Barillas and Shanken (2018) suggest using τ in the range $\tau \in [1.5\ 3]$ and Chib et al. (2018) use $\tau = 3$. As discussed in section 2.5.1 though, we can expect a prior bias towards sparse candidate factor models when the prior means of the alpha's of the candidate models are set to equal zero, as, reflecting the fact that investors want to be compensated for bearing risk, (absolute values of) the true factor means will be positive. Furthermore, as discussed in section 2.5.1, when hyper-parameter k is set too small, the prior bias towards sparse, but potentially statistically invalid, candidate models may be to be too strong, in the sense that the marginal likelihood of less sparse candidate models may be penalized excessively. Thus, although conflicting with economic intuition, we consider values of $\tau \in \{1.5, 2, 3, 5, 10, 20, 30\}$ and investigate the impact of setting $\tau > 3$ on the performance of the Bayesian factor selection methodology.

Table 3: Simulation Study Results Bayesian Factor Selection Methodology I

	τ						
	1.5	2	3	5	10	20	30
			$T = 300$				
Average Accuracy	01 (73)	05 (69)	15 (60)	37 (38)	60 (20)	67 (14)	67 (14)
Minimum Accuracy	00 (99)	00 (99)	00 (98)	00 (98)	11 (82)	41 (52)	42 (53)
			$T = 600$				
Average Accuracy	07 (67)	14 (58)	34 (45)	68 (15)	80 (05)	86 (02)	87 (02)
Minimum Accuracy	00 (99)	00 (99)	00 (98)	10 (90)	56 (40)	77 (10)	74 (10)
			$T = 1200$				
Average Accuracy	19 (55)	40 (37)	50 (30)	88 (15)	92 (05)	94 (00)	95 (00)
Minimum Accuracy	00 (99)	00 (99)	01 (98)	34 (64)	81 (02)	89 (00)	89 (00)

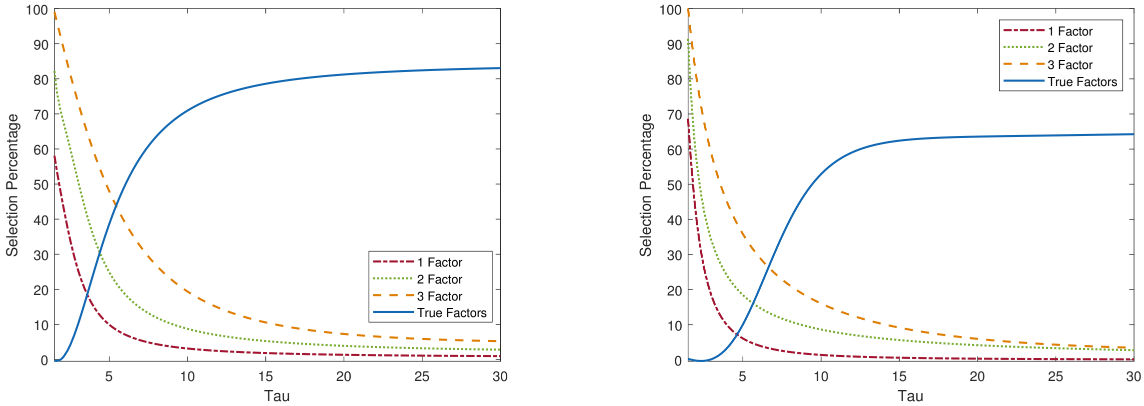
Simulation results Bayesian factor selection methodology. We simulate 13 random DGP's, each DGP is simulated $Z = 100$ times. We apply the Bayesian factor selection methodology to select factors for each simulated DGP, using multiple alternative values for τ. In each "Average Accuracy" row, we display, in plain text, the average selection accuracy observed across the simulated DGP's. In addition, in each "Average Accuracy" row, we display, in (**parentheses**), the average of the percentages of times a sparser, instead of the (DGP implied) true, factor model is selected, observed across the simulated DGP's. In each "Minimum Accuracy" row, we display, in plain text, the minimum selection accuracy observed across the simulated DGP's. In addition, in each "Minimum Accuracy" row, we display, in (**parentheses**), the maximum of the percentages of times a sparser, instead of the (DGP implied) true, factor model is selected, observed across the simulated DGP's. Results are displayed for various sample sizes T. We simulate normally distributed factors.

For each simulated DGP and for each τ, we keep track of the "selection accuracy", defined as the percentage of times the true priced risk factors, as implied by the relevant simulated DGP, are correctly identified out of a total of $Z = 100$ simulation iterations. In addition to the selection accuracy, we keep track of, for each DGP and each τ, the percentage of times (out of $Z = 100$) a sparser factor model (meaning a model with less factors), as opposed to the model implied by the true DGP, is selected. Tables 3 and 12 (Appendix B) display, for each τ, the average of, as well as the minimum of, the selection accuracies observed across the simulated DGP's. In addition, the tables display, in (**parentheses**), for each τ, the average of, as well as the maximum of, the percentages of times a sparser, instead of the (DGP implied) true, factor model is selected, observed across the simulated DGP's. When, for a particular DGP, the percentage of times a sparser, instead of the (DGP implied) true, model is selected is high, the percentage of times we fail to accurately identify all true priced risk factors as implied by the DGP will be high. Thus, a high maximum of the percentages of times a sparser, instead of the true, model is selected implies a low selection accuracy for at least one, or more, DGP('s). Tables 3 and 12 display results in settings with, respectively, normally and student-t distributed factors. Various sample sizes T are considered.

The results displayed in Table 3 are quite surprising. For all relevant sample sizes T, we observe a positive relation between the magnitude of τ and the performance of the Bayesian factor selection methodology, which fades as we increase τ from $\tau = 20$ onwards. For all sample sizes T, both the average as well as the minimum selection accuracy increase as we increase $\tau = 1.5$ to $\tau = 20$. The average as well as the minimum selection accuracy stop increasing as we further increase τ to $\tau = 30$. Interestingly, the overall selection accuracy is dismal when we set $\tau \in \{1.5, 2\}$, as suggested by Barillas and Shanken (2018). In their simulation study, Chib et al. (2018) set $\tau = 3$, and consider sample sizes $T = 600$ and $T = 1200$, with $H = 12$ candidate factors. Chib et al. (2018) obtain average selection accuracies of 30% and 47% for, respectively, sample sizes of $T = 600$ and $T = 1200$, which match our average selection accuracies for these sample sizes. Although Chib et al. (2018) find that the use of their improper priors of the "nuisance" parameters of the candidate models (see Eq. (13) and Eq. (15)), as opposed to the improper priors as specified by Barillas and Shanken (2018), drastically improves upon overall selection accuracy of the Bayesian methodology, we find we can further improve upon overall selection accuracy by increasing τ to $\tau = 20$, which effectively increases k for fixed ShMkt (and H). Furthermore, setting $\tau = 3$ in a setting with sample size $T = 300$, a setting not considered by Chib et al. (2018), results in a dismal overall selection accuracy, and the increase in overall selection accuracy resulting from setting $\tau = 20$ as opposed to $\tau = 3$ is most prominent for this sample size. The results displayed in Table 12 confirm our discussion remains valid in a setting with student-t distributed factors.

The results in Table 3 show that, overall, across the simulated DGP's, the percentage of times a sparser, as opposed to the true, model is selected is excessively high when τ is set in the range $\tau \in [1.5\ 3]$. As a direct result, overall selection accuracy suffers severely (with $\tau \in [1.5\ 3]$). For sample size $T = 300$, the average of the percentages of times a sparser (instead of the true) model is selected, equals 73%, 69% and 60% when τ is set to, respectively, $\tau = 1.5$, $\tau = 2$ and $\tau = 3$. The average of the percentages of times a sparser model is selected reduces to 14% as we increase τ to $\tau = 20$ or $\tau = 30$. As we increase sample size, the average of the percentages of times a sparser model is selected decreases for all values of τ considered. However, we still observe a significant negative relation between the size of τ and the average of the percentages of times a sparser model is selected for the larger sample sizes, and, for the larger sample sizes, we are able to reduce the average of these percentages (close) to 0% when we set $\tau = 20$ or $\tau = 30$. Furthermore, the maximum of the percentages of times a sparser factor model is selected equals 99% when τ is set in the range $\tau \in [1.5\ 3]$, for all sample sizes, and is reduced significantly, for all sample sizes, when we increase τ to $\tau = 20$ or $\tau = 30$. Again, the results displayed in Table 12 confirm our discussion remains valid in a setting with student-t distributed factors.

Figure 1: Simulation Results DGP: WMKT-DHML-MOM-QMJ, WMKT-SMB-DHML-QMJ



We simulate DGP: WMKT-DHML-MOM-QMJ (left hand side) and WMKT-SMB-DHML-QMJ (right hand side), with normally distributed factors and use sample size $T = 300$. Each DGP is simulated $Z = 100$ times. We apply the Bayesian factor selection methodology to select factors for each simulated DGP, using multiple alternative values for τ in the range $\tau \in [1.5\ 30]$. For both DGP's, we plot the percentage of times the true factor model as implied by the DGP is correctly identified, the percentage of times a one factor model is selected, the percentage of times either a one or a two factor model is selected, and the percentage of times either a one, a two or a three factor model is selected against the multiple alternative values for τ considered.

We continue our discussion by simulating individual DGP: WMKT-DHML-MOM-QMJ as well as individual DGP: WMKT-SMB-DHML-QMJ (with $T = 300$ and $\nu = \infty$). For both DGP's, we closely investigate the impact of the magnitude of τ on the selection accuracy of the Bayesian methodology. Figure 1 plots, for both DGP's, the percentage of times the true (DGP implied) factor model is correctly identified, the percentage of times a (false) one factor model is selected, the percentage of times either a one or a two (false) factor model is selected, and the percentage of times either a one, a two or a three (false) factor model is selected against multiple alternative values for τ in the range $\tau \in [1.5\ 30]$. Clearly, for both DGP's, we observe a positive relationship between the magnitude of τ and selection accuracy, in line with the results displayed by Table 3. When τ is set $\tau = 1.5$, we observe, for both DGP's, that the percentage of times a sparser, instead of the true, model is selected approaches 100%. In turn, for $\tau = 1.5$ and both DGP's, we observe a selection accuracy of approximately 0%. For both DGP's, as we increase τ from $\tau = 1.5$ to $\tau = 20$, we decrease the percentage of times a sparser, instead of the true, factor model is selected to approximately 5%, while significantly increasing selection accuracy. The beneficiary effect of increasing τ on selection accuracy fades, for both DGP's, as we keep increasing τ from $\tau = 20$ to $\tau = 30$.

Summarizing, when τ is set in the range $\tau \in [1.5\ 3]$, we find the overall (across the DGP's) percentage of times a sparser, instead of the true, factor model is selected to be excessively high (for the relevant sample sizes), which suggests that setting τ in the range $\tau \in [1.5\ 3]$ results in a excessively strict hyper-parameter k, in turn resulting in an excessively high prior bias towards sparse models. From our discussion in section 2.5.1, we know that, when the prior bias towards sparse models is excessive due to an excessively strict k, sparse models may, a posteriori, still be favoured over less sparse models, even when posterior evidence against the statistical validity of the sparse models is strong. In that case, the excessive prior bias towards sparse models will have a detrimental effect on selection accuracy. Our findings thus strongly suggest to set hyper-parameter k, although conflicting with the theoretical restriction of the potential magnitude of k as discussed in Appendix A.1, to equal a *multiple* of ShMax^2, divided by H:

$$k = (c\text{ShMax})^2/H, \quad c\text{ShMax} = \tau\text{ShMkt},$$

as setting $\text{ShMax} = \tau\text{ShMkt}$ with $\tau > 3$ would imply an unrealistically high maximum Sharpe ratio. This allows us to use larger values of τ, such as $\tau = 20$ (with, for example, $c = 7$), thereby effectively increasing the spreads of the priors of the alpha's of the candidate models and decreasing the excessive prior bias towards sparse factor models, while drastically increasing overall selection accuracy.

For robustness analysis, we also simulate a (fourteenth) one-factor model DGP: WMKT, and investigate whether selection accuracy of the Bayesian methodology is still satisfactory when τ is set $\tau > 3$. Table 13 (Appendix B) displays, for various sample sizes T and normally distributed factors, the selection accuracy for $\tau \in \{1.5, 2, 3, 5, 10, 20, 30\}$. For all sample sizes, we observe the selection accuracy to be similar across all the values of τ considered. Indeed, as we simulate a one-factor model, an excessively strong prior bias towards sparse models, resulting from setting $\tau \in [1.5\ 3]$, no longer negatively impacts selection accuracy.

We proceed to investigate the performance of the Bayesian factor selection methodology, while setting $\tau = 20$, in more detail. Specifically, we analyze the robustness of the methodology in a setting with student-t distributed factors. Table 4 displays, for each simulated DGP, the percentage of times the true priced risk factors are correctly identified, as well as the percentage of times the factor model, as implied by the relevant simulated DGP, ranks among the top five models with the highest marginal likelihoods (in **parentheses**). The Bayesian factor selection methodology performs remarkably well, especially in settings with larger sample sizes of $T = 600$ and $T = 1200$. In the setting with normally distributed factors, we observe the Bayesian methodology to be least accurate when we simulate the DGP I-III-IV-VII-IIX-IX with a sample size of $T = 300$. In this setting, the selection accuracy of the methodology is 41% and the factor model, as implied by the true DGP, is ranked 65 times out of a total of $Z = 100$ simulation iterations among the top-5 models with the highest marginal likelihoods. In the setting with normally distributed factors, we observe the selection accuracy of the Bayesian methodology to range from 75% to 95% and from 90% to 98% across the simulated DGP's, when considering sample sizes of $T = 600$ and $T = 1200$, respectively.

Table 4: Simulation Study Results Bayesian Factor Selection Methodology II

Priced Factors	$T = 300$		$T = 600$		$T = 1200$	
	$\nu = 5$	$\nu = \infty$	$\nu = 5$	$\nu = \infty$	$\nu = 5$	$\nu = \infty$
I-V	67 (88)	78 (93)	78 (98)	89 (99)	92 (97)	96 (100)
I-IX	66 (86)	80 (96)	75 (95)	95 (100)	90 (99)	93 (99)
I-VI	26 (54)	49 (80)	57 (82)	83 (97)	80 (100)	90 (100)
I-IIX	65 (92)	88 (95)	82 (96)	86 (100)	91 (99)	89 (99)
I-II-V	48 (74)	68 (89)	78 (99)	89 (100)	94 (100)	94 (100)
I-II-IIX	79 (89)	89 (98)	91 (99)	89 (99)	93 (99)	91 (100)
I-III-VII	32 (60)	45 (75)	69 (92)	79 (97)	95 (100)	90 (100)
I-IIX-IX	58 (78)	82 (91)	84 (95)	87 (96)	91 (99)	97 (100)
I-II-III-IIX	31 (77)	49 (92)	62 (96)	82 (96)	83 (98)	90 (100)
I-II-IV-IIX	57 (84)	62 (91)	82 (99)	90 (99)	81 (99)	98 (100)
I-II-VI-IIX	32 (80)	57 (91)	67 (92)	84 (97)	91 (99)	94 (99)
I-IV-VII-IIX	71 (85)	81 (90)	88 (95)	94 (99)	93 (99)	94 (100)
I-III-IV-VII-IIX-IX	20 (51)	41 (65)	62 (83)	77 (97)	86 (97)	97 (100)

Simulation results Bayesian factor selection methodology (with $\tau = 20$). The left hand side of the table displays sets of true priced risk factors implied by the DGP's we simulate. Each DGP is simulated $Z = 100$ times. The numerical entries displayed in plain text and within (**parentheses**) respectively give the percentage of times the true priced risk factors are correctly identified and the percentage of times the factor model, as implied by the true DGP, ranks among the top five models with the highest marginal likelihoods. Results are displayed for various sample sizes T and d.o.f. of the student-t distribution ν.

The Bayesian factor selection methodology exhibits adequate robustness in a setting with student-t (with $\nu = 5$ d.o.f.), as opposed to normally, distributed factors. The largest loss in selection accuracy resulting from simulating student-t, as opposed to normally, distributed factors is observed when we simulate the DGP I-III-IV-VII-IIX-IX with a sample size of $T = 300$. We observe an accuracy loss of 50% in this setting. As the sample size T increases, accuracy losses resulting from simulating student-t, as opposed to normally, distributed factors fade. The accuracy loss resulting from simulating student-t distributed factors decreases from 50% to 11% when we simulate DGP I-III-IV-VII-IIX-IX with sample size $T = 1200$, as opposed to $T = 300$.

Furthermore, as the sample size T increases, the percentage of times the true risk factors are identified seems to converge to 100%, for each simulated DGP, in settings with normally, as well as student-t, distributed factors.

In addition to investigating the performance of the Bayesian methodology in terms of selection accuracy, we may investigate the economic loss that results from selecting a wrong model, i.e. selecting a model that is not equivalent to the model implied by the true DGP. This of interest as, judging from the results displayed in Table 4, the Bayesian methodology will not attain a perfect selection accuracy for relevant sample sizes. We assume investors are mean-variance investors that seek to minimize portfolio variance given a target portfolio mean, and quantify the economic loss resulting from selecting a wrong model as the percentage loss in Sharpe ratio resulting from constructing the mean-variance efficient portfolio with selected priced risk factors as opposed to true (DGP implied) priced risk factors. Mean-variance efficient portfolios may be constructed using either true moments, usually unknown in practice, or estimated moments. We calculate Sharpe ratio's of constructed portfolios using true moments. Table 14 (Appendix B) displays the average percentage loss in Sharpe ratio, given that the Bayesian methodology (with $\tau = 20$) selects the wrong model, resulting from constructing the mean-variance efficient portfolio with selected, as opposed to true, priced risk factors, across the simulated DGP's. Percentage losses displayed in plain text and in (**parentheses**) are calculated in a setting where portfolio weights are constructed with, respectively, true and estimated moments.

In a setting where portfolio weights are constructed using true moments and $T = 300$, the average percentage loss in Sharpe ratio ranges from 15% to 35% and from 5% to 32% across the DGP's, when we set $\nu = 5$ and $\nu = \infty$, respectively. As we increase the sample size T to $T = 1200$ (and calculate portfolio weights using true moments), the average percentage loss in Sharpe ratio decreases to 0% across most of the DGP's. Thus, when sample size is sufficiently large, we can expect the Bayesian methodology to identify all priced risk factors with (approximately) 100% probability, even when the Bayesian methodology is not expected to attain perfect model selection accuracy at the relevant sample size. A set of selected factors that contains all true priced risk factors will still be wrong, in the sense that it does not equal the true set of priced risk factors as implied by the DGP, when the set of selected factors contains redundant non-risk factors. When a set of selected risk factors is wrong the true mean-variance efficient portfolio may still be constructed using the set of selected factors with 100% probability if and only if the set of selected factors contains all true priced risk factors, in addition to redundant non-risk factors, with 100% probability. Thus, given that a selected model is wrong, but the incurred percentage loss in Sharpe ratio equals 0% with 100% probability, the set of selected factors must contain all true priced risk factors, in addition to non-risk factors, with 100% probability.

The average percentage loss in Sharpe ratio increases across all DGP's when portfolio weights are constructed using estimated, as opposed to true, moments, even for the largest sample size $T = 1200$. This reflects the fact that, in a setting where portfolio weights are constructed using estimated, as opposed to true, moments a loss in Sharpe ratio may be incurred when the set of selected factors contains redundant non-risk factors, in addition to the true set of priced risk factors, as the selection of additional redundant factors increases the difficulty of moment estimation, thereby effectively increasing the difficulty of estimating the true weights of the mean-variance efficient portfolio.

Table 5: Simulation Study Results GMM Based Factor Selection Methodology

Priced Factors	$T = 300$				$T = 1200$			
	N_1	N_2	N_3	N_4	N_1	N_2	N_3	N_4
I-V	2	7	11	8	8	12	11	18
I-IX	5	7	8	7	8	13	20	24
I-VI	3	2	5	1	5	20	10	10
I-IIX	12	19	37	22	35	41	40	44
I-II-V	2	4	16	22	13	10	32	22
I-II-IIX	6	19	47	55	23	47	62	58
I-III-VII	0	2	1	0	1	3	5	3
I-IIX-IX	4	12	15	13	6	18	27	22
I-II-III-IIX	2	7	13	18	3	13	24	25
I-II-IV-IIX	3	10	20	20	6	30	34	26
I-II-VI-IIX	1	3	13	5	4	13	30	16
I-IV-VII-IIX	4	9	20	25	14	24	38	35
I-III-IV-VII-IIX-IX	0	0	4	6	4	7	24	8

Simulation results GMM factor selection methodology. Sets of true priced risk factors implied by the DGP's we simulate are displayed on the left hand side. Each DGP is simulated $Z = 100$ times. Numerical entries give the percentage of times the true priced risk factors are correctly identified. Results are displayed for various sample sizes T, various test-assets set sizes $N_1 = 20$, $N_2 = 60$, $N_3 = 100$ and $N_4 = 160$, and $\nu = 5$ d.o.f. of the student-t distribution.

We turn our discussion to the simulation study results of the GMM based factor selection methodology of Cochrane (2005). Table 5 displays simulation results of the GMM based factor selection methodology, in a setting with student-t ($\nu = 5$) distributed factors and test-assets. For each simulated DGP, we display the percentage of times the true priced risk factors are correctly identified. We find the selection accuracy of the GMM based methodology to be wanting. For the sample size $T = 300$, selection accuracy of the GMM based methodology ranges from 0% to 12%, from 0% to 19%, from 1% to 47% and from 0% to 55% across the various simulated DGP's, when considering the test-assets set sizes of N_1, N_2, N_3 and N_4, respectively. The overall selection accuracy of the methodology increases when one of the larger test-assets set sizes, N_3 or N_4, is used as opposed to one of the smaller test-assets set sizes, N_1 and N_2. Using the largest test-assets set size, N_4, and sample size $T = 300$, we observe a selection accuracy smaller than 25% for 11 out of the 13 simulated DGP's. This in sharp contrast with the observed selection accuracy of the Bayesian methodology (with $\tau = 20$) in the setting with student-t distributed factors and sample size $T = 300$. We observe the selection accuracy of the Bayesian methodology to be larger than 25% for all but a single one of the simulated DGP's in this setting.

When we increse the sample size to $T = 1200$, we observe an increase in the selection accuracy of the GMM based methodology, for all relevant test-assets set sizes. The GMM based methodology does not achieve the same levels of selection accuracy the Bayesian methodology (with $\tau = 20$) achieves when we use the larger sample size of $T = 1200$ (in the setting with student-t distributed factors and test-assets) though. Using one of the larger test-assets set sizes N_3 or N_4, the selection accuracy of the GMM based methodology ranges, respectively, from 5% to 62% or from 3% to 58% across the simulated DGP's. The selection accuracy of the Bayesian methodology (with $\tau = 20$) ranges from 86% to 95% across simulated DGP's. Furthermore, in the setting with student-t distributed factors and test-assets, for both sample sizes $T = 300$ and $T = 1200$, we observe that the Bayesian methodology (with $\tau = 20$) attains a higher selection accuracy across all the simulated DGP's, as compared to the GMM methodology, regardless of which of our test-assets set sizes is relevant.

Table 15, see Appendix B, displays simulation results of the GMM based factor selection methodology, in a setting with normally distributed factors and test-assets. Although the GMM based methodology does not rely on assumptions about the distributions of the test-assets and candidate factors, we observe the GMM based methodology attains a higher overall selection accuracy across the simulated DGP's when test-assets and candidate factors follow normal, as opposed to student-t, distributions. We still find the selection accuracy of the GMM based methodology, as compared to the Bayesian methodology (with $\tau = 20$), to be wanting in this setting, however.

As discussed in section 4.1, we only simulate a DGP $\mathcal{D}_j$ in case all the estimates of the parameters in parameter vector $b_j = [b_j^1, ..., b_j^{K_j}]'$ of the model implied by DGP $\mathcal{D}_j$ differ significantly from 0 (we use a significance level of 1%). In that case, we can reject the null $H_0 : b_j^i = 0$, for all $i = 1, ..., K_j$. In case we cannot reject the null for a particular $i = 1, ..., K_j$, either the respective parameter b_j^i is actually zero, or the data does not provide enough support to reject the null, although the respective parameter b_j^i is actually non-zero. In the former case, DGP $\mathcal{D}_j$ is invalid, as discussed in section 4.1. In the latter case, the DGP $\mathcal{D}_j$ is actually valid, but factor selection methodologies might have trouble identifying the priced risk factor associated to b_j^i when simulating the DGP with a sample size that is sufficiently small.

We proceed the discussion with an example. We estimate the parameters of the model implied by the DGP with WMKT-RMW-Dollar (I-V-X) as the set of true priced risk factors, Table 6 displays the estimated parameter b of the model implied by the DGP. We assume that factors and test-assets are normally distributed, and use test-assets set size N_4. The estimated direct effects of WMKT and RMW on the SDF are significant, using a 1% significance level, but the estimated direct effect of Dollar on the SDF is not. Even at a significance level of 10%, the estimated direct effect of Dollar remains insignificant. It might well be the case that the Dollar factor does not have a direct effect on the SDF. It might also be the case that the Dollar factor actually has a direct effect on the SDF, but that the effect is too subtle, in the sense that the data does not provide sufficient evidence for the actual presence of the effect.

Table 6: Estimated parameter b of the model implied by DGP: WMKT-RMW-Dollar

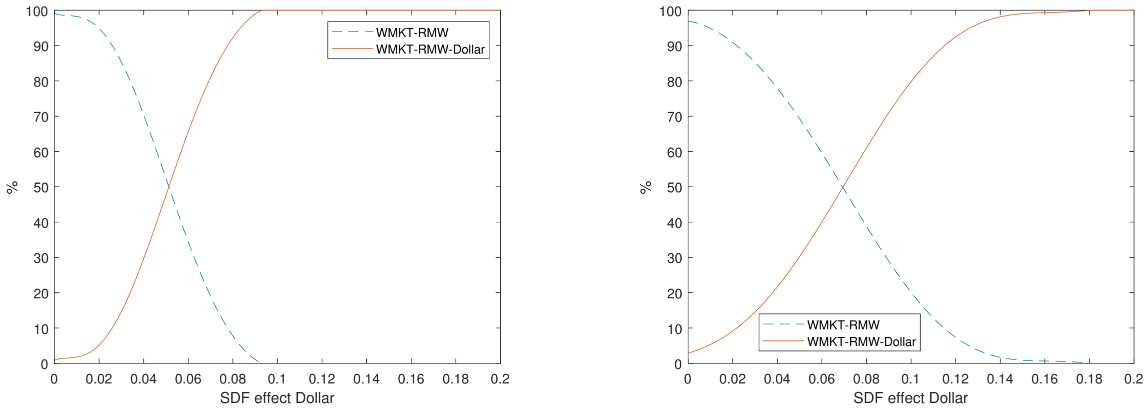
Factor	WMKT	RMW	Dollar
$\hat{b}$	0.071	0.240	0.018
	(0.018)	(0.046)	(0.034)

ML estimate of the parameter b of the model implied by the DGP with WMKT-RMW-Dollar as the set of true priced risk factors. Standard errors are displayed in parentheses. Factors and test-assets are assumed to be normally distributed.

We proceed by simulating the DGP with WMKT-RMW-Dollar as the set of true priced risk factors, using sample sizes $T = 1200$ and $T = 12000$. Factors and test-assets are normally distributed, we use test-assets set size N_4. We manually set the parameter governing the direct effect of Dollar on the SDF, denoted by b^{Dlr}, at various values in the interval [0 0.2]. For each value of b^{Dlr} considered, we simulate the respective DGP $Z = 100$ times. Figures 2 and 6 (Appendix B) display, for both the Bayesian (with $\tau = 20$) and the GMM based factor selection methodologies, for values of b^{Dlr} in interval [0 0.2], the percentage of times WMKT and RMW are selected, as well as the percentage of times WMKT, RMW and Dollar are selected. The factor selection methodologies are applied in a setting where the set of candidate risk factors solely consists of the factors WMKT, RMW and Dollar.

In case b^{Dlr} is set $b^{Dlr} = 0$, the DGP is invalid in the sense that the implied priced risk factor Dollar has no direct effect on the SDF. The Bayesian and GMM based methodologies only identify WMKT and RMW as priced factors, in all simulation iterations, in that case, as observed in Figures 2 and 6. In case b^{Dlr} is set at a value in the interval (0 0.2] (for example b^{Dlr} is set at the ML estimate $b^{Dlr} = 0.018$), the DGP is valid in the sense that the implied priced risk factor Dollar has a direct effect on the SDF. Judging from Figures 2 and 6, the Bayesian and GMM based methodologies have trouble identifying the direct effect of Dollar on the SDF though, for the smaller values of b^{Dlr} in the interval (0 0.2]. For the sample size $T = 1200$, the Bayesian and the GMM based methodologies respectively fail to identify the Dollar factor as a priced risk factor in all simulation iterations when b^{Dlr} is set $b^{Dlr} < 0.1$ and $b^{Dlr} < 0.16$. Although the selection accuracy of the Bayesian as well as the GMM based methodologies increases as we increase the sample size to an epic size of $T = 12000$, the methodologies still fail to consistently identify Dollar as a priced risk factor in the simulation iterations for small enough values of b^{Dlr} in the interval (0 0.2]. The Bayesian and GMM based methodologies may thus fail to identify a priced risk factor when its direct effect on the SDF is too subtle, in the sense that (simulated) data fails to provide sufficient evidence for the presence of the effect.

Figure 2: Simulation Results DGP: WMKT-RMW-Dollar, $T = 1200$

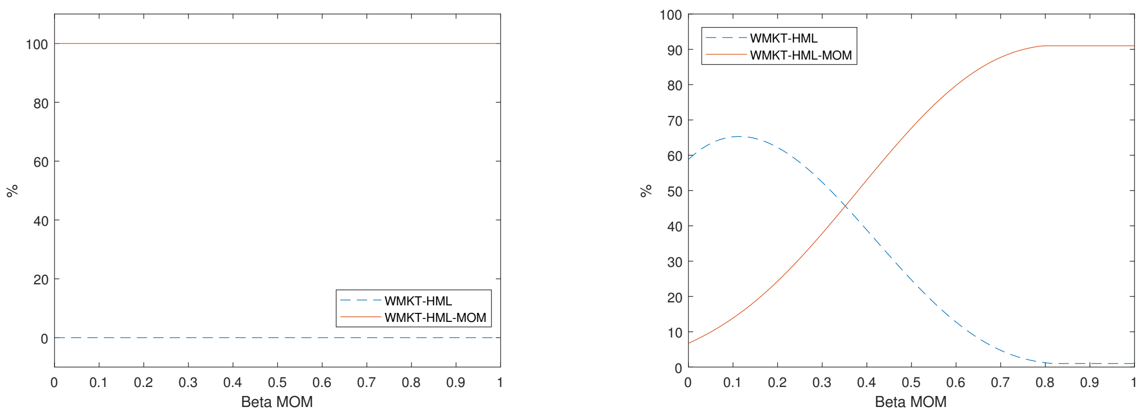


We simulate the DGP: WMKT-RMW-Dollar, with normally distributed factors and test-assets (we use N_4 test-assets and sample size $T = 1200$), $Z = 100$ times, for various values of b^{Dlr} (the direct effect of Dollar on the SDF). We use the Bayesian (with $\tau = 20$) and GMM based factor selection methodologies to select priced risk factors each simulation iteration. The set of candidate factors solely consists of WMKT, RMW and Dollar. For each relevant value of b^{Dlr}, the percentage of times WMKT and RMW are selected as well as the percentage of times WMKT, RMW and Dollar are selected are displayed. On the left: Bayesian methodology results. On the right: GMM based methodology results.

As discussed in section 4.1, we expect test-asset based factor selection methodologies, such as the GMM based methodology, to be unable to accurately identify a priced risk factor, when it is unrelated to the relevant set of test-assets. We illustrate this discussion with an example. We estimate the parameters of the model implied by the DGP with WMKT-HML-MOM (I-III-VII) as the set of true priced risk factors. We assume that factors and test-assets are normally distributed, and use test-assets set size N_1. Estimated direct effects of factors WMKT, HML and MOM on the SDF are all significant, at the 1% level. Let β^{WMKT}, β^{HML} and β^{MOM} denote $N_1 \times 1$ parameter vectors of the model implied by the DGP that govern the direct effects of, respectively, WMKT, HML and MOM on the N_1 test-assets. We test $H_0 : \beta^{WMKT} = 0$, $H_0 : \beta^{HML} = 0$ and $H_0 : \beta^{MOM} = 0$ with F-tests, and do not reject $H_0 : \beta^{MOM} = 0$ at a 10% significance level. We reject the null hypotheses $H_0 : \beta^{WMKT} = 0$ and $H_0 : \beta^{HML} = 0$ at a 1% significance level. In case $\beta^{MOM} = 0$, the MOM factor is a priced risk factor, as it has a direct effect on the SDF, but is unrelated to the relevant set of N_1 test-assets.

We simulate the DGP with WMKT-HML-MOM as the set of true priced risk factors, using sample sizes $T = 1200$ and $T = 12000$. Factors and test-assets are normally distributed, we use test-assets set size N_1. We set parameter $\beta^{MOM} = c\boldsymbol{\iota}$, where $\boldsymbol{\iota}$ denotes a $N_1 \times 1$ vector of ones and c denotes a constant, and consider various values for c in the interval [0 1]. For each value of c considered, we simulate the respective DGP $Z = 100$ times. Figure 3 and 7 (Appendix B) display, for both the Bayesian (with $\tau = 20$) and the GMM based factor selection methodologies, for values of c in interval [0 1], the percentage of times WMKT and HML are selected, as well as the percentage of times WMKT, HML and MOM are selected. The factor selection methodologies are applied in a setting where the set of candidate factors solely consists of WMKT, HML and MOM.

Figure 3: Simulation Results DGP: WMKT-HML-MOM, $T = 1200$



We simulate the DGP: WMKT-HML-MOM, with normally distributed factors and test-assets (we use N_1 test-assets and sample size $T = 1200$), $Z = 100$ times, for various values c, the constant governing $\beta^{MOM} = c\boldsymbol{\iota}$. We use the Bayesian (with $\tau = 20$) and GMM based factor selection methodologies to select priced risk factors each simulation iteration. The set of candidate factors solely consists of WMKT, HML and MOM. For each relevant value of c, the percentage of times WMKT and HML are selected as well as the percentage of times WMKT, HML and MOM are selected are displayed. On the left: Bayesian methodology results. On the right: GMM based methodology results.

The GMM based methodology fails to select all three factors WMKT, HML and MOM in more than 90% of the simulation iterations in case $\beta^{MOM} = 0$, even when we use an epic sample size of $T = 12000$. As we increase the direct effect of MOM on the test-assets, the selection accuracy of the GMM based methodology increases. However, when the direct effect of MOM on the test-assets is sufficiently subtle, in the sense that (simulated) data does not provide sufficient evidence for the presence of the effect, we still find the selection accuracy of the GMM based methodology to be wanting. For example, we observe selection accuracies of 10% and 30% for sample sizes $T = 1200$ and $T = 12000$, respectively, in case $\beta^{MOM} = 0.1\boldsymbol{\iota}$. The Bayesian methodology does not depend on the use of test-asset data, and does not suffer from a loss in selection accuracy when MOM is unrelated, or only subtly related, to our set of N_1 test-assets. For both sample sizes $T = 1200$ and $T = 12000$, the Bayesian methodology attains a selection accuracy of 100%, regardless of the strength of the effect of MOM on the N_1 test-assets.

5 Empirical Study

We apply our factor selection methodologies to select priced risk factors from our set of candidate factors discussed in section 3, and evaluate the pricing performance of candidate factor models.

5.1 Factor Model Selection

Table 7 displays our set of $H = 13$ candidate factors considered in the empirical study, each factor is assigned to a roman numeral for ease of reference.

Table 7: Candidate Factors Empirical Study

W	L	SMB	HML	DHML	RMW	CMA	MOM	QMJ	BAB	Dllr	Crry	GT
I	II	III	IV	V	VI	VII	IIX	IX	X	XI	XII	XIII

Roman Numerals Corresponding To Factors. WMKT is abbreviated with W and is assigned to roman numeral I. LWMKT is abbreviated with L and is assigned to roman numeral II. The Dollar, Carry and Global Tail factors are respectively abbreviated with Dllr, Crry and GT.

Using our set of $H = 13$ candidate factors, a total of $J = 2^{13} - 1 = 8191$ candidate factor models of the form specified in Eq. (11) can be constructed. We apply the Bayesian (marginal likelihood based) factor selection methodology to filter out priced risk factors from our set of candidate factors, and, following our discussion in section 4.2, specify hyper-parameter k, as

$$k = (c\text{ShMax})^2/H, \quad c\text{ShMax} = \tau Sh(WMKT), \quad \tau = 20,$$

with ShMax denoting the maximum attainable Sharpe ratio, and Sh(WMKT) denoting the sample Sharpe ratio over the WMKT portfolio.

We use the Bayesian factor selection methodology to identify factor models with the highest posterior probabilities (we use equal prior weights) out of the set of J candidate factor models. The upper part of Table 8 displays the top-8 models, as ranked by their posterior probabilities. We refer to the factor model with rank x, as ranked by posterior probability, as (factor) model x, in our following discussion, for ease of reference. The posterior probability of model 1 equals 23%, and is 2 times higher than the posterior probability of model 2 and up to 10 times higher than the posterior probabilities of the other top-8 models displayed in Table 8.

Table 8: Posterior Model Probabilities and Cumulative Factor Probabilities

Priced Factors	Posterior Prob.	Priced Factors	Posterior Prob.
1: I-III-IV-V-IIX-IX	0.230	5: I-III-IV-V-IIX-IX-XIII	0.031
2: I-III-V-IIX-IX	0.111	6: I-III-V-VI-IIX-IX	0.028
3: I-III-IV-V-VI-IIX-IX	0.088	7: II-III-IV-V-IIX-IX-XI-XII	0.027
4: II-III-IV-V-IIX-IX-XI	0.048	8: I-III-IV-V-VII-IIX-IX	0.022

Cumulative Factor Probabilities

I	II	III	IV	V	VI	VII	IIX	IX	X	XI	XII	XIII
0.76	0.28	1.00	0.68	0.99	0.26	0.09	0.94	1.00	0.01	0.22	0.19	0.09

We use the Bayesian factor selection methodology (with $\tau = 20$) to identify the models with the highest posterior probabilities out of the set of total potential models that can be constructed using our set of H candidate factors. The upper part of the table displays the top-8 models as ranked by their posterior probabilities. For each of these models, we display the posterior probability, as well as the set of factors, that are proposed (by the respective model) to be priced risk factors. The lower part of the table displays cumulative factor probabilities. The cumulative probability of a factor is defined as the sum of posterior probabilities of models that include the factor.

The lower part of Table 8 displays cumulative factor probabilities. The cumulative factor probability is defined as the sum of posterior probabilities of models that include the factor. The cumulative probability of a factor can be interpreted as the posterior probability that the factor is a priced risk factor. We compute the posterior probability that either one, but not both, of the WMKT or LWMKT factors is a (are) priced risk factor(s) as the sum of posterior probabilities of models that include either one, but not both, of the factors and find it to equal 96%. This finding is in line with our expectation that one, but not both, of the global market factors is a (are) priced risk factor(s). The cumulative factor probabilities of SMB, DHML, QMJ and MOM are close to, or equal to, 100%, indicating that these factors are very likely to be priced risk factors, in addition to one of the global market factors, as well. These are precisely the factors that, in addition to WMKT and HML, make up the factor model with the highest posterior probability. The posterior probability that both HML and DHML are priced risk factors, computed as the sum of posterior probabilities of models that include both factors, is 68%, which is surprising as both factors aim to capture the value effect. As the sample correlation of HML and DHML (Table 11) equals 0.68, the factors may carry different information related to the value effect.

The question remains as to which factors should be selected as priced risk factors. We could simply select the factors that make up model 1 as priced risk factors, but judging from our discussion in section 4, we have no guarantee that the set of factors that make up the model with the highest posterior probability will equal the set of true priced risk factors. Therefore, it seems prudent to consider multiple potential sets of priced risk factors. For example, we could consider eight potential sets of priced risk factors, where each factor model in the top-8 (in terms of posterior probability) is made up by one of the potential sets of factors.

Table 9: GRS Tests Nested Factor Models

Model Pair: Nested vs. General	RW
I-III-V-IIX-IX vs. I-III-IV-V-IIX-IX	0.46
I-III-V-IIX-IX vs. I-III-IV-V-VI-IIX-IX	0.00
I-III-V-IIX-IX vs. I-III-IV-V-IIX-IX-XIII	4.57
I-III-V-IIX-IX vs. I-III-V-VI-IIX-IX	2.74
I-III-V-IIX-IX vs. I-III-IV-V-VII-IIX-IX	0.00
II-III-IV-V-IIX-IX-XI vs. II-III-IV-V-IIX-IX-XI-XII	37.00
II-III-V-IIX-IX-XI-XII vs. II-III-IV-V-IIX-IX-XI-XII	0.46
II-III-V-IIX-IX vs. II-III-V-IIX-IX-XI-XII	91.00

We test, for each listed pair of factor models (each pair contains one nested and one general model), whether the nested factor model prices factors that are excluded from the nested model, but included in the general model, by the means of rolling window GRS tests. For each pair of factor models, RW reports the share of 60-month rolling windows where the GRS test rejects the null that the nested factor model prices the excluded factors (expressed in percentages). We use a significance level of 5%.

We observe that many of the high ranking factor models (as ranked by posterior probability) are actually nested in other high ranking factor models. For example, factor model 2 is nested in models 1, 3, 5, 6 and 8, while factor model 4 is nested in model 7. As discussed in section 2.3.1, a factor model that is nested in a more general model is superior to the general model if the nested factor model actually prices excluded factors that are included in the general model. We proceed to test whether factor model 2 outperforms the more general models 1, 3, 5, 6 and 8, and whether factor model 4 outperforms the more general model 7. Table 9 reports results of rolling window GRS tests. We test, in a conditional setting, for each listed pair of factor models (each pair contains one nested and one general model), whether the nested factor model prices factors that are excluded from the nested model, but included in the general model, by the means of rolling window GRS tests. For each pair of factor models, we report the share of 60-month rolling windows where the GRS test rejects the null that the nested factor model prices the excluded factors. A significance level of 5% is used. We use a conditional setting and 60-month rolling windows to account for potential time-variation of factor loadings and risk premia, in the spirit of Brusa et al. (2014) and Lewellen and Nagel (2006).

Judging from our rolling window (RW) GRS test results, factor model 2 likely dominates each of the more general models 1, 3, 5, 6 and 8 in a conditional setting. For each of the more general models of interest, we are able to reject the null that factor model 2 prices excluded factors, in less than 5% of the rolling windows. On the other hand, in a conditional setting, factor model 4 does not seem to adequately price factor Carry (XII), which is included in more general model 7. In 37% of the rolling windows, we reject the null that factor model 4 prices Carry (XII). Factor model 7 is therefore likely to be superior to factor model 4. We also test whether a nested version of factor model 7, II-III-V-IIX-IX-XI-XII, which excludes HML (IV), is able to (conditionally) price HML (IV). RW GRS test results indicate the restricted version of model 7 is indeed able to (condtionally) price HML (IV), as we are able to reject the null that the restricted version of model 7 prices HML (IV) in less than 1% of the rolling windows.

In addition to the Bayesian factor selection methodology, we may apply the GMM based methodology to select priced risk factors out of our set of $H = 13$ candidate factors. Results based on the GMM based methodology may well be unreliable though, as we found the selection accuracy of the methodology to be wanting in our simulation study (section 4). Table 16 (Appendix B) displays, using each of our four distinct sets of test assets (discussed in section 3.2), the factors that are selected by the GMM based methodology as priced risk factors. Only when we use our aggregate set of $N_4 = 160$ test-assets, results seem plausible, and the GMM methodology selects II, III, V, IIX and IX to be priced risk factors. The factor model II-III-V-IIX-IX is nested in model II-III-V-IIX-IX-XI-XII, but is unable to adequately (conditionally) price excluded factors XI and XII, judging from the RW GRS test result displayed in Table 9.

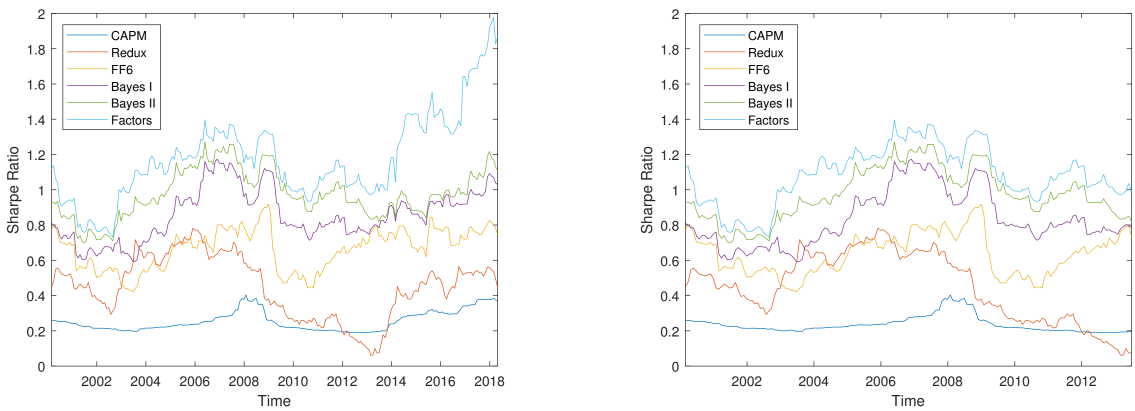
We end up with two potential sets of priced risk factors. The first one being I-III-V-IIX-IX, and the second one being II-III-V-IIX-IX-XI-XII. The first set contains the (global) market factor WMKT and all the factors with cumulative probabilities of (approximately) 100%: SMB, DHML, MOM and QMJ. The second set contains, in addition to the factors with cumulative probabilities of (approximately) 100%, the factors that make up the CAPM Redux model of Brusa et al. (2014): (global) market factor LWMKT and currency factors Dollar and Carry. We refer to the factor model made up by the factors in the first set and the factor model made up by the factors in the second set as the Bayes I model and the Bayes II model, respectively.

5.2 Evaluating Factor Models

We proceed to evaluate the pricing performance of candidate factor models. Our set of candidate factor models consists of the Bayes I and Bayes II models, and three prominent factor models proposed in the literature: the Word CAPM model, the (International) CAPM Redux model of Brusa et al. (2014) and a 6 factor model based on the WMKT factor, the 4 (global) factors of Fama and French (2015) and the (global) momentum factor of Carhart (1997), which we refer to as the FF6 model. We thus consider a total of five candidate factor models, or five candidate (priced risk) factor sets. First, we compare the relative pricing performance of our candidate factor models. As discussed in section 2.3, we should compare the relative pricing performance of factor models by evaluating the pricing performance of the models with respect to (excluded) factors. Second, we evaluate the pricing performance of the candidate factor models in an absolute setting. We investigate whether the factor models fully explain the cross-section of expected excess-returns of test-assets and (excluded) factors.

To gain insight into the relative pricing performance of the candidate factor models, we evaluate, for each candidate factor set, the maximum sample Sharpe ratio (over the portfolio) of the factors. In case a factor model prices all excluded factors, the maximum Sharpe ratio of all $H = 13$ candidate factors will equal the maximum Sharpe ratio of the set of factors that make up the respective factor model. Otherwise, the maximum Sharpe ratio of all factors will be higher. The maximum Sharpe ratio of the factors that make up the candidate factor model with the highest pricing performance w.r.t. excluded factors, will be closest to the maximum Sharpe ratio of all factors, as compared to the maximum Sharpe ratios of the other candidate factor sets. To account for potential time-variation of means and variances, we use a 60-month rolling window (RW) to construct sample Sharpe-ratio's at each time point in our sample period. For each candidate factor set, as well as the complete set of candidate factors, Figure 4 displays the evolution of the RW maximum sample Sharpe ratio over time.

Figure 4: Rolling Window Sharpe Ratio's



The Figure displays, for each candidate factor set (or model), the evolution of the 60-month rolling window (RW) maximum sample Sharpe ratio (over the portfolio) of the respective candidate factors over time. "Factors" displays the evolution of the RW maximum sample Sharpe ratio of all candidate factors. We display the evolution of the RW maximum sample Sharpe ratio's in the period February 2000 - April 2018 and sub-period February 2000 - May 2013 on the left and right hand side, respectively.

We observe that the RW maximum (sample) Sharpe ratio (over the portfolio of the factors) of the Bayes II model closely matches the RW maximum Sharpe ratio of all factors during the whole time period February 2000 - May 2013. The RW maximum Sharpe ratio of the Bayes I model does not match the RW maximum Sharpe ratio of all factors as closely as the RW maximum Sharpe ratio of the Bayes II model does during the time period February 2000 - May 2013, but clearly matches the RW maximum Sharpe ratio of all factors closer than the RW maximum Sharpe ratio's of the CAPM, CAPM Redux and FF6 models during the whole length of the time period. During the time period May 2013 - April 2018, we observe that the RW maximum Sharpe ratio of all factors starts to increase over time and that it diverges from the

RW maximum Sharpe ratio's of the candidate factor models over time. During the whole time period May 2013 - April 2018, the RW maximum Sharpe ratio's of the Bayes I and Bayes II models still outperform the RW maximum Sharpe ratio's of the other candidate factor models though. The results suggest that, in a conditional setting, as compared to the other candidate models, the Bayes II model performs best in terms of relative pricing performance, followed by the Bayes I model. Furthermore, it may well be the case that, in a conditional setting, the Bayes II factor model prices all excluded factors without any pricing error in the time period February 2000 - May 2013. In the time period May 2013 - April 2018 this is likely not the case, as the RW maximum Sharpe ratio of all factors is up to twice as large as the RW maximum Sharpe ratio of the Bayes II model during the time period.

Table 10: Results GRS Tests

Model	ExF	Test-Assets						
		Market	Value	Growth	MEBM	MEINV	MEMOM	MEOP
I: Rolling Window								
CAPM	99	95	90	91	95	89	99	90
Redux	94	91	83	85	91	82	99	80
FF6	84	75	67	57	77	63	96	79
Bayes I	37	34	42	20	53	43	91	56
Bayes II	25	23	32	13	42	36	85	44
II: Full Sample								
CAPM	13.63	7.54	6.78	6.90	7.90	7.84	9.43	7.88
Redux	14.37	7.29	6.52	6.65	7.65	7.59	9.19	7.63
FF6	11.98	5.59	4.86	4.97	6.04	5.99	7.48	6.03
Bayes I	2.13*	2.69	2.10	2.19	3.36	3.32	4.51	3.35
Bayes II	1.66**	2.60	1.98	2.08	3.32	3.27	4.52	3.30

Part I of the Tables displays, for each candidate factor model and for several sets of test-assets, the share of 60-month rolling windows where the GRS test rejects the null that the factor model prices the set of test-assets (expressed in percentages) during the full sample period February 1995 - April 2018. A significance level of 5% is used. Part II of the Tables display GRS test statistics of full sample GRS tests. GRS test statistics with p-values higher than 5% and 10% are, respectively denoted with a $*$ and a $**$. For each candidate factor model, the test-asset set "ExF" refers to the set of factors that are excluded from the factor model (but included in our total set of H candidate factors). To combat the the small T versus large N problem resulting from using 60-month rolling windows, we split our set of all $N_4 = 160$ test assets up into seven smaller test-asset sets. The Market, Value and Growth sets respectively consist of the country market indices, country growth indices and country value indices discussed in section 3.2. The MEBM, MEINV, MEMOM and MEOP sets respectively consist of the global portfolios formed by bi-variate sorts on ME and BE/ME, bi-variate sorts on ME and INV, bi-variate sorts on ME and MOM and bi-variate sorts on ME and OP, as discussed in section 3.2. For each factor model, the set of excluded factors "ExF" is included in all the sets of our test-assets.

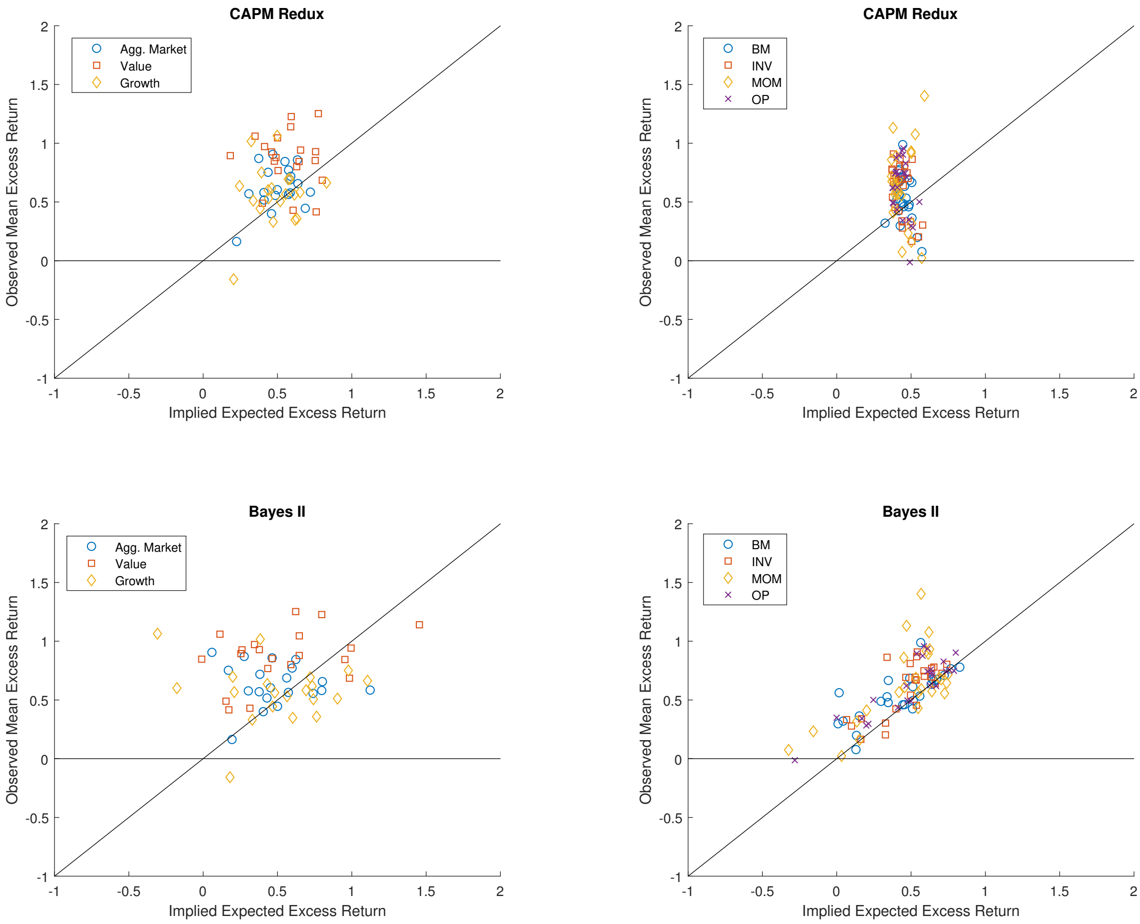
We proceed to formally test whether our candidate factor models price all (excluded) factors and test-assets by the means of full-sample GRS tests as well as rolling window GRS tests. Rolling window GRS tests allow us to account for potential time-variation of factor loadings and risk premia. Part I of Tables 10 and 17 (Appendix B) displays, for each candidate factor model and for several sets of test-assets, the share of 60-month rolling windows where the GRS test rejects the null that the factor model prices the set of test-assets during, respectively, our full sample period and the sub-sample period February 1995 - May 2013. A significance level of 5% is used. Part II of the respective Tables displays GRS test statistics of full sample GRS tests. For each candidate factor model, the test-asset set "ExF" refers to the set of factors that are excluded from the factor model (but included in our total set of H candidate factors). To combat the small T versus large N problem resulting from using 60-month rolling windows, we split our aggregate set of $N_4 = 160$ test assets up into seven smaller test-asset sets. The Market, Value and Growth sets respectively consist of the Country Market indices, Country Growth indices and Country Value indices discussed in section 3.2. The MEBM, MEINV, MEMOM and MEOP sets respectively consist of the global portfolios formed by bi-variate sorts on ME and BE/ME, bi-variate sorts on ME and INV, bi-variate sorts on ME and MOM and bi-variate sorts on ME and OP, as discussed in section 3.2. For each factor model, the set of excluded factors "ExF" is included in all the sets of our test-assets, as a factor model should adequately price *all* assets, that is, test-assets *and* (excluded) factors.

In the full sample period, the share of rolling windows where the GRS test rejects the null that the excluded factors are priced equals 84%, 94%, 99%, 37% and 25% for, respectively, the FF6 model, the CAPM Redux model, the (World) CAPM model and the Bayes I and Bayes II models. In the sub-sample period February 1995 - May 2013, the rolling window GRS test results for excluded factors are similar for the CAPM, CAPM Redux and FF6 models, as compared to the full sample period. As compared to the full sample period, the Bayes II and Bayes I models perform better in the sub-sample though, in the sense that the share of rolling windows where the GRS test rejects the null that the excluded factors are priced is much lower in the sub-sample than in the full-sample, for both models. In the sub-sample, the share of rolling windows where the GRS test rejects the null that the Bayes II model prices the excluded factors is only 1%. These results are in line with our discussion of the rolling window maximum sample Sharpe ratio's of the candidate factor sets (models). The results of the rolling window GRS tests support the claim that, in a conditional setting, during the full-sample period, as compared to the other models, the pricing performance of the Bayes II model w.r.t. excluded factors is most satisfactory, followed by the Bayes I model. Although it is likely the case that the Bayes II model is not able to (conditionally) price the excluded factors without pricing error during the entire full-sample period, the rolling window GRS tests support the claim that it may well be the case that the Bayes II model (conditionally) prices the excluded factors without pricing error during the sub-sample period February 1995 - May 2013. When we consider full-sample GRS tests, we only fail to reject the null that excluded factors are adequately priced for the Bayes I and Bayes II models, when using a significance level of 5%. For all other models, we reject the null at the 5% significance level. This holds for the sample period February 1995 - April 2018 as well as the sub-sample period February 1995 - May 2013.

Our results clearly indicate that the Bayes II model performs best in terms of relative pricing performance, as compared to the other models, followed by the Bayes I model. We proceed the discussion by evaluating the pricing performance of our candidate factor models in an absolute setting. When considering full sample GRS test results (Tables 10 and 17), we reject the null that test-assets are priced for each of our candidate models, and each of our test-asset sets at a significance level of 5%. Judging from the results of the rolling window (RW) GRS tests (during our full sample period) displayed in Table 10, none of the factor models are likely able to (conditionally) price all of the seven sets of test-assets (all test-asset sets include excluded factors) during the entirety of our full sample period. For the CAPM, CAPM Redux and FF6 models, the share of rolling windows where the GRS test rejects the null that test-assets are priced is higher than 57%, for all sets of test-assets. These models are likely unable to (conditionally) price any of the sets of test-assets during the majority of the (or even during the whole) full sample period. Although the Bayes I and Bayes II models likely fail to (conditionally) price the MEMOM test-assets during the majority of the (or even during the whole) full sample period, as for both models the share of rejected RW GRS tests is higher than 85% for this set of test-assets, the Bayes I and Bayes II models may be able to (conditionally) price some of the other sets of test-assets during a (significant) sub-sample of the full-sample period. For all test-asset sets other than MEMOM, the share of rejected RW GRS tests is lower than 44% for the Bayes II model and lower than 56% for the Bayes I model. The Bayes II model is thus most likely to be able to (conditionally) price some of the test-asset sets, other than MEMOM, during a (significant) sub-sample of the full sample period.

To gain further insight into the pricing performance of the candidate factor models we plot, for each factor model, realized expected excess-returns of our entire cross-section of $N_4 = 160$ test-assets against the predicted expected excess-returns of the same test-assets as predicted by the factor model (we do not include excluded factors). We use 60-month rolling windows to estimate factor loadings of the factor models and to estimate (conditional) means of test-asset and factor excess-returns. Each rolling window, mean test-asset excess-returns are predicted, for each factor model, by multiplying estimated (conditional) factor loadings by the corresponding estimated (conditional) factor means. All rolling window predicted mean excess-returns and rolling window realized sample mean excess-returns are then averaged over the total number of rolling windows. Averaged realized mean excess-returns are then plotted against averaged predicted mean excess-returns, for each factor model. Figure 5 displays plots for the CAPM Redux and Bayes II models and Figure 8 (Appendix B) displays plots for the CAPM, FF6 and Bayes I models.

Figure 5: Realized versus Predicted Expected excess-returns



For each of the factor models, the Figure plots realized expected excess-returns of our entire cross-section of test-assets against the predicted expected excess-returns of the same test-assets as predicted by the candidate factor model (excluded factors are not included in the Figure). We use 60-month rolling windows to estimate factor loadings of the factor models and to estimate (conditional) means of test-asset and factor excess-returns. Each rolling window, mean test-asset excess-returns are predicted, for each of the factor models, by multiplying estimated (conditional) factor loadings by the corresponding estimated (conditional) factor means. All rolling window predicted mean excess-returns and rolling window realized sample mean excess-returns are then averaged over the total number of rolling windows. Averaged predicted mean excess-returns are plotted against averaged realized mean excess-returns, for each of the factor models.

Figures 5 and 8 allow us to visually evaluate the extent to which our candidate factor models are, on average, able to explain differences in the expected excess-returns across the cross-section of test-assets. A factor model is able to explain the differences in the expected excess-returns across the cross-section of test-assets, in case the difference between realized expected excess-returns of any pair of test-assets is properly matched by the difference between predicted expected excess-returns of the pair. Judging from Figures 5 and 8, on average, the (World) CAPM is barely able to explain any of the differences in the expected excess-returns across the cross-section of test-assets. Although the CAPM Redux, as opposed to CAPM, seems, on average, to be able to explain some of the differences in the expected excess-returns of the Country Market, Country Value and Country Growth indices, differences in the expected excess-returns of the MEBM, MEINV, MEMOM and MEOP portfolios are, on average, not explained by the CAPM Redux at all. When closely investigating the realized expected excess-returns of the Country Market indices, we observe that, on average, the FF6 explains almost none of the differences in these expected excess-returns. When the Country Market indices are concerned, the CAPM Redux thus seems, on average, to be better able to explain differences in expected excess-returns, as compared to FF6. When test-assets other than the Country Market indices are concerned, the FF6 model is, on average, better able to explain differences in expected excess-returns, as compared to the CAPM Redux. The Bayes I model seems to, on average, (conditionally) overprice most test-assets, as realized expected excess-returns of most test-assets are, on average, higher than corresponding predicted expected excess-returns. Overall, the Bayes II model seems, on average, best able to explain the differences in expected excess-returns across the cross-section of test-assets, as compared to the other models. Although the Bayes II model, on average, does not explain all the differences in expected excess-returns across the cross-section of test-assets, there is, on average, no particular set of test-assets with differences in expected excess-returns that are left, for the most part, unexplained by the model.

Judging from Figures 5 and 8, all candidate factor models (conditionally) price many of our test-assets with, on average, substantial pricing error. The results of rolling window GRS tests (Table 10) indicate that the Bayes I and Bayes II models may be able to (conditionally) price the Market, Value and Growth test-asset sets (including excluded factors), during a (significant) sub-sample of our full sample period. This conflicts with the (substantial) average pricing errors of the Bayes I and Bayes II models observed in Figures 5 and 8 (concerning the Country Market, Country Value and Country Growth indices). Therefore, concerning the Bayes I and Bayes II models, it is likely the case that, either, we are unable to reject a false null hypothesis of no pricing errors for the Market, Value and Growth test-asset sets during a substantial share of our rolling windows, or differences between the relevant predicted and realized expected excess-returns differ substantially during the rolling windows. Either way, our findings support the claim that none of our candidate factor models are able to (conditionally) price all test-assets during the entire full sample period.

6 Conclusion

The relevance for the identification of priced risk factors on the international level has increased tremendously in the last couple of decades, along with an increasing share of investors with foreign equity holdings and an increase in (global) candidate factors, as proposed by the literature, that help explain anomalies of the (World) CAPM. The quest to identify priced (excess-return) risk factors that fully explain the cross-section of asset expected excess-returns corresponds to the quest to find the mean-variance efficient portfolio (Huberman and Kandel (1987)). The identification of priced (excess-return) risk factors on the international level is therefore relevant from an explanatory viewpoint as well as from the viewpoint of an investor who aims to find the mean-variance efficient portfolio on the international level.

The recent research of Barillas and Shanken (2018) and Chib et al. (2018) provides a Bayesian, marginal likelihood based, factor selection methodology that enables us to filter out priced risk factors from a set of candidate (excess-return) factors. We extend upon the research of Barillas and Shanken (2018) and Chib et al. (2018), and argue that the specification of the priors of the unconditional means of proposed priced risk factors, or alpha's, (across the candidate models) implies a prior bias towards sparse factor models, as, conflicting with economic intuition, the prior means of the alpha's are set to equal zero. We argue that excessively narrow spreads of the priors of the alpha's may imply an excessive prior bias towards sparse models.

Indeed, in a simulation study, we find the precision of the marginal likelihood based factor selection methodology of Chib et al. (2018) and Barillas and Shanken (2018) to be wanting when the hyper-parameter k, governing the spreads of the priors of the alpha's, is set to equal the squared maximum (attainable) Sharpe ratio (over the portfolio) of the candidate factors, divided by the number of candidate factors, as proposed by Barillas and Shanken (2018). When hyper-parameter k is specified as suggested by Barillas and Shanken (2018), the marginal likelihood based factor selection methodology tends to excessively favour sparser factor models, as opposed to the true factor model as implied by the simulated DGP. The finding in turn suggests the specification of the priors of the alpha's as proposed by Barillas and Shanken (2018) implies a prior bias towards sparse factor models that is too excessive. We find we can substantially improve upon the precision of the marginal likelihood based factor selection methodology by setting hyper-parameter k, although conflicting with theoretical restrictions on the potential magnitude of k as derived by Barillas and Shanken (2018), equal to a *multiple* of the squared maximum Sharpe ratio of the candidate factors, divided by the number of candidate factors, thereby effectively increasing the spreads of the priors of the alpha's and decreasing the prior bias towards sparse factor models. Using our specification of the priors of the alpha's, we find the precision of the marginal likelihood based factor selection methodology to be robust in a setting with student-t, as opposed to normally, distributed factors and to be much more satisfactory than the precision of the GMM based factor selection methodology of Cochrane (2005).

It may well be prudent to further investigate the specification of the priors of the alpha's though, as our new specification of the priors that allows for larger spreads actually violates theoretical restrictions on the potential magnitude of hyper-parameter k. The key insight here is that we have to violate the theoretical restrictions in order to compensate for the (excessive) prior bias towards sparse factor models otherwise implied by setting the prior means of the alpha's to equal zero, which is in conflict with economic theory in the first place. Investors demand compensation for bearing risk, and we can a priori expect (absolute values) of expected excess-returns of traded factors to be positive. Although a prior bias towards sparse factor models, resulting from setting the prior means of the alpha's to equal zero, is in line with a preference of a sparse over a less-sparse factor model, unless sufficient posterior evidence is available against its statistical validity, it may well be prudent to re-evaluate the specification of the prior means and prior variances of the alpha's across the candidate models in further research, such that they are more in harmony with economic theory.

In our empirical study, we apply the marginal likelihood based factor selection methodology of Chib et al. (2018), using priors for the alpha's with increased spreads, to select priced risk factors out of a set of prominent global (excess-return) factors proposed in the literature. We find that the maximum attainable (sample) Sharpe ratios over two portfolios of, respectively, two selected factor sets outperform maximum attainable Sharpe ratio's over portfolios of, respectively, factor sets of several prominent factor models proposed in the literature. The two selected factor models also outperform the factor models proposed in the literature in terms of pricing performance w.r.t. excluded factors. We find the seven factor model consisting of the LWMKT, Dollar and Carry factors of Brusa et al. (2014) and the SMB, DHML, QMJ and MOM factors of, respectively, Fama and French (1993), Asness and Frazzini (2013), Asness et al. (2019) and Carhart (1997) performs best in terms of pricing performance w.r.t. excluded factors. The set of factors making up the seven factor model is thus superior in terms of approximating the true mean-variance efficient portfolio on the international level, as compared to other popular factor sets proposed in the literature. Although we find the seven factor model's overall ability to explain differences across the cross-section of expected excess-returns of global stock portfolios to be more satisfactory as compared to other prominent factor models proposed in the literature, it remains a challenge to fully explain the cross-section of expected excess-returns of global stocks, as none of our considered factor models are likely able to price all of our global stock portfolios without pricing error.

References

Asness, C. and Frazzini, A. (2013). The devil in hml's details. *The Journal of Portfolio Management*, 39: 49 – 68.

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

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68: 929 – 985.

Back, K. E. (2015). *Asset Pricing and Portfolio Choice Theory*. Oxford University Press.

Barillas, F. and Shanken, J. (2017). Which alpha? *Review of Financial Studies*, 30: 1316 – 1338.

Barillas, F. and Shanken, J. (2018). Comparing asset pricing models. *The Journal of Finance*, 73: 715 – 754.

Brusa, F., Ramdadorai, T., and Verdelhan, A. (2014). The international capm redux. *Saïd Business School Working Paper*.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Financial Economics*, 52: 57 – 82.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90: 1313 – 1321.

Chib, S. and Zeng, X. (2018). Which factors are risk factors in asset pricing? a model scan framework. *Journal of Business & Economic Statistics*, forthcoming.

Chib, S., Zeng, X., and Zhao, L. (2018). On comparing asset pricing models. *The Journal of Finance*, forthcoming.

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

Dumas, B. and Solnik, B. (1995). The world price of foreign exchange risk. *Journal of Finance*, 50: 445 – 479.

Fama, E. F. (1965). The behavior of stock market prices. *The Journal of Business*, 38: 34 – 105.

Fama, E. F. and French, K. E. (1992). The cross section of expected stock returns. *Journal of Financial Economics*, 47: 427 – 465.

Fama, E. F. and French, K. E. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33: 3 – 56.

Fama, E. F. and French, K. E. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105: 457 – 472.

Fama, E. F. and French, K. E. (2015). A five-factor asset pricing model. *Journal of Financial*

Economics, 116: 1 –22.

Fan, Z., Londono, J. M., and Xiao, X. (2019). Us equity tail risk and currency risk premia. *International Finance Discussion Papers.*

Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111: 1 – 25.

Gibbons, M., Ross, S., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57: 1121 – 1151.

Hansen, L. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy*, 99: 225 – 262.

Hou, K., Karolyi, G., and Kho, B. (2011). What factors drive global stock returns? *Review of Financial Studies*, 24: 2527 – 2574.

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

Huberman, G. and Kandel, S. (1987). Mean variance spanning. *Journal of Finance*, 42: 873 – 888.

Jeffreys, H. (1961). *Theory of Probability.* Oxford University Press, New York.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48: 65 – 91.

Lewellen, J. and Nagel, S. (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82: 289 – 314.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47: 13 – 37.

Pukthuanthong, K., Roll, R., and Subrahmanyam, A. (2019). A protocol for factor identification. *Review of Financial Studies*, 32: 1573 – 1607.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19: 425 – 442.

7 Appendix

A Auxiliary Derivations

A.1 Specifying the Spread of the Prior of Parameter Alpha

We consider a set of H normally distributed (candidate) excess-return risk factors, captured in $(H \times 1)$ vector f_t:

$$f_t = \alpha + \epsilon_t, \quad \epsilon_t \sim N_H(0, \boldsymbol{\Sigma}).$$

We consider variance matrix $\boldsymbol{\Sigma}$ given for illustrative purposes, and specify the prior distribution of α as

$$\pi(\alpha) = \phi_H(\alpha | 0, k\boldsymbol{\Sigma}). \tag{20}$$

Back (2015) shows that the squared maximum attainable Sharpe ratio over any portfolio of assets equals

$$\text{Sh}^2 = \mu' \boldsymbol{\Omega}^{-1} \mu,$$

with μ denoting the mean of the excess-returns of the assets and $\boldsymbol{\Omega}$ denoting the variance matrix of the excess-returns of the assets. Thus, the squared maximum attainable Sharpe ratio over the portfolio of our H factors, denoted by $\text{Sh}(f)^2$, equals

$$\text{Sh}(f)^2 = \alpha' \boldsymbol{\Sigma}^{-1} \alpha.$$

Barillas and Shanken (2018) argue that, in case $H = 1$, k equals the prior expectation of the squared alpha divided by residual variance $\boldsymbol{\Sigma} = \sigma^2$, or the expected squared Sharpe ratio. In case $H = 1$, prior (20) implies

$$E(\alpha^2) = k\sigma^2, \quad k = E(\alpha^2/\sigma^2), \quad k = E(\text{Sh}(f)^2).$$

Barillas and Shanken (2018) argue that, in general, with $H > 1$ factors, the quadratic form $\alpha'(k\boldsymbol{\Sigma})^{-1}\alpha$ is distributed as chi-square with H d.o.f. Thus

$$E(\alpha'(k\boldsymbol{\Sigma})^{-1}\alpha) = H, \quad k = E(\alpha'(\boldsymbol{\Sigma})^{-1}\alpha)/H, \quad k = E(\text{Sh}(f)^2)/H.$$

Assuming our set of H factors span the true mean-variance efficient portfolio, and assuming the market portfolio is not mean-variance efficient, our best guess for the prior expectation of the squared maximum Sharpe ratio over the portfolio of the factors, which equals the squared Sharpe ratio of the mean-variance efficient portfolio, ShMax^2, is a multiple of the squared Sharpe ratio of the market portfolio, ShMkt^2:

$$E(\text{Sh}(f)^2) = \text{ShMax}^2 = (\tau \text{ShMkt})^2.$$

A.2 Inverse Map Derivation

Let factor model $\mathcal{M}_1$ stand for the factor model in which all of the H candidate factors are (proposed) priced risk factors (we omit the pricing equation of the test-assets for simplicity)

$$f_{1,t} = \alpha_1 + \epsilon_{1,t}, \quad \epsilon_{1,t} \sim N_H(0, \boldsymbol{\Sigma}_1). \tag{21}$$

The "nuisance" parameter η_1 of this model is $\eta_1 = \sigma_1 = \text{vech}(\boldsymbol{\Sigma}_1)$. Factor model $\mathcal{M}_j$ $(j > 2)$ is written as (omitting the pricing equation of the test-assets)

$$\begin{aligned} f_{j,t} &= \alpha_j + \epsilon_{j,t}, \quad \epsilon_{j,t} \sim N_{K_j}(0, \boldsymbol{\Sigma}_j), \\ \tilde{f}_{j,t} &= \boldsymbol{\beta}_{j,f} f_{j,t} + \nu_{j,t}, \quad \varepsilon_{j,t} \sim N_{M_j}(0, \tilde{\boldsymbol{\Sigma}}_j), \end{aligned}$$

with $f_{j,t}$ and $\tilde{f}_{j,t}$ denoting, respectively, a vector of (proposed) priced risk factors and a vector of non-risk factors. The "nuisance" parameter η_j of model $\mathcal{M}_j$ $(j > 2)$ is

$$\eta_j = (\beta'_{j,f}, \sigma'_j, \tilde{\sigma}'_j)', \quad \beta_{j,f} = \text{vec}(\boldsymbol{\beta}_{j,f}), \quad \sigma_j = \text{vech}(\boldsymbol{\Sigma}_j), \quad \tilde{\sigma}_j = \text{vech}(\tilde{\boldsymbol{\Sigma}}_j).$$

We will proceed to derive the inverse map that connects "nuisance" parameters η_1 and η_j $(j > 2)$

$$\eta_1 = g_j^{-1}(\eta_j).$$

We rewrite model $\mathcal{M}_j$ $(j > 2)$ by substituting the expression for $f_{j,t}$ into the expression of $\tilde{f}_{j,t}$:

$$\begin{pmatrix} f_{j,t} \\ \tilde{f}_{j,t} \end{pmatrix} = \begin{pmatrix} \alpha_j \\ \boldsymbol{\beta}_{j,f}\alpha_j \end{pmatrix} + \begin{pmatrix} \epsilon_{j,t} \\ \boldsymbol{\beta}_{j,f}\epsilon_{j,t} + \varepsilon_{j,t} \end{pmatrix}, \tag{22}$$

with

$$\begin{pmatrix} \epsilon_{j,t} \\ \boldsymbol{\beta}_{j,f}\epsilon_{j,t} + \varepsilon_{j,t} \end{pmatrix} \sim N_H(0, \boldsymbol{V}_j), \quad \boldsymbol{V}_j = \begin{pmatrix} \boldsymbol{\Sigma}_j & \boldsymbol{\Sigma}_j \boldsymbol{\beta}'_{j,f} \\ \boldsymbol{\beta}_{j,f}\boldsymbol{\Sigma}_j & \tilde{\boldsymbol{\Sigma}}_j + \boldsymbol{\beta}_{j,f}\boldsymbol{\Sigma}_j \boldsymbol{\beta}_{j,f} \end{pmatrix}. \tag{23}$$

Under the assumption that model $\mathcal{M}_j$ $(j > 2)$ is correctly specified, i.e. non-risk factors $\tilde{f}_{j,t}$ are priced by proposed priced risk factors $f_{j,t}$ without pricing error, model $\mathcal{M}_j$, as given by equations (22) and (23), is equivalent to model $\mathcal{M}_1$, as given by equation (21). Thus, assuming model $\mathcal{M}_j$ $(j > 2)$ is correctly specified, $\boldsymbol{\Sigma}_1$ equals $\boldsymbol{V}_j$

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \boldsymbol{\Sigma}_j & \boldsymbol{\Sigma}_j \boldsymbol{\beta}'_{j,f} \\ \boldsymbol{\beta}_{j,f}\boldsymbol{\Sigma}_j & \tilde{\boldsymbol{\Sigma}}_j + \boldsymbol{\beta}_{j,f}\boldsymbol{\Sigma}_j \boldsymbol{\beta}'_{j,f} \end{pmatrix}.$$

Alternatively, in vech form:

$$\sigma_1 = \begin{pmatrix} \sigma_j \\ (\boldsymbol{\Sigma_j} \otimes I_{M_j})\beta_{j,f} \\ \tilde{\sigma}_j + \text{vech}(\boldsymbol{\beta}_{j,f}\boldsymbol{\Sigma}_j \boldsymbol{\beta}'_{j,f}) \end{pmatrix}, \tag{24}$$

with I_{M_j} denoting the $(M_j \times M_j)$ identity matrix (and M_j being the number of non-risk factors in $\tilde{f}_{j,t}$). The set of vector equations given in Eq. (24) constitute the inverse map

$$\eta_1 = \sigma_1 = g_j^{-1}(\eta_j).$$

A.3 Marginal Likelihood Derivation

We derive a closed form expression for the marginal likelihood of factor model $\mathcal{M}_j$, $j = 1, ..., J$:

$$\begin{aligned} f_{j,t} &= \alpha_j + \epsilon_{j,t}, \quad \epsilon_{j,t} \sim N_{K_j}(0, \boldsymbol{\Sigma}_j), \\ \tilde{f}_{j,t} &= \boldsymbol{\beta}_{j,f} f_{j,t} + \nu_{j,t}, \quad \nu_{j,t} \sim N_{M_j}(0, \tilde{\boldsymbol{\Sigma}}_j), \\ r_t &= \boldsymbol{\beta}_r f_t^* + u_t, \quad u_t \sim N_N(0, \tilde{\boldsymbol{\Sigma}}_r), \end{aligned} \tag{25}$$

with shocks $\epsilon_{j,t}$, $\nu_{j,t}$ and u_t being mutually independent. Vectors f_t^* and r_t respectively denote a $(H \times 1)$ vector of candidate factors and a $(N \times 1)$ vector of test-assets. Vectors $f_{j,t}$ and $\tilde{f}_{j,t}$ respectively denote a $(K_j \times 1)$ vector of proposed priced risk factors and a $(M_j \times 1)$ vector of implied non-risk factors. Let $\beta_{j,f} = \text{vec}(\boldsymbol{\beta}_{j,f})$ and $\beta_r = \text{vec}(\boldsymbol{\beta}_r)$ respectively denote the vectorizations of $\boldsymbol{\beta}_{j,f}$ and $\boldsymbol{\beta}_r$. Let $\sigma_j = \text{vech}(\boldsymbol{\Sigma}_j)$, $\tilde{\sigma}_j = \text{vech}(\tilde{\boldsymbol{\Sigma}}_j)$ and $\sigma_r = \text{vech}(\tilde{\boldsymbol{\Sigma}}_r)$ denote the half vectorizations of the three covariance matrices. The parameters of model $\mathcal{M}_j$, $j = 1, ..., J$, are then

$$\theta_j = (\alpha_j', \eta_j', \beta_r', \sigma_r')' \in (\Theta_{\alpha_j}',\ \Theta_{\eta_j}',\ \Theta_{\beta_r}',\ \Theta_{\sigma_r}')', \quad \eta_j = (\beta_{j,f}', \sigma_j', \tilde{\sigma}_j')' \in \Theta_{\eta_j},$$

where Θ_{α_j}, Θ_{η_j}, Θ_{β_r} and Θ_{σ_r} respectively denote the parameter spaces of α_j, η_j, β_r and σ_r. We specify the prior density of parameter θ_j:

$$p(\theta_j|\mathcal{M}_j) = \pi(\alpha_j|\mathcal{M}_j, \eta_j)\psi(\eta_j|\mathcal{M}_j, \beta_r)\psi_r(\beta_r, \sigma_r).$$

The conditional prior of α_j, $\pi(\alpha_j|\mathcal{M}_j, \eta_j)$, is specified as a proper density:

$$\pi(\alpha_j|\mathcal{M}_j, \eta_j) = \phi_{K_j}(\alpha_j|0, k\boldsymbol{\Sigma}_j),$$

with $\phi_d(\cdot|\mu, \boldsymbol{\Sigma})$ denoting the pdf of the d-dimensional multivariate normal distribution with mean μ and covariance matrix $\boldsymbol{\Sigma}$. The prior of parameter η_j is specified as the uninformative prior (we set $c = 1$):

$$\psi(\eta_j|\mathcal{M}_j) = |\boldsymbol{\Sigma}_j|^{-\frac{2K_j - H + 1}{2}} |\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{H+1}{2}}$$

We leave the prior of parameters β_r and σ_r, $\psi_r(\beta_r, \sigma_r)$ unspecified, the following derivations hold for any arbitrary specification of the prior $\psi_r(\beta_r, \sigma_r)$.

Let $\boldsymbol{Y} = (y_1, ..., y_T)'$ (with $y_t = ((f_t^*)', r_t')'$) denote the observation matrix of all candidate risk factors and test-assets. The marginal likelihood of model $\mathcal{M}_j$ is given as

$$m(\boldsymbol{Y}|\mathcal{M}_j) = \int_{\Theta_{\sigma_r}} \int_{\Theta_{\beta_r}} \int_{\Theta_{\eta_j}} \int_{\Theta_{\alpha_j}} p(\boldsymbol{Y}|\mathcal{M}_j, \theta_j)\pi(\alpha_j|\mathcal{M}_j, \eta_j)\psi(\eta_j|\mathcal{M}_j, \beta_r)\psi_r(\beta_r, \sigma_r)\mathrm{d}\theta_j. \tag{26}$$

The density $p(\boldsymbol{Y}|\mathcal{M}_j,\theta_j)$ is the likelihood function implied by factor model $\mathcal{M}_j$ (Eq. (25)). Let $\boldsymbol{F}^* = (f_1^*,...,f_T^*)'$, $\boldsymbol{F} = (f_{j,1},...,f_{j,T})'$, $\tilde{\boldsymbol{F}} = (\tilde{f}_{j,1},...,\tilde{f}_{j,T})'$ and $\boldsymbol{R} = (r_1,...,r_T)'$ denote observation matrices of, respectively, all candidate factors, priced risk factors, non-risk factors and test-assets. Let I_T denote the $T\times T$ identity matrix, and ι_T a $(T\times 1)$ vector of ones. The likelihood function $p(\boldsymbol{Y}|\mathcal{M}_j,\theta_j)$ can be split up due to the independence of the error terms of model $\mathcal{M}_j$:

$$p(\boldsymbol{Y}|\mathcal{M}_j,\theta_j) = \phi_{T\times K_j}(\boldsymbol{F}|\iota_T\alpha_j',\boldsymbol{\Sigma}_j\otimes I_T)\phi_{T\times M_j}(\tilde{\boldsymbol{F}}|\boldsymbol{F}\boldsymbol{\beta}_{j,f}',\tilde{\boldsymbol{\Sigma}}_j\otimes I_T)\phi_{T\times N}(\boldsymbol{R}|\boldsymbol{F}^*\boldsymbol{\beta}_r',\tilde{\boldsymbol{\Sigma}}_r\otimes I_T),$$

with $\phi_{c\times d}(\cdot|\boldsymbol{M},\boldsymbol{S}\otimes\boldsymbol{L})$ denoting the pdf of a matricvariate normally distributed random $c\times d$ matrix with mean matrix $\boldsymbol{M}$ $(c\times d)$ and with covariance matrix $\boldsymbol{S}\otimes\boldsymbol{L}$ $(\boldsymbol{S}: d\times d,\ \boldsymbol{L}: c\times c)$. Plugging in the likelihood function $p(\boldsymbol{Y}|\mathcal{M}_j,\theta_j)$ into the marginal likelihood, Eq. (26), results in the following expression for the marginal likelihood:

$$m(\boldsymbol{Y}|\mathcal{M}_j) = m(\boldsymbol{F}|\mathcal{M}_j)m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)m(\boldsymbol{R}),$$

with

$$\begin{aligned}
m(\boldsymbol{F}|\mathcal{M}_j) &= \int_{\Theta_{\alpha_j}}\int_{\Theta_{\sigma_j}} \phi_{T\times K_j}(\boldsymbol{F}|\iota_T\alpha_j',\boldsymbol{\Sigma}_j\otimes I_T)\phi_{K_j}(\alpha_j|0,k\boldsymbol{\Sigma}_j)|\boldsymbol{\Sigma}_j|^{-\frac{2K_j-H+1}{2}}\mathrm{d}\alpha_j\mathrm{d}\sigma_j\\
m(\tilde{\boldsymbol{F}}|\mathcal{M}_j) &= \int_{\Theta_{\tilde{\sigma}_j}} \phi_{T\times M_j}(\tilde{\boldsymbol{F}}|\boldsymbol{F}\boldsymbol{\beta}_{j,f}',\tilde{\boldsymbol{\Sigma}}_j\otimes I_T)|\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{H+1}{2}}\mathrm{d}\tilde{\sigma}_j\\
m(\boldsymbol{R}) &= \int_{\Theta_{\sigma_r}}\int_{\Theta_{\beta_r}} \phi_{T\times N}(\boldsymbol{R}|\boldsymbol{F}^*\boldsymbol{\beta}_r',\tilde{\boldsymbol{\Sigma}}_r\otimes I_T)\psi_r(\beta_r,\sigma_r)\mathrm{d}\beta_r\mathrm{d}\sigma_r.
\end{aligned}$$

The expression $m(\boldsymbol{F}|\mathcal{M}_j)$ corresponds to the marginal likelihood of the sub-model of the priced risk factors, as implied by model $\mathcal{M}_j$:

$$f_{j,t} = \alpha_j + \epsilon_{j,t},\quad \epsilon_{j,t}\sim N_{K_j}(0,\boldsymbol{\Sigma}_j),$$

with the following prior density for the sub-model parameters:

$$p(\alpha_j,\sigma_j|\mathcal{M}_j) = \phi_{K_j}(\alpha_j|0,k\boldsymbol{\Sigma}_j)|\boldsymbol{\Sigma}_j|^{-\frac{2K_j-H+1}{2}}.$$

The expression $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ corresponds to the marginal likelihood of the sub-model of the non-risk factors, as implied by model $\mathcal{M}_j$:

$$\tilde{f}_{j,t} = \boldsymbol{\beta}_{j,f}f_{j,t} + \nu_{j,t},\quad \nu_{j,t}\sim N_{M_j}(0,\tilde{\boldsymbol{\Sigma}}_j),$$

with the following prior density for the sub-model parameters:

$$p(\tilde{\sigma}_j,\beta_{j,f}|\mathcal{M}_j) = |\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{H+1}{2}}.$$

The expression $m(\boldsymbol{R})$ corresponds to the marginal likelihood of the sub-model of the test-assets

$$r_t = \boldsymbol{\beta}_r f_t^* + u_t,\quad u_t\sim N_N(0,\tilde{\boldsymbol{\Sigma}}_r),$$

with an (unspecified) arbitrary prior density for the sub-model parameters, denoted by $\psi_r(\beta_r, \sigma_r)$. The sub-model of the test-assets is identical across all candidate models $\mathcal{M}_j$, $j = 1, ..., J$. Therefore, the marginal likelihood of the sub-model of the test-assets, $m(\boldsymbol{R})$, is identical across all candidate models $\mathcal{M}_j$, $j = 1, ..., J$, as well.

As marginal likelihood $m(\boldsymbol{R})$ is identical across all candidate factor models $\mathcal{M}_j$, $j = 1, ..., J$, the marginal likelihood $m(\boldsymbol{R})$ always cancels out when constructing ratio's of marginal likelihoods (Bayes factors) of candidate factor models. When comparing the marginal likelihoods of candidate factor models, we thus may omit the inclusion of test-assets in our candidate factor models (resulting in an empty vector r_t with $N = 0$) and simply calculate the marginal likelihood of candidate model $\mathcal{M}_j$, $j = 1, ..., J$, as

$$m(\boldsymbol{Y}|\mathcal{M}_j) = m(\boldsymbol{F}|\mathcal{M}_j)m(\tilde{\boldsymbol{F}}|\mathcal{M}_j).$$

We proceed to derive closed form expressions for marginal likelihoods $m(\boldsymbol{F}|\mathcal{M}_j)$ and $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$, in turn resulting in a closed form expression for marginal likelihood $m(\boldsymbol{Y}|\mathcal{M}_j)$.

The expression $m(\boldsymbol{F}|\mathcal{M}_j)$ corresponds to the marginal likelihood of the sub-model of the priced risk factors, as implied by model $\mathcal{M}_j$:

$$f_{j,t} = \alpha_j + \epsilon_{j,t}, \quad \epsilon_{j,t} \sim N_{K_j}(0, \boldsymbol{\Sigma}_j), \quad A_j = \alpha_j', \tag{27}$$

with the following prior density for the sub-model parameters:

$$p(\alpha_j, \sigma_j|\mathcal{M}_j) = \phi_{K_j}(\alpha_j|0, k\boldsymbol{\Sigma}_j)|\boldsymbol{\Sigma}_j|^{-\frac{2K_j-H+1}{2}},$$

which may be written as

$$p(\alpha_j, \sigma_j|\mathcal{M}_j) = (2\pi)^{-\frac{1}{2}K_j}k^{-\frac{1}{2}K_j}|\boldsymbol{\Sigma}_j|^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\text{tr}[\boldsymbol{\Sigma}_j^{-1}(A_j'k^{-1}A_j)]\Big)|\boldsymbol{\Sigma}_j|^{-\frac{K_j-M_j+1}{2}}, \tag{28}$$

with $\text{tr}[\boldsymbol{M}]$ denoting the trace operator of square matrix $\boldsymbol{M}$. The likelihood function of the sub-model of priced risk factors, implied by $\mathcal{M}_j$ and given by Eq. (27), is given as

$$p(\boldsymbol{F}|\alpha_j, \sigma_j, \mathcal{M}_j) = \phi_{T\times K_j}(\boldsymbol{F}|\iota_T\alpha_j', \boldsymbol{\Sigma}_j \otimes I_T),$$

$$p(\boldsymbol{F}|\alpha_j, \sigma_j, \mathcal{M}_j) = (2\pi)^{-\frac{1}{2}K_jT}|\boldsymbol{\Sigma}_j|^{-\frac{T}{2}}\exp\Big(-\frac{1}{2}\text{tr}[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{F}-\iota_TA_j)'(\boldsymbol{F}-\iota_TA_j)]\Big). \tag{29}$$

Marginal likelihood $m(\boldsymbol{F}|\mathcal{M}_j)$ is, following Chib (1995), calculated as

$$m(\boldsymbol{F}|\mathcal{M}_j) = \frac{p(\boldsymbol{F}|\alpha_j, \sigma_j, \mathcal{M}_j)p(\alpha_j, \sigma_j|\mathcal{M}_j)}{p(\alpha_j|\boldsymbol{F}, \sigma_j, \mathcal{M}_j)p(\sigma_j|\boldsymbol{F}, \mathcal{M}_j)}, \tag{30}$$

with $p(\alpha_j|\boldsymbol{F}, \sigma_j, \mathcal{M}_j)$ denoting the conditional posterior of α_j, conditional on σ_j, and $p(\sigma_j|\boldsymbol{F}, \mathcal{M}_j)$ denoting the marginal posterior distribution of σ_j. Substituting closed form expressions of

$p(\alpha_j, \sigma_j|\mathcal{M}_j)$, $p(\boldsymbol{F}|\alpha_j, \sigma_j, \mathcal{M}_j)$, $p(\alpha_j|\boldsymbol{F}, \sigma_j, \mathcal{M}_j)$ and $p(\sigma_j|\boldsymbol{F}, \mathcal{M}_j)$ in Eq. (30) will result in a closed form expression for $m(\boldsymbol{F}|\mathcal{M}_j)$.

We proceed our discussion by deriving closed form expressions of $p(\alpha_j|\boldsymbol{F}, \sigma_j, \mathcal{M}_j)$ and $p(\sigma_j|\boldsymbol{F}, \mathcal{M}_j)$. First, we determine the kernel of the joint posterior distribution of α_j and σ_j

$$p(\alpha_j, \sigma_j|\boldsymbol{F}, \mathcal{M}_j) \propto p(\alpha_j, \sigma_j|\mathcal{M}_j)p(\boldsymbol{F}|\alpha_j, \sigma_j, \mathcal{M}_j)$$

$$p(\alpha_j, \sigma_j|\boldsymbol{F}, \mathcal{M}_j) \propto |\boldsymbol{\Sigma}_j|^{-\frac{T+K_j-M_j+2}{2}} \exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(A_j'k^{-1}A_j)]\Big)\exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{F}-\iota_T A_j)'(\boldsymbol{F}-\iota_T A_j)]\Big).$$

We re-write

$$(\boldsymbol{F} - \iota_T A_j)'(\boldsymbol{F} - \iota_T A_j) + A_j'k^{-1}A_j = (\boldsymbol{W} - \boldsymbol{V}A_j)'(\boldsymbol{W} - \boldsymbol{V}A_j),$$

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{F} \\ 0_{1\times K_j} \end{pmatrix}, \quad \boldsymbol{V} = \begin{pmatrix} \iota_T \\ k^{-\frac{1}{2}} \end{pmatrix},$$

with $0_{1\times K_j}$ a $(1\times K_j)$ vector of zeroes. Thus, we are able to re-write the kernel of $p(\alpha_j, \sigma_j|\boldsymbol{F}, \mathcal{M}_j)$ as

$$p(\alpha_j, \sigma_j|\boldsymbol{F}, \mathcal{M}_j) \propto |\boldsymbol{\Sigma}_j|^{-\frac{T+K_j-M_j+2}{2}} \Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{W} - \boldsymbol{V}A_j)'(\boldsymbol{W} - \boldsymbol{V}A_j)]\Big).$$

Using the decomposition rule

$$(\boldsymbol{W} - \boldsymbol{V}A_j)'(\boldsymbol{W} - \boldsymbol{V}A_j) = (\boldsymbol{W} - \boldsymbol{V}\tilde{A}_j)'(\boldsymbol{W} - \boldsymbol{V}\tilde{A}_j) + (A_j - \tilde{A}_j)'(\boldsymbol{V}'\boldsymbol{V})(A_j - \tilde{A}_j), \qquad (31)$$

with $\tilde{A}_j = (\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'\boldsymbol{W}$, we re-write the kernel of $p(\alpha_j, \sigma_j|\boldsymbol{F}, \mathcal{M}_j)$ as

$$p(\alpha_j, \sigma_j|\boldsymbol{F}, \mathcal{M}_j) \propto |\boldsymbol{\Sigma}_j|^{-\frac{T+K_j-M_j+2}{2}} \exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{W} - \boldsymbol{V}\tilde{A}_j)'(\boldsymbol{W} - \boldsymbol{V}\tilde{A}_j)]\Big) \exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(A_j - \tilde{A}_j)'(T + k^{-1})(A_j - \tilde{A}_j)]\Big),$$

with $\tilde{A}_j = (T+k^{-1})^{-1}\iota_T\boldsymbol{F}$. Thus, the kernel of the marginal posterior distribution $p(\sigma_j|\boldsymbol{F}, \mathcal{M}_j)$ is given as

$$p(\sigma_j|\boldsymbol{F}, \mathcal{M}_j) \propto \int_{\Theta_{\alpha_j}} p(\alpha_j, \sigma_j|\boldsymbol{F}, \mathcal{M}_j)\mathrm{d}\alpha_j,$$

which can be re-written as

$$p(\sigma_j|\boldsymbol{F}, \mathcal{M}_j) \propto |\boldsymbol{\Sigma}_j|^{-\frac{T+K_j-M_j+2}{2}} \exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{W} - \boldsymbol{V}\tilde{A}_j)'(\boldsymbol{W} - \boldsymbol{V}\tilde{A}_j)]\Big)|\boldsymbol{\Sigma}_j(T + k^{-1})^{-1}|^{\frac{1}{2}} \int_{\Theta_{\alpha_j}} |\boldsymbol{\Sigma}_j(T + k^{-1})^{-1}|^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(A_j - \tilde{A}_j)'(T + k^{-1})(A_j - \tilde{A}_j)]\Big)\mathrm{d}\alpha_j.$$

Using the fact that

$$\int_{\Theta_{\alpha_j}} |\boldsymbol{\Sigma}_j(T + k^{-1})^{-1}|^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(A_j - \tilde{A}_j)'(T + k^{-1})(A_j - \tilde{A}_j)]\Big)\mathrm{d}\alpha_j$$

integrates to a constant as the integrand is the kernel of a multivariate normal distribution with mean $\tilde{A}_j$ and variance $(T+k^{-1})^{-1}\boldsymbol{\Sigma}_j$, and the fact that $|\boldsymbol{\Sigma}_j(T+k^{-1})^{-1}| = |\boldsymbol{\Sigma}_j|(T+k^{-1})^{-K_j}$, we re-write the kernel of the marginal posterior distribution $p(\sigma_j|\boldsymbol{F},\mathcal{M}_j)$ as

$$p(\sigma_j|\boldsymbol{F},\mathcal{M}_j) \propto |\boldsymbol{\Sigma}_j|^{-\frac{T+K_j-M_j+1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[(\boldsymbol{W}-\boldsymbol{V}\tilde{A}_j)'(\boldsymbol{W}-\boldsymbol{V}\tilde{A}_j)]\Big),$$

which we identify to be the kernel of an inverted Wishart distribution with parameter matrix $\boldsymbol{S}_j = (\boldsymbol{W}-\boldsymbol{V}\tilde{A}_j)'(\boldsymbol{W}-\boldsymbol{V}\tilde{A}_j)$, and $T-M_j$ d.o.f.:

$$p(\sigma_j|\boldsymbol{F},\mathcal{M}_j) = c^{-1}|\boldsymbol{S}_j|^{\frac{T-M_j}{2}}|\boldsymbol{\Sigma}_j|^{-\frac{T-M_j+K_j+1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}\boldsymbol{S}_j]\Big), \tag{32}$$

with $c = \Big[2^{\frac{(T-M_j)K_j}{2}}\Gamma_{K_j}\Big(\frac{T-M_j}{2}\Big)\Big]$, and with $\Gamma_d(\cdot)$ denoting the d dimensional multivariate gamma function. Note that we may re-write parameter $\boldsymbol{S}_j$ as

$$\boldsymbol{S}_j = \sum_{t=1}^{T}(f_{j,t}-\hat{\alpha}_j)(f_{j,t}-\hat{\alpha}_j)' + \frac{k^{-1}T}{T+k^{-1}}\hat{\alpha}_j\hat{\alpha}_j',$$

with α_j denoting the OLS estimate of α_j.

By conditioning the kernel of $p(\alpha_j,\sigma_j|\boldsymbol{F},\mathcal{M}_j)$ on σ_j, we write, using the decomposition rule of Eq. (31), the kernel of the conditional posterior distribution of α_j as

$$p(\alpha_j|\boldsymbol{F},\sigma_j,\mathcal{M}_j) \propto \exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(A_j-\tilde{A}_j)'(T+k^{-1})(A_j-\tilde{A}_j)]\Big),$$

which we identify to be the kernel of multivariate normal distribution with mean $\tilde{A}_j$ and variance $(T+k^{-1})^{-1}\boldsymbol{\Sigma}_j$. Thus

$$p(\alpha_j|\boldsymbol{F},\sigma_j,\mathcal{M}_j) = \Big(\frac{1}{2\pi}\Big)^{\frac{K_j}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}(T+k^{-1})^{\frac{K_j}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(A_j-\tilde{A}_j)'(T+k^{-1})(A_j-\tilde{A}_j)]\Big).$$

Substituting the closed form expression of $p(\alpha_j|\boldsymbol{F},\sigma_j,\mathcal{M}_j)$ and the closed form expressions of $p(\alpha_j,\sigma_j|\mathcal{M}_j)$, $p(\boldsymbol{F}|\alpha_j,\sigma_j,\mathcal{M}_j)$ and $p(\sigma_j|\boldsymbol{F},\mathcal{M}_j)$ respectively given by equations (28), (29) and (32) in equation (30) gives the closed form expression of $m(\boldsymbol{F}|\mathcal{M}_j)$ as given by Eq. (16).

We turn the discussion towards the derivation of the closed form expression of $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$. The expression $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ corresponds to the marginal likelihood of the sub-model of the non-risk factors, as implied by model $\mathcal{M}_j$:

$$\tilde{f}_{j,t} = \boldsymbol{\beta}_{j,f}f_{j,t} + \nu_{j,t}, \quad \nu_{j,t} \sim N_{M_j}(0,\tilde{\boldsymbol{\Sigma}}_j), \quad \boldsymbol{B}_j = \boldsymbol{\beta}_{j,f}', \tag{33}$$

with the following prior density for the sub-model parameters:

$$p(\tilde{\sigma}_j,\beta_{j,f}|\mathcal{M}_j) = |\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{K_j+M_j+1}{2}}. \tag{34}$$

In the subsequent discussion, we treat observation matrix $\boldsymbol{F} = (f_{j,1}, ..., f_{j,T})'$ as given. The likelihood function of the sub-model of the non-risk factors, implied by model $\mathcal{M}_j$ and given by Eq. (33), is given as

$$p(\tilde{\boldsymbol{F}}|\tilde{\sigma}_j, \beta_{j,f}, \mathcal{M}_j) = \phi_{T\times M_j}(\tilde{\boldsymbol{F}}|\boldsymbol{F}\boldsymbol{\beta}'_{j,f}, \tilde{\boldsymbol{\Sigma}}_j \otimes I_T),$$

$$p(\tilde{\boldsymbol{F}}|\tilde{\sigma}_j, \beta_{j,f}, \mathcal{M}_j) = (2\pi)^{-\frac{1}{2}M_jT}|\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{T}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}(\tilde{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{B}_j)'(\tilde{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{B}_j)]\Big). \quad (35)$$

Again, we calculate marginal likelihood $m(\tilde{\boldsymbol{F}}|\mathcal{M}_j)$ as

$$m(\tilde{\boldsymbol{F}}|\mathcal{M}_j) = \frac{p(\tilde{\boldsymbol{F}}|\tilde{\sigma}_j, \beta_{j,f}, \mathcal{M}_j)p(\tilde{\sigma}_j, \beta_{j,f}|\mathcal{M}_j)}{p(\beta_{j,f}|\tilde{\boldsymbol{F}}, \tilde{\sigma}_j, \mathcal{M}_j)p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j)}, \quad (36)$$

with $p(\beta_{j,f}|\tilde{\boldsymbol{F}}, \tilde{\sigma}_j, \mathcal{M}_j)$ denoting the conditional posterior of $\beta_{j,f}$, conditional on $\tilde{\sigma}_j$, and $p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j)$ denoting the marginal posterior of $\tilde{\sigma}_j$.

We derive closed form expressions of $p(\beta_{j,f}|\tilde{\boldsymbol{F}}, \tilde{\sigma}_j, \mathcal{M}_j)$ and $p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j)$. We determine the kernel of the joint posterior distribution of $\beta_{j,f}$ and $\tilde{\sigma}_j$

$$p(\tilde{\sigma}_j, \beta_{j,f}|\tilde{\boldsymbol{F}}, \mathcal{M}_j) \propto p(\tilde{\boldsymbol{F}}|\tilde{\sigma}_j, \beta_{j,f}, \mathcal{M}_j)p(\tilde{\sigma}_j, \beta_{j,f}|\mathcal{M}_j)$$

$$p(\tilde{\sigma}_j, \beta_{j,f}|\tilde{\boldsymbol{F}}, \mathcal{M}_j) \propto |\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{T+K_j+M_j+1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}(\tilde{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{B}_j)'(\tilde{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{B}_j)]\Big).$$

The kernel of the marginal posterior $p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j)$ is given as

$$p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j) \propto \int_{\Theta_{\beta_{j,f}}} p(\tilde{\sigma}_j, \beta_{j,f}|\tilde{\boldsymbol{F}}, \mathcal{M}_j)\mathrm{d}\beta_{j,f},$$

which can be re-written as

$$p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j) \propto |\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{T+K_j+M_j+1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}(\tilde{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{B}}_j)'(\tilde{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{B}}_j)]\Big)|\tilde{\boldsymbol{\Sigma}}_j \otimes (\boldsymbol{F}'\boldsymbol{F})^{-1}|^{\frac{1}{2}}$$
$$\int_{\Theta_{\beta_{j,f}}} |\tilde{\boldsymbol{\Sigma}}_j \otimes (\boldsymbol{F}'\boldsymbol{F})^{-1}|^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)'(\boldsymbol{F}'\boldsymbol{F})(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)]\Big)\mathrm{d}\beta_{j,f},$$

using the decomposition rule

$$(\tilde{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{B}_j)'(\tilde{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{B}_j) = (\tilde{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{B}}_j)'(\tilde{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{B}}_j) + (\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)'(\boldsymbol{F}'\boldsymbol{F})(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j), \quad (37)$$

with $\hat{\boldsymbol{B}}_j = (\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}\tilde{\boldsymbol{F}}$. Using the fact that

$$\int_{\Theta_{\beta_{j,f}}} |\tilde{\boldsymbol{\Sigma}}_j \otimes (\boldsymbol{F}'\boldsymbol{F})^{-1}|^{-\frac{1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)'(\boldsymbol{F}'\boldsymbol{F})(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)]\Big)\mathrm{d}\beta_{j,f}$$

integrates to a constant as the integrand is the kernel of a matricvariate normal distribution with mean $\hat{\boldsymbol{B}}_j$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}_j \otimes (\boldsymbol{F}'\boldsymbol{F})^{-1}$, and $|\tilde{\boldsymbol{\Sigma}}_j \otimes (\boldsymbol{F}'\boldsymbol{F})^{-1}| = |\tilde{\boldsymbol{\Sigma}}_j|^{K_j}|(\boldsymbol{F}'\boldsymbol{F})^{-1}|^{M_j}$, we re-write the kernel of the marginal posterior $p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j)$ as

$$p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j) \propto |\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{T+M_j+1}{2}}\exp\Big(-\frac{1}{2}\mathrm{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}(\tilde{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{B}}_j)'(\tilde{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{B}}_j)]\Big),$$

which we recognise to be the kernel of an inverted Wishart distribution with parameter matrix $\tilde{\boldsymbol{S}}_j = (\tilde{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{B}}_j)'(\tilde{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{B}}_j) = \sum_{t=1}^{T}(\tilde{f}_{j,t} - \hat{\boldsymbol{\beta}}_{j,f} f_{j,t})(\tilde{f}_{j,t} - \hat{\boldsymbol{\beta}}_{j,f} f_{j,t})'$ (with $\hat{\boldsymbol{\beta}}_{j,f}$ the OLS estimate of $\boldsymbol{\beta}_{j,f}$) and T d.o.f.:

$$p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j) = c^{-1}|\tilde{\boldsymbol{S}}_j|^{\frac{T}{2}}|\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{T+M_j+1}{2}}\exp\Big(-\frac{1}{2}\text{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}\tilde{\boldsymbol{S}}_j]\Big), \tag{38}$$

with $c = \Big[2^{\frac{TM_j}{2}}\Gamma_{M_j}\Big(\frac{T}{2}\Big)\Big]$.

By conditioning the kernel of $p(\tilde{\sigma}_j, \beta_{j,f}|\tilde{\boldsymbol{F}}, \mathcal{M}_j)$ on $\tilde{\sigma}_j$, we write, using the decomposition rule of Eq. (37), the kernel of the conditional posterior of $\beta_{j,f}$ as

$$p(\beta_{j,f}|\tilde{\boldsymbol{F}}, \tilde{\sigma}_j, \mathcal{M}_j) \propto \exp\Big(-\frac{1}{2}\text{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)'(\boldsymbol{F}'\boldsymbol{F})(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)]\Big),$$

which we recognise to be the kernel of a matricvariate normal distribution with mean $\hat{\boldsymbol{B}}_j$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}_j \otimes (\boldsymbol{F}'\boldsymbol{F})^{-1}$:

$$p(\beta_{j,f}|\tilde{\boldsymbol{F}}, \tilde{\sigma}_j, \mathcal{M}_j) = \Big(\frac{1}{2\pi}\Big)^{\frac{M_jK_j}{2}}|\tilde{\boldsymbol{\Sigma}}_j|^{-\frac{K_j}{2}}|(\boldsymbol{F}'\boldsymbol{F})^{-1}|^{-\frac{M_j}{2}}\exp\Big(-\frac{1}{2}\text{tr}[\tilde{\boldsymbol{\Sigma}}_j^{-1}(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)'(\boldsymbol{F}'\boldsymbol{F})(\boldsymbol{B}_j - \hat{\boldsymbol{B}}_j)]\Big).$$

Substituting the closed form expression of $p(\beta_{j,f}|\tilde{\boldsymbol{F}}, \tilde{\sigma}_j, \mathcal{M}_j)$ and the closed form expressions of $p(\tilde{\sigma}_j, \beta_{j,f}|\mathcal{M}_j)$, $p(\tilde{\boldsymbol{F}}|\tilde{\sigma}_j, \beta_{j,f}, \mathcal{M}_j)$ and $p(\tilde{\sigma}_j|\tilde{\boldsymbol{F}}, \mathcal{M}_j)$ respectively given by equations (34), (35) and (38) in Eq. (36) gives the closed form expression of $m(\tilde{\boldsymbol{F}}_j|\mathcal{M}_j)$ as given by Eq. (16).

A.4 Testing for the effect of a factor in a SUR factor model

We consider the factor regression model

$$r_t = \alpha + \boldsymbol{\beta} f_t + \epsilon_t, \quad \epsilon_t \sim N_N(0, \boldsymbol{\Sigma}), \quad \boldsymbol{\beta} = [\beta^1, ..., \beta^K],$$

with r_t and f_t respectively denoting a $(N \times 1)$ vector of test-assets (or non-risk factors) and a $(K \times 1)$ vector of (proposed) priced risk factors. Matrix $\boldsymbol{\beta}$ is a matrix of $N \times K$ regression coefficients, with β^i, $i = 1, ..., K$, denoting the vector of N regression coefficients that correspond with the same factor f_t^i, $f_t = [f_t^1, ..., f_t^K]'$. In case factors f_t price all test-assets r_t, it holds $\alpha = 0$. Let $\boldsymbol{R} = (r_1, ..., r_T)'$, $\boldsymbol{F} = (f_1, ..., f_T)'$ respectively denote observation matrices of test-assets and priced risk factors. Let $\boldsymbol{E} = (\epsilon_1, ..., \epsilon_T)'$ denote the (non-observed) matrix of disturbances. Let regressor matrix $\boldsymbol{X}$ be defined as: $\boldsymbol{X} = (\iota_T\ \boldsymbol{F})$, with ι_T denoting a $(T \times 1)$ vector of ones. The factor regression model can be written as a SUR (seemingly unrelated regression) model:

$$\boldsymbol{R} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \quad \text{vec}(\boldsymbol{E}) \sim N_{TN}(0, \boldsymbol{\Sigma} \otimes I_T), \quad \boldsymbol{B} = (\alpha\ \boldsymbol{\beta})'.$$

The OLS estimator of $\boldsymbol{B}$ is

$$\hat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{R}.$$

Let us denote row i, $i = 1, ..., K+1$ of $\boldsymbol{B}$ and $\hat{\boldsymbol{B}}$ with $\boldsymbol{B}_i$ and $\hat{\boldsymbol{B}}_i$, respectively. Thus, $\boldsymbol{B}_{i+1}$, $i = 1, ..., K$, equals the transpose of β^i, the set of N regression coefficients corresponding with factor f_t^i. It can be shown that the distribution of the OLS estimator of $\boldsymbol{B}_i$, $\hat{\boldsymbol{B}}_i$, is given by

$$\hat{\boldsymbol{B}}_i' \sim N_N(\boldsymbol{B}_i', q_{ii}\boldsymbol{\Sigma}),$$

with q_{ii} denoting the i-th diagonal element of matrix $\boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$. To test for the effect of factor f_t^i, $i = 1, ..., K$, on the N test-assets, we can base a test on the hypothesis $H_0 : \boldsymbol{B}_{i+1}' = a$. If, for example, we aim to test whether the effect of factor f_t^i differs significantly from 0, we test $H_0 : \boldsymbol{B}_{i+1}' = 0$. The quadratic form

$$(\hat{\boldsymbol{B}}_i' - a)'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{B}}_i' - a)/q_{ii} \sim \chi^2(N), \tag{39}$$

is a chi-squared distributed random variable with N d.o.f.. The chi-squared test depends on the covariance matrix $\boldsymbol{\Sigma}$, and is exact in case $\boldsymbol{\Sigma}$ is given. Alternatively, $\boldsymbol{\Sigma}$ may be estimated

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-K}\hat{\boldsymbol{E}}'\hat{\boldsymbol{E}}, \quad \hat{\boldsymbol{E}} = \boldsymbol{R} - \boldsymbol{X}\hat{\boldsymbol{B}},$$

and has the property

$$(T-K)\hat{\boldsymbol{\Sigma}} \sim \mathrm{W}_N(\boldsymbol{\Sigma}, T-K),$$

with $\mathrm{W}_N(\boldsymbol{\Sigma}, T-K)$ denoting the Wishart distribution with $(N \times N)$ scale matrix $\boldsymbol{\Sigma}$ and $T-K$ d.o.f.. Plugging in the estimator of $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\Sigma}}$, into the quadratic form given in Eq. (39) gives the statistic

$$t^2 = (\hat{\boldsymbol{B}}_i' - a)'\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{B}}_i' - a)/q_{ii} \sim \mathrm{T}^2(N, T-K),$$

with $\mathrm{T}^2(N, T-K)$ denoting Hotelling's T^2-distribution with parameters N and $T-K$. The t^2 statistic can be scaled such that the F-statistic

$$F = \frac{T-M-K+1}{N(T-K)}t^2 \sim F(N, T-K-N+1),$$

is a F-distributed random variable with N and $T-K-N+1$ d.o.f.. All derived results are exact under the assumption of normally distributed disturbances. If we relax the assumption, results still hold asymptotically.

B Auxiliary Graphs and Tables

Data

Table 11: Correlation Matrix Candidate Factors

	WMKT	LWMKT	SMB	HML	DHML	RMW	CMA	MOM	QMJ	BAB	DLLR	CRRY	GT
WMKT	1.00												
LWMKT	0.97	1.00											
SMB	-0.18	-0.16	1.00										
HML	-0.17	-0.15	0.02	1.00									
DHML	0.12	0.13	0.00	0.68	1.00								
RMW	-0.43	-0.43	-0.26	0.20	-0.02	1.00							
CMA	-0.40	-0.37	-0.03	0.73	0.41	0.20	1.00						
MOM	-0.24	-0.25	0.13	-0.25	-0.74	0.14	-0.04	1.00					
QMJ	-0.70	-0.70	-0.27	0.08	-0.28	0.75	0.32	0.31	1.00				
BAB	-0.31	-0.24	0.22	0.45	0.08	0.40	0.40	0.24	0.32	1.00			
DLLR	0.28	0.51	0.09	0.04	0.16	-0.15	-0.08	-0.13	-0.33	0.12	1.00		
CRRY	0.40	0.38	0.00	-0.05	0.09	-0.11	-0.16	-0.11	-0.26	-0.11	0.26	1.00	
GT	-0.19	-0.19	0.00	-0.07	-0.05	0.08	-0.08	0.01	0.13	0.05	-0.14	-0.25	1.00

Sample correlation matrix of the set of candidate factors. DLLR, CRRY and GT are abbreviations of Dollar, Carry and Global Tail respectively.

Simulation Study

Table 12: Simulation Study Results Bayesian Factor Selection Methodology I

	τ						
	1.5	2	3	5	10	20	30
			$T = 300$				
Average Accuracy	00 (75)	02 (73)	07 (67)	20 (52)	40 (35)	50 (30)	51 (31)
Minimum Accuracy	00 (99)	00 (99)	00 (99)	01 (96)	12 (82)	20 (77)	21 (77)
			$T = 600$				
Average Accuracy	03 (72)	08 (66)	21 (51)	50 (28)	70 (11)	75 (09)	76 (09)
Minimum Accuracy	00 (99)	00 (99)	00 (99)	03 (96)	36 (59)	57 (43)	56 (40)
			$T = 1200$				
Average Accuracy	10 (65)	20 (53)	39 (37)	78 (17)	85 (03)	91 (01)	92 (01)
Minimum Accuracy	00 (99)	00 (99)	01 (99)	10 (85)	60 (35)	80 (06)	79 (06)

Simulation results Bayesian factor selection methodology. We simulate 13 random DGP's, each DGP is simulated $Z = 100$ times. We apply the Bayesian factor selection methodology to select factors for each simulated DGP, using multiple alternative values for τ. In each "Average Accuracy" row, we display, in plain text, the average selection accuracy observed across the simulated DGP's. In addition, in each "Average Accuracy" row, we display, in (**parentheses**), the average of the percentages of times a sparser, instead of the (DGP implied) true, factor model is selected, observed across the simulated DGP's. In each "Minimum Accuracy" row, we display, in plain text, the minimum selection accuracy observed across the simulated DGP's. In addition, in each "Minimum Accuracy" row, we display, in (**parentheses**), the maximum of the percentages of times a sparser, instead of the (DGP implied) true, factor model is selected, observed across the simulated DGP's. Results are displayed for various sample sizes T. We simulate student-t distributed factors (with $\nu = 5$ d.o.f.).

Table 13: Simulation Study Results Bayesian Factor Selection Methodology DGP: WMKT

	τ						
	1.5	2	3	5	10	20	30
	$T = 300$						
Observed Selection Accuracy	55	58	58	58	59	62	62
	$T = 600$						
Observed Selection Accuracy	71	71	67	67	65	71	73
	$T = 1200$						
Observed Selection Accuracy	93	90	89	85	85	90	90

Simulation results Bayesian factor selection methodology. We simulate DGP: WMKT, the DGP is simulated $Z = 100$ times. We apply the Bayesian factor selection methodology to select factors for the simulated DGP, using multiple alternative values for τ. For each τ, we display the observed selection accuracy: the percentage of times the true model WMKT is identified. Results are displayed for various sample sizes T. We simulate normally distributed factors.

Table 14: Average Loss in Sharpe Ratio, Bayesian Factor Selection Methodology

Priced Factors	$T = 300$		$T = 600$		$T = 1200$	
	$\nu = 5$	$\nu = \infty$	$\nu = 5$	$\nu = \infty$	$\nu = 5$	$\nu = \infty$
I-V	15 (50)	05 (25)	08 (40)	00 (11)	00 (10)	00 (02)
I-IX	16 (35)	08 (10)	03 (23)	00 (05)	00 (05)	00 (05)
I-VI	28 (35)	17 (43)	16 (31)	08 (18)	05 (25)	00 (05)
I-IIX	15 (35)	00 (18)	06 (10)	00 (10)	00 (03)	00 (03)
I-II-V	34 (55)	32 (50)	11 (25)	04 (21)	00 (07)	00 (02)
I-II-IIX	20 (38)	04 (13)	00 (05)	00 (05)	00 (03)	00 (03)
I-III-VII	28 (48)	08 (44)	09 (39)	05 (30)	00 (23)	00 (05)
I-IIX-IX	18 (40)	06 (35)	12 (20)	00 (04)	00 (01)	00 (01)
I-II-III-IIX	35 (53)	28 (51)	25 (50)	15 (33)	03 (20)	00 (10)
I-II-IV-IIX	33 (55)	10 (49)	15 (40)	00 (25)	00 (15)	00 (02)
I-II-VI-IIX	35 (52)	30 (47)	28 (48)	16 (35)	00 (23)	00 (14)
I-IV-VII-IIX	34 (53)	11 (33)	04 (15)	00 (05)	00 (07)	00 (03)
I-III-IV-VII-IIX-IX	29 (38)	13 (21)	06 (11)	00 (00)	00 (00)	00 (00)

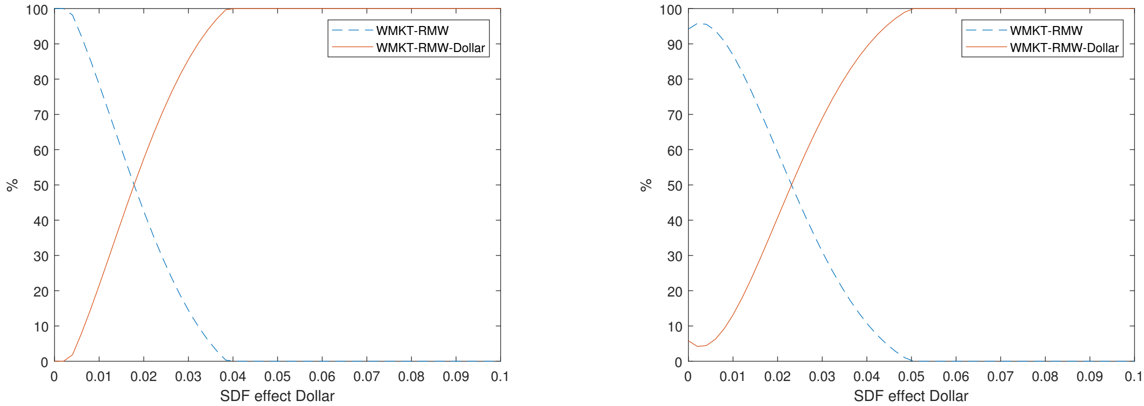
Simulation results Bayesian factor selection methodology (with $\tau = 20$). The left hand side of the table displays sets of true priced risk factors implied by the DGP's we simulate. Each DGP is simulated $Z = 100$ times. Each numerical entry displays the average percentage loss in Sharpe ratio (given that the wrong model has been selected) resulting from constructing the mean-variance efficient portfolio with selected, as opposed to true, priced risk factors. Percentage losses displayed in plain text and in (**parentheses**) are calculated in a setting where portfolio weights are constructed with, respectively, true and estimated moments. Sharpe ratio's of constructed portfolios are calculated using true moments. Results are displayed for various sample sizes T and d.o.f. of the t-distribution ν.

Table 15: Simulation Study Results GMM Based Factor Selection Methodology

Priced Factors	$T = 300$				$T = 1200$			
	N_1	N_2	N_3	N_4	N_1	N_2	N_3	N_4
I-V	9	22	27	26	21	19	26	32
I-IX	7	18	20	32	13	25	24	38
I-VI	5	7	7	6	12	13	15	19
I-IIX	25	37	58	49	38	46	47	54
I-II-V	5	16	33	33	18	18	34	31
I-II-IIX	19	39	65	70	41	70	70	68
I-III-VII	1	4	5	8	2	3	8	8
I-IIX-IX	9	22	26	28	17	29	26	31
I-II-III-IIX	0	12	22	19	8	23	32	27
I-II-IV-IIX	3	25	44	27	17	21	27	31
I-II-VI-IIX	2	8	22	19	5	27	29	28
I-IV-VII-IIX	8	18	41	35	18	40	36	35
I-III-IV-VII-IIX-IX	0	0	8	9	7	20	37	35

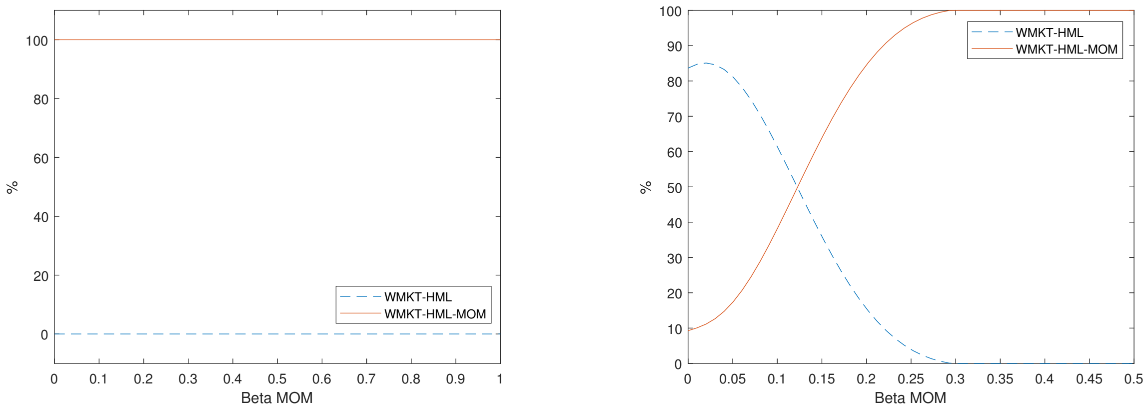
Simulation results GMM factor selection methodology. Sets of true priced risk factors implied by the DGP's we simulate are displayed on the left hand side. Each DGP is simulated $Z = 100$ times. Numerical entries give the percentage of times the true priced risk factors are correctly identified. Results are displayed for various sample sizes T, various test-assets set sizes $N_1 = 20$, $N_2 = 60$, $N_3 = 100$ and $N_4 = 160$, and $\nu = \infty$ d.o.f. of the student-t distribution.

Figure 6: Simulation Results DGP: WMKT-RMW-Dollar, $T = 12000$



We simulate the DGP: WMKT-RMW-Dollar, with normally distributed factors and test-assets (we use N_4 test-assets and sample size $T = 12000$), $Z = 100$ times, for various values of b^{Dlr} (the direct effect of Dollar on the SDF). We use the Bayesian (with $\tau = 20$) and GMM based factor selection methodologies to select priced risk factors each simulation iteration. The set of candidate factors solely consists of WMKT, RMW and Dollar. For each relevant value of b^{Dlr}, the percentage of times WMKT and RMW are selected as well as the percentage of times WMKT, RMW and Dollar are selected are displayed. On the left: Bayesian methodology results. On the right: GMM based methodology results.

Figure 7: Simulation Results DGP: WMKT-HML-MOM, $T = 12000$



We simulate the DGP: WMKT-HML-MOM, with normally distributed factors and test-assets (we use N_1 test-assets and sample size $T = 12000$), $Z = 100$ times, for various values c, the constant governing $\beta^{MOM} = c\iota$. We use the Bayesian (with $\tau = 20$) and GMM based factor selection methodologies to select priced risk factors each simulation iteration. The set of candidate factors solely consists of WMKT, HML and MOM. For each relevant value of c, the percentage of times WMKT and HML are selected as well as the percentage of times WMKT, HML and MOM are selected are displayed. On the left: Bayesian methodology results. On the right: GMM based methodology results.

Table 16: GMM Based Factor Selection

Test-assets	Priced Risk Factors	Test-assets	Priced Risk Factors
Set 1	VII	Set 3	IX-XI
Set 2	II	Set 4	II-III-V-IIX-IX

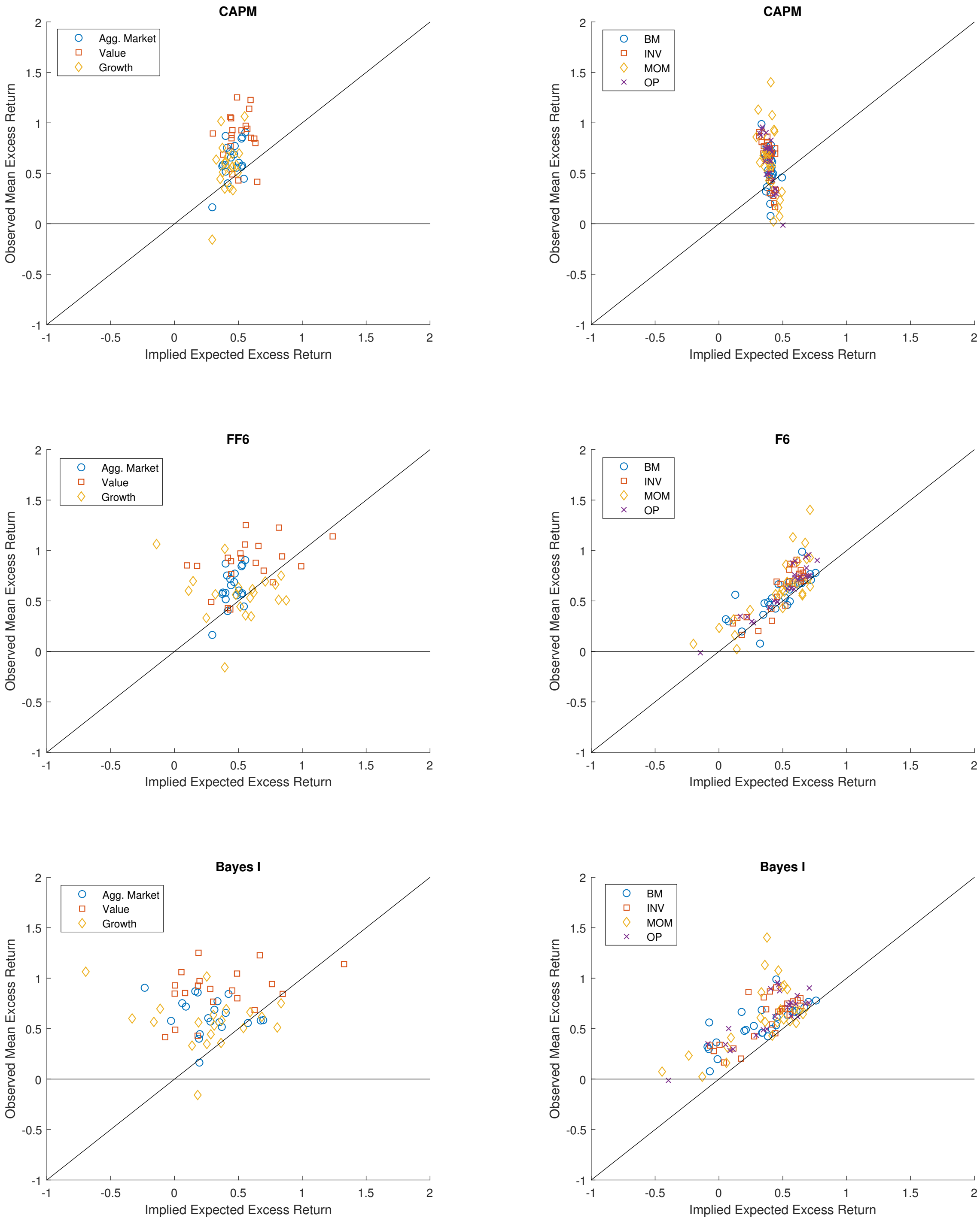
We use the GMM based methodology to select priced risk factors out of our set of $H = 13$ candidate factors. The table displays, for each of our test-asset sets (discussed in section 3.2), which factors the GMM based methodology selects as priced risk factors.

Table 17: Results GRS Tests, sub-sample period

Model	ExF	Test-Assets						
		Market	Value	Growth	MEBM	MEINV	MEMOM	MEOP
I: Rolling Window								
CAPM	99	93	86	89	95	85	99	86
Redux	91	88	80	84	91	75	98	74
FF6	78	81	72	62	73	50	95	73
Bayes I	20	35	44	18	45	26	87	44
Bayes II	1	20	28	10	32	17	79	30
II: Full Sample								
CAPM	10.30	6.20	5.48	5.59	6.69	6.52	7.66	6.20
Redux	9.95	5.67	4.96	5.07	6.18	6.01	7.13	5.69
FF6	8.91	4.70	4.00	4.12	5.25	5.09	6.17	4.78
Bayes I	2.10*	2.65	2.06	2.16	3.32	3.18	4.07	2.93
Bayes II	1.13**	2.44	1.84	1.93	3.15	3.01	3.91	2.75

Part I of the Tables displays, for each candidate factor model and for several sets of test-assets, the share of 60-month rolling windows where the GRS test rejects the null that the factor model prices the set of test-assets (expressed in percentages) during during the sub-sample period February 1995 - May 2013. A significance level of 5% is used. Part II of the Tables display GRS test statistics of full sample GRS tests (during the sub-sample period February 1995 - May 2013). For each candidate factor model, the test-asset set "ExF" refers to the set of factors that are excluded from the factor model (but included in our total set of H candidate factors). GRS test statistics with p-values higher than 5% and 10% are, respectively denoted with a $*$ and a $**$. To combat the the small T versus large N problem resulting from using 60-month rolling windows, we split our set of all $N_4 = 160$ test assets up into seven smaller test-asset sets. The Market, Value and Growth sets respectively consist of the country market indices, country growth indices and country value indices discussed in section 3.2. The MEBM, MEINV, MEMOM and MEOP sets respectively consist of the global portfolios formed by bi-variate sorts on ME and BE/ME, bi-variate sorts on ME and INV, bi-variate sorts on ME and MOM and bi-variate sorts on ME and OP, as discussed in section 3.2. For each factor model, we add the set of excluded factors "ExF" to all the sets of our test-assets.

Figure 8: Realized versus Predicted Expected excess-returns



For each of the factor models, the Figure plots realized expected excess-returns of our entire cross-section of test-assets against the predicted expected excess-returns of the same test-assets as predicted by the candidate factor model (excluded factors are not included in the Figure). We use 60-month rolling windows to estimate factor loadings of the factor models and to estimate (conditional) means of test-asset and factor excess-returns. Each rolling window, mean test-asset excess-returns are predicted, for each of the factor models, by multiplying estimated (conditional) factor loadings by the corresponding estimated (conditional) factor means. All rolling window predicted mean excess-returns and rolling window realized sample mean excess-returns are then averaged over the total number of rolling windows. Averaged predicted mean excess-returns are plotted against averaged realized mean excess-returns, for each of the factor models.