Spring 5-15-2020

Essays on Asset Pricing: A Model Comparison Perspective

Lingxiao Zhao
Washington University in St. Louis

,

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Economics

Dissertation Examination Committee:
Siddhartha Chib, Co-Chair
Werner Ploberger, Co-Chair
Gaetano Antinolfi
John Nachbar
Guofu Zhou

Essays on Asset Pricing: A Model Comparison Perspective

by
Lingxiao Zhao

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2020
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, and most of all, I would like to thank Siddhartha Chib for his expertise, assistance, guidance, and patience throughout the process of writing this thesis. Without his help, these papers would not have been possible. I would also thank my co-chair Werner Ploberger and committee member Guofu Zhou for their support, suggestion, and encouragement. Thank you to John Nachbar and Gaetano Antinolfi for their valuable feedback on the manuscript.

I would like to extend my sincere gratitude to exceptional faculty in the Department of Economics, Olin Business School, and Center for High-Performance Computing. I would like to give special thanks to Xiaming Zeng and Malcolm Tobias for their meaningful help.

Finally, I would like to thank my family for their mental and financial support; and to my cat Gulu and my friends, thanks for keeping me company on long walks.

Lingxiao Zhao

Washington University in St. Louis

May 2020

Dedicated to my daughter, Trucy.

Thank you for coming into my life.

I love you with all my heart.

ABSTRACT OF THE DISSERTATION

Essays on Asset Pricing: A Model Comparison Perspective

by

Lingxiao Zhao

Doctor of Philosophy in Economics

Washington University in St. Louis, 2019

Professor Siddhartha Chib, Co-Chair

Professor Werner Ploberger, Co-Chair

In my dissertation, I focus on theoretical and empirical asset pricing from a Bayesian model comparison perspective.

In the first Chapter, revisiting the framework of Barillas and Shanken (2018), BS henceforth, we show that the Bayesian marginal likelihood-based model comparison method in that paper is unsound: the priors on the nuisance parameters across models must satisfy a change of variable property for densities that is violated by the Jeffreys priors used in the BS method. Extensive simulation exercises confirm that the BS method performs unsatisfactorily. We derive a new class of improper priors on the nuisance parameters, *starting from a single improper prior*, which leads to valid marginal likelihoods and model comparisons. The performance of our marginal likelihoods is significantly better, allowing for reliable Bayesian work on which factors are risk factors in asset pricing models.

In the second Chapter, starting from the twelve distinct risk factors in four well-established asset pricing models, a pool we refer to as the winners, we construct and compare 4,095 asset pricing models and find that the model with the risk factors, Mkt, SMB, MOM, ROE, MGMT, and PEAD, performs the best in terms of Bayesian posterior probability, out-of-sample predictability, and Sharpe ratio. A more extensive model comparison of 8,388,607 models, constructed from the

twelve winners plus eleven principal components of anomalies unexplained by the winners, shows the benefit of incorporating information in genuine anomalies in explaining the cross-section of expected equity returns.

Chapter 1

On Comparing Asset Pricing Model

Siddhartha Chib, Xiaming Zeng, and Lingxiao Zhao[1]

In this paper we revisit the framework of Barillas and Shanken (2018), BS henceforth, and show that the Bayesian marginal likelihood-based model comparison method in that paper is unsound. In particular, we show that in this comparison of asset pricing models, in which the nuisance parameters $\{\eta_j\}$ across models are connected by invertible mappings, the priors on the nuisance parameters across models must satisfy a certain change of variable property for densities that is violated by the off-the-shelf Jeffrey' priors used in the BS method. Hence, the BS "marginal likelihoods" each depend on an arbitrary constant, which voids the ranking of models by the size of the marginal likelihoods and invalidates any conclusions drawn from such a method about the underlying data-generating process (DGP). In the online appendix of their paper, BS discuss an alternative method for calculating marginal likelihoods with their improper priors, which they call the permutation method. This more involved method is not used in the paper but, as we show below, it is also unsound and as a result leads to invalid marginal likelihoods.

[1]Siddhartha Chib (corresponding author, chib@wustl.edu) is at the Olin Business School, Washington University in St. Louis. Xiaming Zeng is an Investment Professional. Lingxiao Zhao is at the Department of Economics, Washington University in St. Louis. We are grateful to the Editor (Stefan Nagel) and two anonymous reviewers for their constructive and helpful comments.

We conduct extensive simulation exercises using two experiments. In the first, we match eight potential risk factors to the excess market return (Mkt), size (SMB), value (HML), profitability (RMW) and investment (CMA) factors proposed by Fama and French (1993, 2015), the profitability (ROE) and investment (IA) factors in the q-factor model proposed by Hou, Xue, and Zhang (2015), and the Carhart (1997) momentum (MOM) factor. In the second, we match twelve potential risk factors to the eight factors above as well as the Asness and Frazzini (2013) quality minus junk (QMJ) factor, the Pastor and Stambaugh (2003) liquidity (LIQ) factor, the Frazzini and Pedersen (2014) betting against beta (BAB) factor, and another version of value factor (HMLD) proposed by Asness, Frazzini, and Pedersen (2019). Given the prejudged status of the Mkt factor as a risk factor, we have $2^7 = 128$ candidate models in the first experiment and $2^{11} = 2,048$ candidate models in the second. We repeat our comparison exercises over 100 simulated replications of the data for sample sizes of 600, 1,200 and 12,000, 120,000 and 1.2 million for each of 30 (55) true DGPs in the first (second) experiment. In the first experiment the BS method has some success when the sample size is 1.2 million, but in the second experiment the BS method fails to locate any of the true DGPs even once in 100 replications for any sample size, including the epic sample size of 1.2 million.

In a significant advance, we derive a new class of improper priors on the nuisance parameters, *starting from a single improper prior*, with the property that the improper priors in this class necessarily share the same arbitrary constant c. This class of priors leads to valid marginal likelihoods and, in turn, valid model comparisons. The construction of this class of improper priors is summarized in Proposition 2 below. As we detail, the ability of the resulting marginal likelihoods to pick the true DGPs is significantly better.

We also discuss an extension of our method to the more general class of model comparisons in which the status of the Mkt factor as a risk factor is also in doubt. Chib and Zeng (2019) have recently developed a method for conducting such comparisons that is based on proper priors, each derived from a single proper prior, and student-t distributions of the factors. The approach in this paper, though closely related to that of Chib and Zeng (2019), requires fewer prior inputs, and together pave the way for reliable Bayesian work on which factors are risk factors in asset pricing models.

The rest of the paper is organized as follows. In Section I, we outline the BS method for calculating marginal likelihoods. In Section II we discuss the issues that arise in calculating marginal likelihoods with improper priors, and in Proposition 2 we provide a class of improper priors on nuisance parameters that lead to valid marginal likelihoods. In Section III, we derive the priors and marginal likelihoods that satisfy Proposition 2 (which we refer to as the Chib, Zeng, and Zhao priors and marginal likelihoods) for the problem of comparing asset pricing models. Section IV contains further critical discussion of the BS method, and Section V and VI present results from extensive simulation experiments on the performance of the BS and Chib, Zeng, and Zhao methods, respectively. Section VII concludes. Appendices contain additional details relevant for the discussion in the main text.

1.1 BS Method

In the method of BS, one starts with a collection of K (traded) potential risk factors. The market factor (Mkt) is one of these K factors and is prejudged to be a risk factor. We will relax this assumption in our method below. A particular asset pricing model arises by choosing one or more of the remaining $K-1$ factors as risk factors. The model-space thus contains $J = 2^{(K-1)}$ models. Let $\mathscr{M}_j$, $j = 1,2,...,J$, represents any one of the possible models. It is defined by the vector of risk factors $\{\text{Mkt}, f_j\}$ of size L_j and the vector of non-risk factors f_j^* of size $(K-L_j)$. Note that f is indexed by j because what goes into f is what varies across models. Then, letting t denote a particular point in the sample, $t = 1,2,...,n$, each model in the model-space is given by

$$f_{j,t} = \alpha_j + \beta_j \text{Mkt}_t + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \sim \mathscr{N}_{L_j-1}(0, \Sigma_j)$$

$$f^*_{j,t} = \begin{pmatrix} \beta^*_{j,m} \, B^*_{j,f} \end{pmatrix} \begin{pmatrix} \text{Mkt}_t \\ f_{j,t} \end{pmatrix} + \varepsilon^*_{j,t}, \quad \varepsilon^*_{j,t} \sim \mathcal{N}_{K-L_j}(0, \Sigma^*_j),$$

where an intercept vector is absent from the $f^*_{j,t}$ model because of the pricing restrictions and the error terms $\varepsilon_{j,t}$ and $\varepsilon^*_{j,t}$ are assumed to be mutually independent and independently distributed across t. Lowercase letters denote vectors and uppercase letters matrices (of dimensions that are suppressed for convenience). Let $\beta^*_{j,f} = \mathsf{vec}(B^*_{j,f})$ denote the column-vectorized form of $B^*_{j,f}$, and $\sigma_j = \mathsf{vech}(\Sigma_j)$ and $\sigma^*_j = \mathsf{vech}(\Sigma^*_j)$ the half or unique element vectorizations of the two covariance matrices. Then the parameters of $\mathcal{M}_j$ are

$$\theta_j = (\alpha_j, \beta_j, \beta^*_{j,m}, \beta^*_{j,f}, \sigma_j, \sigma^*_j) \in \Theta_{\theta_j},$$

of which

$$\eta_j = \left(\beta_j, \beta^*_{j,m}, \beta^*_{j,f}, \sigma_j, \sigma^*_j\right)$$

are the nuisance parameters of $\mathcal{M}_j$. We let Θ_{θ_j} and Θ_{η_j} denote the parameter spaces of θ_j and η_j, respectively, these being obvious by context.

BS suppose that the prior density of θ_j is given by

$$p_{BS}(\theta_j|\mathcal{M}_j) = \pi_{BS}(\alpha_j|\mathcal{M}_j, \eta_j)\psi_{BS}(\eta_j|\mathcal{M}_j), \tag{1.1.1}$$

where

$$\pi_{BS}(\alpha_j|\mathcal{M}_j, \eta_j) = \mathcal{N}_{L-1}(\alpha_j|0, k\Sigma_j) \tag{1.1.2}$$

$$\psi_{BS}(\eta_j|\mathcal{M}_j) = |\Sigma_j|^{-L_j/2}|\Sigma^*_j|^{-(K-L_j+1)/2} \tag{1.1.3}$$

and $k > 0$ controls the spread of the prior on α_j. Thus, in this prior $\pi_{BS}(\alpha_j|\mathcal{M}_j, \eta_j)$ is a proper density and $\psi_{BS}(\eta_j|\mathscr{M}_j)$ is an improper density (which comes from Jeffreys rule). The proportionality sign of this improper density is replaced by equality because BS set the constant of proportionality to one.

Under this prior, BS calculate the marginal likelihood of each of the J models. The marginal likelihood is the integral of the sampling density (the likelihood function) with respect to the prior. If we let

$$y_{1:T} = (f_1, f_1^*, ..., f_T, f_T^*)$$

denote the sample data on the factors over T time periods, the marginal likelihood of $\mathscr{M}_j$ is given by the expression

$$m(y_{1:T}|\mathscr{M}_j) \triangleq \int_{\Theta_{\eta_j}} \int_{\Theta_{\alpha_j}} p(y_{1:T}|\mathscr{M}_j, \theta_j)\pi(\alpha_j|\mathscr{M}_j, \eta_j)\psi(\eta_j|\mathscr{M}_j)\mathrm{d}\theta_j, \tag{1.1.4}$$

which because of the independence of the errors and the independence of the priors, can be split into two pieces as follows:

$$m(y_{1:T}|\mathscr{M}_j) = m(f_{1:T}|\mathscr{M}_j)m(f_{1:T}^*|\mathscr{M}_j),$$

where each term on the right-hand side (RHS) is in closed form under the above assumptions. BS take the log of $m(y_{1:T}|\mathscr{M}_j)$, $j = 1, ..., J$, to screen for the best model.

1.2 Marginal Likelihoods with Improper Priors

In general, improper priors invalidate Bayesian model comparisons by marginal likelihoods. An improper prior is one whose integral over the parameter space is not finite. As a result, multiplying an improper density by *any* arbitrary positive constant produces the *same* improper density. In other words, because $\psi_{BS}(\eta_j|\mathscr{M}_j)$ is an improper distribution, $c_j\psi_{BS}(\eta_j|\mathscr{M}_j)$ is the *same* improper prior for any $c_j > 0$. This means that the marginal likelihood is indeterminate since it depends on an arbitrary $c_j > 0$.

Fixing c_j at some value does not (in general) solve the problem because the resulting Bayes factor depends on that choice. Thus, the choice of BS,

$$c_j = 1\ ,\ j = 1, 2, ..., J,$$

is not a panacea. In defense of this choice, in footnote 9 of their paper, BS make a reference to nuisance parameters that are common across models. It is known that improper priors can be used in the calculation of the marginal likelihood for parameters that are common across models *and* that have the same support in each model. To see this, suppose that the nuisance parameters $\eta_j = (\beta_j, \beta^*_{j,m}, \beta^*_{j,f}, \sigma_j, \sigma^*_j)$ do not vary by model, and that their parameter spaces Θ_η are also common across models. In that case,

$$m(y_{1:T}|\mathscr{M}_j) = \int_{\Theta_\eta} \int_{\Theta_{\alpha_j}} p(y_{1:T}|\mathscr{M}_j, \theta_j)\pi(\alpha_j|\mathscr{M}_j, \eta)c\psi(\eta)\mathrm{d}\theta_j. \tag{1.2.1}$$

Thus, in comparing any two models, since the same constant c appears in the prior density of the common nuisance parameters, the constant c cancels out. This simple argument is the basis of the following proposition.

PROPOSITION 1. *If the nuisance parameters are common across models and have the same support in each model, then the collection of marginal likelihoods*

$$\{\ m(y_{1:T}|\mathscr{M}_1)\ , ...,\ m(y_{1:T}|\mathscr{M}_J)\}$$

with a common improper prior on the common nuisance parameter are valid and comparable.

The setting of BS, however, does not correspond to this common parameter-common support case because the nuisance parameters η_j do, in fact, vary by model, and the parameter spaces on which the improper prior is defined also vary by model. This can be easily seen from the model formulation. In the BS method each nuisance parameter is given its own Jeffreys prior that has its own constant ($c_j = 1$), which renders the marginal likelihoods indeterminate.

For improper priors to work, the improper priors must be such that they *necessarily* share the same constant across models. How can one make the different priors share the same constant when the nuisance parameters differ? This can be achieved by taking advantage of the fact that the nuisance parameters $\{\eta_j\}$ in this problem are connected by invertible maps. Chib and Zeng (2019) exploit this feature to derive *proper* priors across models from a single proper prior. In the current context with improper priors, we proceed as follows.

- We first derive the invertible map, as well as the Jacobian of the transformation, that connects the nuisance parameters η_1 of a model that we call $\mathscr{M}_1$, and the nuisance parameters η_j of a generic model that we refer to as $\mathscr{M}_j$.
- Next we give the nuisance parameters η_1 of $\mathscr{M}_1$ a Jeffreys prior.
- Then, for every other model $j > 1$, we derive the prior on η_j by a change of variable from that *single* prior density.

The resulting improper prior densities then *necessarily* share the same constant, which means that marginal likelihoods calculated with these priors are valid and comparable as that common constant appears in each marginal likelihood and, hence, cancels out in taking ratios or log differences. This construction, which is new to the literature, is stated next.

PROPOSITION 2. *Consider a collection of J models $\mathscr{M}_1, \ldots, \mathscr{M}_J$. Suppose that the nuisance parameters η_1 of model $\mathscr{M}_1$ are connected to the nuisance parameters η_j of $\mathscr{M}_j$ $(j > 1)$ by the invertible mapping $\eta_j = g_j(\eta_1)$, with the inverse mapping given by*

$$\eta_1 = g_j^{-1}(\eta_j). \tag{1.2.2}$$

Let

$$c\psi(\eta_1|\mathscr{M}_1)$$

denote an arbitrary chosen improper prior on η_1 *in model* $\mathscr{M}_1$ *with an arbitrary constant c. Let*

$$\tilde{\psi}(\eta_j|\mathscr{M}_j) = c\psi(g_j^{-1}(\eta_j)|\mathscr{M}_1)\left|\mathrm{det}\left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'}\right)\right|,\ j = 2, 3, ..., J, \tag{1.2.3}$$

denote the improper priors obtained by the change of variable formula from the first prior, where the last term is the absolute value of the Jacobian of the transformation. Finally, let

$$m(y_{1:T}|\mathscr{M}_1) = \int_{\Theta_{\eta_1}}\int_{\Theta_{\alpha_1}} p(y_{1:T}|\mathscr{M}_1, \theta_1)\pi(\alpha_1|\mathscr{M}_1, \eta_1)c\psi(\eta_1|\mathscr{M}_1)\mathrm{d}\theta_1$$

and

$$\tilde{m}(y_{1:T}|\mathscr{M}_j) = \int_{\Theta_{\eta_j}}\int_{\Theta_{\alpha_j}} p(y_{1:T}|\mathscr{M}_j, \theta_j)\pi(\alpha_j|\mathscr{M}_j, \eta_j)\tilde{\psi}(\eta_j|\mathscr{M}_j)\mathrm{d}\theta_j \tag{1.2.4}$$

denote the marginal likelihoods of $\mathscr{M}_1$ *and* $\mathscr{M}_j$*,* $j > 1$*, computed using* $c\psi(\eta_1|\mathscr{M}_1)$ *and* $\tilde{\psi}(\eta_j|\mathscr{M}_j)$*, respectively. Then the collection of marginal likelihoods*

$$\{m(y_{1:T}|\mathscr{M}_1),\ \tilde{m}(y_{1:T}|\mathscr{M}_2), ...,\ \tilde{m}(y_{1:T}|\mathscr{M}_J)\}$$

are valid and comparable.

The proof of this proposition is straightforward. Inserting the definition of the improper prior $\tilde{\psi}(\eta_j|\mathscr{M}_j)$ into $\tilde{m}(y_{1:T}|\mathscr{M}_j)$, we get

$$\tilde{m}(y_{1:T}|\mathscr{M}_j) = \int_{\Theta_{\eta_j}}\int_{\Theta_{\alpha_j}} p(y_{1:T}|\mathscr{M}_j, \theta_j)\pi(\alpha_j|\eta_j)c\psi(g_j^{-1}(\eta_j)|\mathscr{M}_1)\left|\mathrm{det}\left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'}\right)\right|\mathrm{d}(\alpha_j, \eta_j). \tag{1.2.5}$$

Since the same constant c appears in the RHS of each marginal likelihood in the collection, the marginal likelihoods are comparable.

It is worth noting that a reader of this paper argued that the priors in the BS collection are valid because the nuisance parameters are connected by invertible maps. As proof of this claim, the reader used a change-of-variable argument. This proof is incorrect, however, because the improper priors in the BS collection do not take advantage of the invertible mapping, but the idea that the change of variable property should play a role is relevant, though, as we have shown, the change of variable property has to be enforced on the priors across models, as it is not an automatic consequence of the invertible mapping between the nuisance parameters.

To emphasize the latter point, what Proposition 2 states is that *the improper priors across models have to be constructed from the prior of one model by the change of variable formula* for densities. Provided one follows this construction, the same constant c appears on the RHS of each marginal likelihood. Any improper prior across models that is not constructed in this way will violate the change of variable condition and hence necessarily entail an arbitrary constant, rendering the marginal likelihood comparison void.

1.3 Improper Priors and Valid Marginal Likelihoods

We now derive the collection of improper priors that respect Proposition 2 and calculate the marginal likelihoods with these priors. To derive the class of priors according to the construction given in Proposition 2, we first derive the invertible map that connects the nuisance parameters η_1 of a model we call $\mathscr{M}_1$ and the nuisance parameters η_j of a generic model that we refer to as $\mathscr{M}_j$. We then derive the Jacobian of the transformation, followed by the prior density of η_j by the construction given in Proposition 2. We refer to the priors and marginal likelihoods that emerge from our method as the Chib, Zeng, and Zhao (CZZ) priors and marginal likelihoods.

1.3.1 Derivation of the CZZ Priors

To facilitate the calculations, we specify the $J = 2^{(K-1)}$ models $\left\{\mathscr{M}_j\right\}_{j=1}^{J}$ as follows.

- $\mathscr{M}_1$ denotes the model in which all K factors are risk factors, following Chib and Zeng (2019),
- $\mathscr{M}_j$, $j = 2, 3, \ldots, J-1$, denotes the models in which $\{\text{Mkt}, f_j\}$ are the risk factors (i.e., f_j is nonempty), and
- $\mathscr{M}_J$ denotes the model in which $\{\text{Mkt}\}$ is the only risk factor (i.e., f_J is empty).

We now apply the construction given in Proposition 2. By definition, $\mathscr{M}_1$ is the model

$$f_{1,t} = \alpha_1 + \beta_1 \text{Mkt}_t + \varepsilon_{1,t}, \quad \varepsilon_{1,t} \sim \mathscr{N}_{K-1}(0, \Sigma_1) \tag{1.3.1}$$

with $f^*_{1,t}$ empty. Let $\sigma_1 = \mathsf{vech}(\Sigma_1)$. Then the nuisance parameters of $\mathscr{M}_1$ are given by

$$\eta_1 = (\beta_1, \sigma_1).$$

Next, consider model $\mathscr{M}_j$, $j = 2, 3, \ldots, J-1$, which we can write as

$$f_{j,t} = \alpha_j + \beta_j \text{Mkt}_t + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \sim \mathscr{N}_{L_j - 1}(0, \Sigma_j) \tag{1.3.2}$$

$$f^*_{j,t} = \begin{pmatrix} \beta^*_{j,m} & B^*_{j,f} \end{pmatrix} \begin{pmatrix} \text{Mkt}_t \\ f_{j,t} \end{pmatrix} + \varepsilon^*_{j,t}, \quad \varepsilon^*_{j,t} \sim \mathscr{N}_{K-L_j}(0, \Sigma^*_j) \tag{1.3.3}$$

with nuisance parameters given by

$$\eta_j = \left(\beta_j, \beta^*_{j,m}, \beta^*_{j,f}, \sigma_j, \sigma^*_j\right).$$

Plugging the model in (1.3.2) into (1.3.3), we get

$$f^*_{j,t} = B^*_{j,f}\alpha_j + \left(\beta^*_{j,m} + B^*_{j,f}\beta_j\right) \text{Mkt}_t + \left(\varepsilon^*_{j,t} + B^*_{j,f}\varepsilon_{j,t}\right). \tag{1.3.4}$$

Vectorising equations (1.3.2) and (1.3.4), we have that

$$\begin{pmatrix} f_{j,t} \\ f^*_{j,t} \end{pmatrix} = \begin{pmatrix} \alpha_j \\ B^*_{j,f}\alpha_j \end{pmatrix} + \begin{pmatrix} \beta_j \\ \beta^*_{j,m} + B^*_{j,f}\beta_j \end{pmatrix} \mathrm{Mkt}_t + \tilde{\varepsilon}_{j,t}, \tag{1.3.5}$$

where

$$\tilde{\varepsilon}_{j,t} \sim \mathscr{N}_{K-1}\left(0, \begin{pmatrix} \Sigma_j & \Sigma_j B^{*\prime}_{j,f} \\ B^*_{j,f}\Sigma_j & \Sigma^*_j + B^*_{j,f}\Sigma_j B^{*\prime}_{j,f} \end{pmatrix}\right).$$

Comparing the parameters of equations (1.3.1) and (1.3.5), we see that the nuisance parameters of $\mathscr{M}_1$ and $\mathscr{M}_j$, $j = 2,3,...,J-1$, are related as follows:

$$\beta_1 = \begin{pmatrix} \beta_j \\ \beta^*_{j,m} + B^*_{j,f}\beta_j \end{pmatrix} \tag{1.3.6}$$

$$\Sigma_1 = \begin{pmatrix} \Sigma_j & \Sigma_j B^{*\prime}_{j,f} \\ B^*_{j,f}\Sigma_j & \Sigma^*_j + B^*_{j,f}\Sigma_j B^{*\prime}_{j,f} \end{pmatrix}, \tag{1.3.7}$$

or in vech form,

$$\sigma_1 = \begin{pmatrix} \sigma_j \\ \left(\Sigma_j \otimes I_{K-L_j}\right)\beta^*_{j,f} \\ \sigma^*_j + \mathrm{vech}\left(B^*_{j,f}\Sigma_j B^{*\prime}_{j,f}\right) \end{pmatrix}. \tag{1.3.8}$$

The set of vector equations in (1.3.6) and (1.3.8) constitute the inverse map $\eta_1 = g_j^{-1}(\eta_j)$. The determinant of the Jacobian of this transformation can now be derived. By derivations given in Appendix B, we have that

$$\left|\det\left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'}\right)\right| = |\Sigma_j|^{K-L_j}.$$

Following the construction in Proposition 2, let the prior on η_1 in model $\mathscr{M}_1$ be the Jeffreys improper prior

$$c\psi(\eta_1|\mathscr{M}_1) = c|\Sigma_1|^{-\frac{K}{2}}.$$

Then, by the rule for the determinant of a partitioned matrix applied to (1.3.7), the prior of η_j in model $\mathscr{M}_j$, $j = 2, 3, ..., J-1$, is

$$\begin{aligned}
\tilde{\psi}(\eta_j|\mathscr{M}_j) &= c\psi(g_j^{-1}(\eta_j)|\mathscr{M}_1)\left|\det\left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'}\right)\right| \\
&= c\left(\det(\Sigma_j)\det\left(\Sigma_j^* + B_{j,f}^*\Sigma_j B_{j,f}^{*\prime} - B_{j,f}^*\Sigma_j\Sigma_j^{-1}\Sigma_j B_{j,f}^{*\prime}\right)\right)^{-\frac{K}{2}}|\Sigma_j|^{K-L_j} \\
&= c|\Sigma_j|^{-\frac{2L_j-K}{2}}|\Sigma_j^*|^{-\frac{K}{2}}.
\end{aligned}$$

Finally, consider model $\mathscr{M}_J$, which can be written as

$$f_{J,t}^* = \beta_{J,m}^*\text{Mkt}_t + \varepsilon_{J,t}^*, \quad \varepsilon_{J,t}^* \sim \mathscr{N}_{K-1}(0, \Sigma_J^*). \tag{1.3.9}$$

This model is just a special case of $\mathscr{M}_j$ $(j \neq 1)$. It can be easily seen that the Jacobian is equal to one, which implies that the prior of η_J in model $\mathscr{M}_J$ is given by

$$\begin{aligned}
\tilde{\psi}(\eta_J|\mathscr{M}_J) &= c\psi(g_J^{-1}(\eta_J)|\mathscr{M}_1)\left|\det\left(\frac{\partial g_J^{-1}(\eta_J)}{\partial \eta_J'}\right)\right| \\
&= c|\Sigma_J^*|^{-\frac{K}{2}}.
\end{aligned}$$

We have thus proved the following new result.

PROPOSITION 3. *Let the first model $\mathscr{M}_1$ in equation (1.3.1) have the improper prior on η_1 given by*

$$c\psi(\eta_1|\mathscr{M}_1) = c|\Sigma_1|^{-\frac{K}{2}},$$

where c is an arbitrary constant. Then the prior of η_j in $\mathscr{M}_j$, $j = 2, 3, .., J-1$, given by

$$\tilde{\psi}(\eta_j|\mathscr{M}_j) = c|\Sigma_j|^{-\frac{2L_j-K}{2}}|\Sigma_j^*|^{-\frac{K}{2}}$$

and that of η_J in $\mathscr{M}_J$ given by

$$\tilde{\psi}(\eta_J|\mathscr{M}_J) = c|\Sigma_J^*|^{-\frac{K}{2}}$$

satisfy Proposition 2 and lead to comparable marginal likelihoods.

The simplicity of this result should be noted.

1.3.2 CZZ Marginal Likelihoods

The valid marginal likelihoods for models $\mathscr{M}_1, ..., \mathscr{M}_J$ can now be derived. We assume that the prior of $\alpha_j|\mathscr{M}_j, \eta_j$ is the same as in (1.1.2). These marginal likelihoods are in closed form for every model in the model-space. As explained in Proposition 2, the constant c is arbitrary. In the expressions below we set it to equal one. We use the identity of the marginal likelihood introduced in Chib (1995) to simplify the computations of the marginal likelihoods. The calculations are tedious but straightforward, and hence are suppressed.

Consider the typical model $\mathscr{M}_j$ $(j \neq 1, J)$. The log marginal likelihood can be split into two pieces (because of the independence of the errors and the independence of the priors) as follows:

$$\log \tilde{m}(y_{1:T}|\mathscr{M}_j) = \log \tilde{m}(f_{j,1:T}|\mathscr{M}_j) + \log \tilde{m}(f^*_{j,1:T}|\mathscr{M}_j), \tag{1.3.10}$$

where the first term on the RHS is

$$\begin{aligned} &-\frac{(K-L_j)(L_j-1)}{2}\log 2 - \frac{(T-1)(L_j-1)}{2}\log \pi - \frac{(L_j-1)}{2}\log k \\ &-\frac{(L_j-1)}{2}\log|W| - \frac{(T+L_j-K-1)}{2}\log|\Psi_j| + \log \Gamma_{L_j-1}\left(\frac{T+L_j-K-1}{2}\right), \end{aligned} \tag{1.3.11}$$

the second term is

$$
\begin{aligned}
&\frac{(K-L_j)(L_j-1)}{2}\log 2-\frac{(K-L_j)(T-L_j)}{2}\log\pi\\
&-\frac{(K-L_j)}{2}\log|W_j^*|-\frac{(T-1)}{2}\log|\Psi_j^*|+\log\Gamma_{K-L_j}\left(\frac{T-1}{2}\right),
\end{aligned}
\tag{1.3.12}
$$

and

$$
\begin{aligned}
X'X &= \sum_{t=1}^{T}(1\ \mathrm{Mkt}_t)'(1\ \mathrm{Mkt}_t)\ ,\ \Lambda^{-1}=\begin{pmatrix}k^{-1} & 0\\ 0 & 0\end{pmatrix}\\
W &= X'X+\Lambda^{-1}\ ,\ W_j^*=\sum_{t=1}^{T}(\mathrm{Mkt}_t\ f'_{j,t})'(\mathrm{Mkt}_t\ f'_{j,t})\\
\Psi_j &= \sum_{t=1}^{T}(f_{j,t}-\hat{\alpha}_j-\hat{\beta}_j\mathrm{Mkt}_t)(f_{j,t}-\hat{\alpha}_j-\hat{\beta}_j\mathrm{Mkt}_t)'+(\hat{\alpha}_j\ \hat{\beta}_j)\left(X'XW^{-1}\Lambda^{-1}\right)\begin{pmatrix}\hat{\alpha}'_j\\ \hat{\beta}'_j\end{pmatrix}\\
\Psi_j^* &= \sum_{t=1}^{T}(f_{j,t}^*-\hat{\beta}_{j,m}^*\mathrm{Mkt}_t-\hat{B}_{j,f}^*f_{j,t})(f_{j,t}^*-\hat{\beta}_{j,m}^*\mathrm{Mkt}_t-\hat{B}_{j,f}^*f_{j,t})'.
\end{aligned}
$$

In these expressions the hat symbol denotes the least square estimates, and $\Gamma_d(\cdot)$ denotes the d-dimensional multivariate gamma function. Finally, for $\mathscr{M}_1$ the log marginal likelihood is given by (1.3.11), and for $\mathscr{M}_J$ it is given by (1.3.12).

The computations typically take a few seconds to scan our model-space of 2,048 models in the twelve-factor case.

1.3.3 CZZ Marginal Likelihoods: General Case

We briefly note that our method can be extended in two directions: (1) to the more general class of asset pricing model comparisons where the status of the Mkt factor as a risk factor is also in doubt, as in the recent work of Chib and Zeng (2018) where marginal likelihoods are computed based on proper priors and student-t distributions of the factors, and (2) to the case in which the intercept term in the model of the risk factors is given a model-specific prior. This second extension is also motivated by the work of Chib and Zeng (2018).

Let $\tilde{f}$ denote the set of risk factors, and let f^*denote the set of non-risk factors. The model that we describe here differs from those above because Mkt can now enter into $\tilde{f}$ or f^*. Also note that $\tilde{f}$ can never be empty, which means that the total number of models in the model-space is given by $\tilde{J} = 2^K - 1$. As above, suppose that in $\mathscr{M}_1$ all K factors are risk factors,

$$\tilde{f}_{1,t} = \tilde{\alpha}_1 + \tilde{\varepsilon}_{1,t}, \quad \tilde{\varepsilon}_{1,t} \sim \mathscr{N}_K\left(0, \Sigma_1\right). \tag{1.3.13}$$

Let $\sigma_1 = \mathsf{vech}(\Sigma_1)$. Then the nuisance parameters of $\mathscr{M}_1$ are simply

$$\eta_1 = \sigma_1 \, .$$

In model $\mathscr{M}_j$, $j = 2, 3, ..., \tilde{J}$, let $\tilde{f}_j$ denote the risk factors with dimension $L_j \times 1$ and let f_j^* denote the non-risk factors with dimension $(K - L_j) \times 1$. This model is given by

$$\tilde{f}_{j,t} = \tilde{\alpha}_j + \tilde{\varepsilon}_{j,t}, \quad \tilde{\varepsilon}_{j,t} \sim \mathscr{N}_{L_j}\left(0, \Sigma_j\right) \tag{1.3.14}$$

$$f^*_{j,t} = B^*_{j,f} \tilde{f}_{j,t} + \varepsilon^*_{j,t}, \quad \varepsilon^*_{j,t} \sim \mathscr{N}_{K-L_j}\left(0, \Sigma^*_j\right) \tag{1.3.15}$$

with nuisance parameters

$$\eta_j = \left(\beta^*_{j,f}, \sigma_j, \sigma^*_j\right),$$

where $\beta^*_{j,f} = \mathsf{vec}(B^*_{j,f})$, $\sigma_j = \mathsf{vech}(\Sigma_j)$, and $\sigma^*_j = \mathsf{vech}(\Sigma^*_j)$. By calculations that we suppress, we can prove the following result.

PROPOSITION 4. *Let model $\mathscr{M}_1$ in equation (1.3.13) have the improper prior on η_1 given by*

$$c\psi(\eta_1 | \mathscr{M}_1) = c|\Sigma_1|^{-\frac{K+1}{2}},$$

where c is an arbitrary constant. Then the priors of η_j in $\mathscr{M}_j$, $j = 2,3,..,\tilde{J}$, given by

$$\tilde{\psi}(\eta_j|\mathscr{M}_j) = c|\Sigma_j|^{-\frac{2L_j-K+1}{2}}|\Sigma_j^*|^{-\frac{K+1}{2}}$$

satisfy Proposition 2 and lead to comparable marginal likelihoods.

Next, instead of supposing that $\tilde{\alpha}_j$ has a $\mathscr{N}_{L_j}(0, k\Sigma_j)$ prior in which the mean vector is zero, and that the constant k is common across models, we suppose that $\tilde{\alpha}_j$ has the model-specific prior

$$\tilde{\alpha}_j|\mathscr{M}_j \sim \mathscr{N}_{L_j}(\tilde{\alpha}_{j0}, k_j\Sigma_j)\ ,\ j = 1,2,...,\tilde{J},$$

where the prior mean $\tilde{\alpha}_{j0}$ and the multiplier k_j are determined from a training sample (a sample of data prior to the sample used for the model comparisons). In our applications, the training sample consists of the first $tr = 0.1$ (tenth) of the data. If we let $n_t = tr \times T$ denote the size of the this training sample data, then

$$\tilde{\alpha}_{j0} = n_t^{-1}\sum_{t=1}^{n_t}\tilde{f}_{j,t}\ , \tag{1.3.16}$$

which is the average of the risk factors in the training sample data. To determine the model-specific multiplier k_j, we calculate $\hat{\Sigma}_{j0}$, the least square estimate of Σ_j in the training sample, and V_{j0}, the negative inverse Hessian over $\tilde{\alpha}_j$, from the log of the marginal likelihood of the training sample observations $\tilde{f}_{1:n_t}$ (conditioned on $\tilde{\alpha}_j$ but marginalized over Σ_j):

$$\log\tilde{m}(\tilde{f}_{1:n_t}|\mathscr{M}_j, \tilde{\alpha}_j) = \log\int p(\tilde{f}_{1:n_t}|\mathscr{M}_j, \tilde{\alpha}_j, \Sigma_j)\pi(\Sigma_j|\mathscr{M}_j)\mathrm{d}\Sigma_j.$$

After omitting terms that do not involve $\tilde{\alpha}_j$, the above expression can be written as

$$-\frac{(n_t + L_j - K)}{2}\log\det\left(\sum_{t=1}^{n_t}(\tilde{f}_{j,t} - \tilde{\alpha}_j)(\tilde{f}_{j,t} - \tilde{\alpha}_j)'\right).$$

The Hessian matrix (a $L_j \times L_j$ matrix) of the latter function can be computed numerically. Our choice of k_j is the average of the (element-by-element) ratio of the diagonal elements of V_{j0} and $\hat{\Sigma}_{j0}$,

$$k_j = mult \times L_j^{-1} \mathsf{sum}\left(\mathsf{diag}\left(V_{j0}\right)/\mathsf{diag}\left(\hat{\Sigma}_{j0}\right)\right)\ ,\ j = 1,2,...,\tilde{J}, \tag{1.3.17}$$

where $mult = \frac{1-tr}{tr}$ is a multiplier that adjusts for the different sizes of the training and estimation samples. We can now prove the following proposition about the marginal likelihoods for the estimation sample.

PROPOSITION 5. *Under the collection of priors in Proposition 4, with c set equal to one, and* $\tilde{\alpha}_j|\mathscr{M}_j \sim \mathscr{N}_{L_j}(\tilde{\alpha}_{j0}, k_j\Sigma_j)$*, the marginal likelihood of model* $\mathscr{M}_j$*,* $j = 2,3,...,\tilde{J}$*, on the log-scale is given by*

$$\log \tilde{m}(y_{n_t+1:T}|\mathscr{M}_j) = \log \tilde{m}(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j) + \log \tilde{m}(f^*_{j,n_t+1:T}|\mathscr{M}_j), \tag{1.3.18}$$

where the first term on the RHS is

$$\begin{aligned} &-\frac{(K-L_j)L_j}{2}\log 2 - \frac{\tilde{T}L_j}{2}\log\pi - \frac{L_j}{2}\log(\tilde{T}k_j+1) \\ &-\frac{\left(\tilde{T}+L_j-K\right)}{2}\log|\Psi_j| + \log\Gamma_{L_j}\left(\frac{\tilde{T}+L_j-K}{2}\right)\ , \end{aligned} \tag{1.3.19}$$

the second term is

$$\begin{aligned} &\frac{(K-L_j)L_j}{2}\log 2 - \frac{(K-L_j)(\tilde{T}-L_j)}{2}\log\pi \\ &-\frac{\left(K-L_j\right)}{2}\log|W^*_j| - \frac{\tilde{T}}{2}\log|\Psi^*_j| + \log\Gamma_{K-L_j}\left(\frac{\tilde{T}}{2}\right), \end{aligned} \tag{1.3.20}$$

and the log marginal likelihood of $\mathscr{M}_1$ *defined in (1.3.13) is given by (1.3.19). In these expressions,* $\tilde{T} = (T-n_t)$ *and*

$$\Psi_j = \sum_{t=n_t+1}^{T}(\tilde{f}_{j,t}-\hat{\tilde{\alpha}}_j)(\tilde{f}_{j,t}-\hat{\tilde{\alpha}}_j)' + \frac{\tilde{T}}{\tilde{T}k_j+1}\left(\hat{\tilde{\alpha}}_j-\tilde{\alpha}_{j0}\right)\left(\hat{\tilde{\alpha}}_j-\tilde{\alpha}_{j0}\right)'$$

$$W_j^* = \sum_{t=n_t+1}^{T} \tilde{f}_{j,t}\tilde{f}_{j,t}' \ , \ \Psi_j^* = \sum_{t=n_t+1}^{T} (f_{j,t}^* - \hat{B}_{j,f}^* \tilde{f}_{j,t})(f_{j,t}^* - \hat{B}_{j,f}^* \tilde{f}_{j,t})'.$$

As above, the hat symbol denotes the least square estimates, but now calculated using the data beyond the training sample, and $\Gamma_d(\cdot)$ denotes the d-dimensional multivariate gamma function. We emphasize that these marginal likelihoods correspond to the more general model comparison problem, where the status of the Mkt factor as a risk factor is also in doubt. Although we do not report any results in this paper from applying Proposition 5, our experiments show that the model-specific prior $\tilde{\alpha}_j|\mathscr{M}_j \sim \mathscr{N}_{L_j}(\tilde{\alpha}_{j0}, k_j\Sigma_j)$ produces performance gains of up to 20%, for smaller sample sizes, compared to the marginal likelihoods of Proposition 5 based on $\tilde{\alpha}_j|\mathscr{M}_j \sim \mathscr{N}_{L_j}(0, k\Sigma_j)$. Thus, it is our recommendation that future work using our method rely not only on the general model given here but also on the model-specific prior defined by (1.3.16) and (1.3.17).

1.4 Further Comments about the BS Method

It is clear from our Proposition 3 that the off-the-shelf BS Jeffreys priors are different from the priors dictated by Proposition 2, in particular, the BS method's priors involve arbitrary constants that do not cancel out in the calculation of the marginal likelihoods. The reason is that the BS method uses separate Jeffreys priors that are unrelated to the across-models change of variable formula that is used in the construction of Proposition 2. In fact, BS do not derive the general mapping between pairs of models, nor do they derive the general form of the Jacobian of the transformation. Without this information, the required collection of improper priors given in our Proposition 3 cannot be constructed.

1.4.1 Example

Consider an example with three factors, say, the excess market return (Mkt), size (SMB), and value (HML) factors. In this case, there are four possible pricing models that need to be compared simultaneously. Let us consider two of these four models.

In the first model, $\mathscr{M}_1$, suppose that all three factors $\{$Mkt, HML SMB$\}$ are the risk factors. The factor model is now

$$\underbrace{\begin{pmatrix} \mathrm{HML}_t \\ \mathrm{SMB}_t \end{pmatrix}}_{f_t^*:2\times 1} = \underbrace{\begin{pmatrix} \alpha_{1,h} \\ \alpha_{1,s} \end{pmatrix}}_{\alpha_1^*:2\times 1} + \underbrace{\begin{pmatrix} \beta_{1,hm} \\ \beta_{1,sm} \end{pmatrix}}_{\beta_1^*:2\times 1} \mathrm{Mkt}_t + \varepsilon_t^*, \quad \varepsilon_t \sim \mathscr{N}\left(0, \underbrace{\begin{pmatrix} \sigma_{1,h}^2 & \sigma_{1,hs} \\ \sigma_{1,hs} & \sigma_{1,s}^2 \end{pmatrix}}_{\Sigma_1:2\times 2}\right). \tag{1.4.1}$$

In this case, the nuisance parameters are

$$\eta_1 = \left(\beta_{1,hm}, \beta_{1,sm}, \sigma_{1,h}^2, \sigma_{1,hs}, \sigma_{1,s}^2\right)\ , \tag{1.4.2}$$

which is of size five. From (1.1.3), we have that

$$\psi_{BS}(\eta_1|\mathscr{M}_1) = \frac{1}{\left(\sigma_{1,h}^2\sigma_{1,s}^2 - \sigma_{1,hs}^2\right)^{3/2}}. \tag{1.4.3}$$

In the second model, $\mathscr{M}_2$, suppose that $\{\mathrm{Mkt}, \mathrm{HML}\}$ are the risk factors. In this case, the factor model is given by

$$\underbrace{\mathrm{HML}_t}_{f_t:1\times 1} = \underbrace{\alpha_{2,h}}_{\alpha_2:1\times 1} + \underbrace{\beta_{2,hm}}_{\beta_2:1\times 1} \mathrm{Mkt}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathscr{N}(0, \underbrace{\sigma_{2,h}^2}_{\Sigma_2:1\times 1}) \tag{1.4.4}$$

$$\underbrace{\mathrm{SMB}_t}_{f_t^*:1\times 1} = \underbrace{\begin{pmatrix} \beta_{2,sm}^* & \beta_{2,sh}^* \end{pmatrix}}_{\beta_2^*:1\times 2} \begin{pmatrix} \mathrm{Mkt}_t \\ \mathrm{HML}_t \end{pmatrix} + \varepsilon_t^*, \quad \varepsilon_t^* \sim \mathscr{N}(0, \underbrace{\sigma_{2,s}^{*2}}_{\Sigma_2^*:1\times 1}), \tag{1.4.5}$$

where the first subscript of the parameter indicates the model. The specification for SMB has no intercept term due to the pricing restrictions. The set of nuisance parameters in this model is of size five and is given by

$$\eta_2 = \left(\beta_{2,hm}, \beta^*_{2,sm}, \beta^*_{2,sh}, \sigma^2_{2,h}, \sigma^{*2}_{2,s}\right). \tag{1.4.6}$$

The prior density of η_2 from (1.1.3) is

$$\psi_{BS}(\eta_2|\mathscr{M}_2) = \frac{1}{\sigma^2_{2,h}} \frac{1}{\sigma^{*2}_{2,s}}. \tag{1.4.7}$$

The inverse mapping $g_2^{-1}(\cdot)$ in (1.2.2) can be derived by substituting the model of HML into that of SMB in model $\mathscr{M}_2$ and comparing terms with those in model $\mathscr{M}_1$. By elementary algebra we get

$$\begin{aligned}
\beta_{1,hm} &= \beta_{2,hm} && (1.4.8)\\
\beta_{1,sm} &= \beta^*_{2,sm} + \beta^*_{2,sh}\beta_{2,hm} \\
\sigma^2_{1,h} &= \sigma^2_{2,h} \\
\sigma_{1,hs} &= \beta^*_{2,sh}\sigma^2_{2,h} \\
\sigma^2_{1,s} &= \beta^{*2}_{2,sh}\sigma^2_{2,h} + \sigma^{*2}_{2,s}, && (1.4.9)
\end{aligned}$$

where

$$\left|\det\left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_2'}\right)\right| = \sigma^2_{2,h}. \tag{1.4.10}$$

We can now easily check that the BS prior of η_2 in $\mathscr{M}_2$, given in (1.4.7) above, is not equal to the required prior in (1.2.3):

$$\begin{aligned}
\psi_{BS}(\eta_2|\mathscr{M}_2) = \frac{1}{\sigma^2_{2,h}} \frac{1}{\sigma^{*2}_{2,s}} && (1.4.11)\\
\neq \tilde{\psi}(\eta_2|\mathscr{M}_2) = \psi_{BS}(g_j^{-1}(\eta_j)|\mathscr{M}_1)\sigma^2_{2,h} && (1.4.12)\\
= \frac{1}{\left(\sigma^2_{2,h}\sigma^{*2}_{2,s}\right)^{\frac{3}{2}}}\sigma^2_{2,h} = \frac{1}{\sigma_{2,h}\sigma^{*3}_{2,s}}. && (1.4.13)
\end{aligned}$$

Therefore, because $\psi_{BS}(\eta_2|\mathscr{M}_2) \neq \tilde{\psi}(\eta_2|\mathscr{M}_2)$, Proposition 2 is violated.

1.4.2 Permutation Method

A reader of our paper argued that the method given by BS in the appendix of their paper is immune to the flaw discussed above. This method, which is called the permutation method, is more involved and is not used by BS in the analysis given in their paper. We show that, unfortunately, the permutation method also involves arbitrary constants that do not cancel out.

Consider a three-factor world consisting of Mkt, SMB, and HML. Since Mkt is always a risk factor, there are $2! = 2$ possible permutations. In the first permutation the factors are ordered as $\mathscr{P}_1 = \{\text{Mkt, HML, SMB}\}$, and in the second they are ordered as $\mathscr{P}_2 = \{\text{Mkt, SMB, HML}\}$. Under $\mathscr{P}_1$, three nested models can be shown to arise by suitably restricting the parameters of the model

$$\text{HML}_t = a + b\text{Mkt}_t + e \tag{1.4.14}$$

$$\text{SMB}_t = c + d\text{Mkt}_t + g\text{HML}_t + u. \tag{1.4.15}$$

For instance, the model $\mathscr{M}_1|\mathscr{P}_1$ (Mkt, HML, SMB are risk factors) arises by setting $a \neq 0$ and $c \neq 0$, $\mathscr{M}_2|\mathscr{P}_1$ (Mkt, HML are risk factors) arises by setting $a \neq 0$ and $c = 0$, and $\mathscr{M}_3|\mathscr{P}_1$ (Mkt is the only risk factor) arises by setting $a = 0$ and $c = 0$. Since these models are nested, they share the same nuisance parameters. Proposition 1 applies and, for example, the constant c_1 (here the subscript 1 denotes the first permutation) can be carried through for the computation of the marginal likelihoods of these models. Under $\mathscr{P}_2$, three nested models can be shown to arise by suitably restricting the parameters of the model

$$\text{SMB}_t = a' + b'\text{Mkt}_t + e' \tag{1.4.16}$$

$$\text{HML}_t = c' + d'\text{Mkt}_t + g'\text{SMB}_t + u'. \tag{1.4.17}$$

For instance, the model $\mathscr{M}_1|\mathscr{P}_2$ (Mkt,SMB, HML are risk factors) arises by setting $a' \neq 0$ and $c' \neq 0$, $\mathscr{M}_2|\mathscr{P}_2$ (Mkt, SMB are risk factors) arises by setting $a' \neq 0$ and $c = 0$, and $\mathscr{M}_3|\mathscr{P}_2$ (Mkt is the only risk factor) arises by setting $a' = 0$ and $c' = 0$. Again, in comparing these three models, the constant c_2 can be used in the improper prior because this situation corresponds to that of Proposition 1. Notice also that $\mathscr{M}_1|\mathscr{P}_1$ and $\mathscr{M}_1|\mathscr{P}_2$ are the same model, as are $\mathscr{M}_3|\mathscr{P}_1$ and $\mathscr{M}_3|\mathscr{P}_2$. Thus, according to Proposition 1, one can replace c_2 with c_1 in calculating the marginal likelihoods of $\mathscr{M}_1|\mathscr{P}_2$ and $\mathscr{M}_3|\mathscr{P}_2$.

The problem arises in this method in comparing the two distinct models $\mathscr{M}_2|\mathscr{P}_1$ and $\mathscr{M}_2|\mathscr{P}_2$, which lie in two different permutations, because these are not nested by the same model. We are now in the situation corresponding to Proposition 2. The nuisance parameters of these models can be linked, but as we know the Jeffreys prior for BS for the parameters in $\mathscr{M}_2|\mathscr{P}_2$ will not satisfy the change of variable condition, which invalidates the marginal likelihood comparison. In other words, the hidden constants c_1 and c_2, which are not relevant in comparing the models within a given permutation, now do not cancel out, invalidating the comparison of models $\mathscr{M}_2|\mathscr{P}_1$ and $\mathscr{M}_2|\mathscr{P}_2$. The problem gets worse as the number of factors increase. For example, with 12 factors, there are $11! = 39,916,800$ possible permutations and numerous models across those permutations for which the BS priors across permutations violate the change of variable condition. Numerical experiments confirm that, besides being numerically unwieldy, the permutation method suffers from the same performance issues as the method used by BS in their paper. Thus, both methods, the one used by BS in their paper and the one mentioned in their online appendix, are unsound and cannot be used to find risk factors in asset pricing. To avoid duplication in our findings, however, in the next section we maintain our focus on the method that is used by BS in their paper.

1.5 Performance of BS Method

In their paper BS do not provide simulation evidence on the performance of their marginal likelihoods to screen for the correct model. Rectifying this omission is the first order of business. We construct two experiments that mimic real-world factors and situations, apply the BS method for different true DGPs and sample sizes, and report on what we find. The first experiment involves eight factors and a model-space of $J = 128$ models. This is a relatively small-scale problem that should be easy to get right. In this case, we consider 33 DGPs to ensure that our results are not specific to one particular DGP in the model-space. For each DGP, we run the experiment 100 times for several sample sizes, which go up to $T = 1.2$ million. We then record the percentage of times (in those 100 replications) that the true DGP is selected by the BS marginal likelihood. The second experiment follows the same approach for $K = 12$ factors and an associated model-space of $J = 2,048$ models. In this setting, we consider 55 DGPs. For each of these 55 DGPs, we run the experiment 100 times, calculating the marginal likelihood of each of the 2,048 models, and we record the percentage of times the true DGP is selected. These experiments are again conducted for different values of T, where start from $T = 600$ and go up to $T = 1.2$ million. We also subject our new method to the same set of experiments.

1.5.1 Eight-Factor Experiment: $J = 128$

In our first experiment, we consider a problem with eight factors. Our simulations proceed as follows. We match eight factors to the excess market return (Mkt), size (SMB), value (HML), profitability (RMW), and investment (CMA) factors proposed by Fama and French (1993, 2015), the profitability (ROE) and investment (IA) factors in the q-factor model proposed by Hou, Xue, and Zhang (2015), and the Carhart (1997) momentum (MOM) factor. In this setting, there are $2^7 = 128$ possible models depending on the assumption made about the collection of factors that go into f_t (the Mkt factor always being included as one of those factors). To ensure that the results do not depend on a particular DGP, we consider 33 different DGPs for generating the data. For each DGP, we generate 100 replicated data sets. For each of these data sets, we calculate the BS "marginal likelihoods" of the 128 candidate models to see if the true DGP is selected. We repeat these steps for each of the 33 DGPs.

In Table 1.1 we report the percentage of times (in 100 replications of data for each true model) that the true DGP is selected for sample sizes of size T = 600, 1,200, 12,000, 120,000, and 1,200,000 based on the "marginal likelihood" criterion of BS.

The true DGPs are listed by row, and following BS, the value of k in equation (1.1.2) is given by

$$k = (\mathrm{Sh}^2_{\mathsf{max}} - \mathrm{Sh}(\mathrm{Mkt})^2)/7, \tag{1.5.1}$$

where Sh(Mkt) is the Sharpe ratio of the simulated Mkt factor, $\mathrm{Sh}_{\mathsf{max}} = \tau \times \mathrm{Sh}(\mathrm{Mkt})$, and τ is set to 3. We have also tried other values of τ mentioned in BS: 1.25, 1.5, and 2. In each case the associated selection percentages of these are no higher than those with $\tau = 3$. The "marginal likelihood" approach of BS does not select any of the 33 true models even once in 100 replications for samples up to $T = 12,000$, which corresponds to a thousand years of data. The method has some success in detecting a few DGPs for the two largest samples sizes, but this success pertains to sample sizes that are unattainable in practice. One does not observe even this limited success in the next set of experiments with twelve potential risk factors.

1.5.2 Twelve-Factor Experiment $J = 2,048$

The performance of the "marginal likelihood" method of BS worsens as the model-space is enlarged. To illustrate this point, we provide extensive results from our second experiment with twelve potential risk factors. The overall experiment and implementation are similar to those for the eight-factor experiment.

We match our twelve factors to the eight factors in the first experiment, as well as the Asness

This table reports the simulation results on the performance of the BS method with eight potential risk factors. The model-space consists of $J = 128$ models. Each row represents a particular DGP for generating the data. Numerical entries are the percentage of times the true DGP is selected among the 128 candidate models in a repeated sampling experiment, for each of five different sample sizes (indicated by column) and for each of 33 different DGPs (indicated by row). Following BS, $k = (\mathsf{Sh}^2_{\mathsf{max}} - \mathrm{Sh}(\mathrm{Mkt})^2)/7$, where Sh refers to the sharp ratio and $\mathrm{Sh}_{\mathsf{max}} = 3 \times \mathrm{Sh}(\mathrm{Mkt})$.

Risk factors in the true model	Barillas and Shanken (2018) $T =$ 600	$T =$ 1,200	$T =$ 12,000	$T =$ 120,000	$T =$ 1,200,000
Mkt SMB RMW IA MOM	0	0	0	0	0
Mkt RMW IA MOM	0	0	0	0	0
Mkt IA MOM	0	0	0	0	0
Mkt HML ROE MOM	0	0	0	0	25
Mkt SMB RMW CMA MOM	0	0	0	0	0
Mkt RMW CMA MOM	0	0	0	0	0
Mkt CMA MOM	0	0	0	0	0
Mkt SMB HML RMW MOM	0	0	0	0	17
Mkt HML RMW MOM	0	0	0	0	17
Mkt SMB RMW MOM	0	0	0	36	92
Mkt RMW MOM	0	0	0	41	94
Mkt HML MOM	0	0	0	0	52
Mkt MOM	0	0	0	60	94
Mkt SMB ROE IA	0	0	0	0	0
Mkt ROE IA	0	0	0	0	0
Mkt SMB RMW IA	0	0	0	0	0
Mkt RMW IA	0	0	0	0	0
Mkt IA	0	0	0	0	0
Mkt SMB CMA ROE	0	0	0	0	0
Mkt CMA ROE	0	0	0	0	0
Mkt SMB HML ROE	0	0	0	0	32
Mkt HML ROE	0	0	0	0	38
Mkt SMB ROE	0	0	0	31	93
Mkt ROE	0	0	0	31	91
Mkt SMB RMW CMA	0	0	0	0	0
Mkt RMW CMA	0	0	0	0	0
Mkt CMA	0	0	0	0	0
Mkt SMB HML RMW	0	0	0	0	30
Mkt HML RMW	0	0	0	0	38
Mkt SMB RMW	0	0	0	42	92
Mkt RMW	0	0	0	43	96
Mkt HML	0	0	0	0	61
Mkt	0	0	0	60	97

Table 1.1

and Frazzini (2013) quality minus junk (QMJ) factor, the Pastor and Stambaugh (2003) liquidity (LIQ) factor, the Frazzini and Pedersen (2014) betting against beta (BAB) factor, and another version of value factor (HMLD) proposed by Asness et al. (2019). We now have $2^{11} = 2,048$ possible models depending on the assumption made about the collection of factors that go into f_t. As in our first experiment, the parameters of the DGP are fixed at the maximum likelihood

This table reports the simulation results on the performance of the BS method with twelve potential risk factors. The model-space consists of J = 2,048 models.

Risk factors in the true model	Barillas and Shanken (2018)				
	$T =$ 600	$T =$ 1,200	$T =$ 12,000	$T =$ 120,000	$T =$ 1,200,000
Mkt SMB RMW IA MOM	0	0	0	0	0
Mkt RMW IA MOM	0	0	0	0	0
Mkt IA MOM	0	0	0	0	0
Mkt HML ROE MOM	0	0	0	0	0
Mkt SMB RMW CMA MOM	0	0	0	0	0
Mkt RMW CMA MOM	0	0	0	0	0
Mkt CMA MOM	0	0	0	0	0
Mkt SMB HML RMW MOM	0	0	0	0	0
Mkt HML RMW MOM	0	0	0	0	0
Mkt SMB RMW MOM	0	0	0	0	0
Mkt RMW MOM	0	0	0	0	0
Mkt HML MOM	0	0	0	0	0
Mkt MOM	0	0	0	0	0
Mkt SMB ROE IA	0	0	0	0	0
Mkt ROE IA	0	0	0	0	0
Mkt SMB RMW IA	0	0	0	0	0
Mkt RMW IA	0	0	0	0	0
Mkt IA	0	0	0	0	0
Mkt SMB CMA ROE	0	0	0	0	0
Mkt CMA ROE	0	0	0	0	0
Mkt SMB HML ROE	0	0	0	0	0
Mkt HML ROE	0	0	0	0	0
Mkt SMB ROE	0	0	0	0	0
Mkt ROE	0	0	0	0	0
Mkt SMB RMW CMA	0	0	0	0	0
Mkt RMW CMA	0	0	0	0	0
Mkt CMA	0	0	0	0	0
Mkt SMB HML RMW	0	0	0	0	0
Mkt HML RMW	0	0	0	0	0
Mkt SMB RMW	0	0	0	0	0
Mkt RMW	0	0	0	0	0
Mkt HML	0	0	0	0	0
Mkt	0	0	0	0	0
Mkt QMJ	0	0	0	0	0
Mkt LIQ	0	0	0	0	0
Mkt BAB	0	0	0	0	0
Mkt HMLD	0	0	0	0	0
Mkt MOM QMJ	0	0	0	0	0
Mkt IA QMJ	0	0	0	0	0
Mkt QMJ HMLD	0	0	0	0	0
Mkt MOM QMJ HMLD	0	0	0	0	0
Mkt MOM QMJ LIQ	0	0	0	0	0
Mkt CMA MOM LIQ	0	0	0	0	0
Mkt IA MOM QMJ	0	0	0	0	0
Mkt CMA LIQ BAB	0	0	0	0	0
Mkt SMB HML RMW QMJ	0	0	0	0	0
Mkt RMW CMA MOM BAB	0	0	0	0	0
Mkt CMA ROE BAB HMLD	0	0	0	0	0
Mkt RMW QMJ BAB HMLD	0	0	0	0	0
Mkt RMW CMA MOM LIQ BAB	0	0	0	0	0
Mkt RMW ROE QMJ BAB HMLD	0	0	0	0	0
Mkt HML RMW IA MOM BAB HMLD	0	0	0	0	0
Mkt HML RMW IA MOM BAB HMLD	0	0	0	0	0
Mkt HML RMW IA MOM BAB HMLD	0	0	0	0	0
Mkt HML ROE IA LIQ BAB HMLD	0	0	0	0	0

Table 1.2

(ML) values to ensure that the generated data resemble the real data. After checking the statistical significance of the risk factors for each of the 2,048 possible models using methods described in Appendix A, there are 567 models that are possible DGPs in this setting.

To conserve space, we do not report results for each of the 567 DGPs (this information is available from us up on request). For each of the 567 DGPs, we generate 100 data sets for a total of 56,700 data sets. For each of these data sets we calculate 2,048 "marginal likelihoods" using the method of BS, one for each of the 2,048 possible models, and then record the percentage of times the true DGP is selected. The results show that the BS method does not select the correct model even once across the 567 DGPs for any sample size, including the sample size of 1.2 million. Table 1.2 reports the results for 55 of the 567 DGPs, where 33 of these DGPs are the same as those in the first experiment above, and 22 are with the new factors included in the current experiment.

1.6 Performance of the CZZ Method

For comparison, we replicate the above set of experiments using the same set of DGPs and the same data sets, based on the CZZ marginal likelihood method. The results reported in Tables 1.3 and 1.4, for the eight-factor and twelve-factor experiments, respectively, show that the performance of the CZZ method is significantly better even when confronted with meager sample sizes of $T = 600$ and $1,200$. These results demonstrate clearly the performance gains from using the CZZ priors and marginal likelihoods.

1.7 Conclusion

In this paper, we show that the "marginal likelihood" approach of Barillas and Shanken (2018) is unsound on account of its reliance on off-the-shelf Jeffreys improper priors on model-specific nuisance parameters. These priors do not satisfy the required *across-models* change of variable

This table reports the simulation performance of the CZZ method with eight potential risk factors. The model-space consists of $J = 128$ models. Each row represents a particular DGP for generating the data. Numerical entries are the percentage of times the true DGP is selected among the 128 candidate models in a repeated sampling experiment, for each of five different sample sizes (indicated by column) and for each of 33 different DGPs (indicated by row). Following BS, $k = (\mathrm{Sh}^2_{\mathsf{max}} - \mathrm{Sh}(\mathrm{Mkt})^2)/7$, where Sh refers to the sharp ratio and $\mathrm{Sh}_{\mathsf{max}} = 3 \times \mathrm{Sh}(\mathrm{Mkt})$.

	CZZ method		
Risk factors in the true model	$T = 600$	$T = 1{,}200$	$T = 12{,}000$
Mkt SMB RMW IA MOM	62	80	91
Mkt RMW IA MOM	53	72	91
Mkt IA MOM	55	69	85
Mkt HML ROE MOM	45	65	91
Mkt SMB RMW CMA MOM	57	70	96
Mkt RMW CMA MOM	59	74	88
Mkt CMA MOM	51	76	86
Mkt SMB HML RMW MOM	54	75	96
Mkt HML RMW MOM	53	71	89
Mkt SMB RMW MOM	49	61	91
Mkt RMW MOM	49	68	88
Mkt HML MOM	54	70	87
Mkt MOM	51	68	83
Mkt SMB ROE IA	61	68	91
Mkt ROE IA	51	69	85
Mkt SMB RMW IA	61	74	90
Mkt RMW IA	57	69	87
Mkt IA	48	60	84
Mkt SMB CMA ROE	59	69	91
Mkt CMA ROE	51	69	84
Mkt SMB HML ROE	65	76	90
Mkt HML ROE	51	72	86
Mkt SMB ROE	56	75	93
Mkt ROE	49	71	84
Mkt SMB RMW CMA	50	66	92
Mkt RMW CMA	52	68	87
Mkt CMA	51	65	83
Mkt SMB HML RMW	51	69	90
Mkt HML RMW	46	69	84
Mkt SMB RMW	42	61	83
Mkt RMW	52	67	84
Mkt HML	54	66	83
Mkt	49	64	89

Table 1.3

formula, formulated in our Proposition 2, and hence depend on arbitrary constants that invalidate the model comparison by marginal likelihoods.

In a notable advance, we derive a new class of improper priors on the nuisance parameters that follow the construction given in Proposition 2 and hence lead to valid marginal likelihoods and model comparisons. The empirical performance of our new marginal likelihoods is significantly

This table reports the simulation performance of the CZZ method with twelve potential risk factors. The model-space consists of J = 2,048 models.

	CZZ method		
Risk factors in the true model	$T = 600$	$T = 1{,}200$	$T = 12{,}000$
Mkt SMB RMW IA MOM	36	47	79
Mkt RMW IA MOM	19	42	75
Mkt IA MOM	28	31	75
Mkt HML ROE MOM	17	34	74
Mkt SMB RMW CMA MOM	26	42	79
Mkt RMW CMA MOM	23	44	78
Mkt CMA MOM	29	39	75
Mkt SMB HML RMW MOM	21	39	81
Mkt HML RMW MOM	20	40	74
Mkt SMB RMW MOM	26	45	82
Mkt RMW MOM	31	38	79
Mkt HML MOM	35	42	73
Mkt MOM	29	38	71
Mkt SMB ROE IA	36	46	77
Mkt ROE IA	24	35	74
Mkt SMB RMW IA	33	43	78
Mkt RMW IA	25	38	75
Mkt IA	24	37	69
Mkt SMB CMA ROE	34	45	78
Mkt CMA ROE	27	35	71
Mkt SMB HML ROE	33	47	74
Mkt HML ROE	28	42	74
Mkt SMB ROE	32	41	84
Mkt ROE	27	40	73
Mkt SMB RMW CMA	28	47	76
Mkt RMW CMA	27	37	72
Mkt CMA	28	39	73
Mkt SMB HML RMW	24	40	78
Mkt HML RMW	27	40	76
Mkt SMB RMW	27	40	79
Mkt RMW	27	41	71
Mkt HML	22	46	74
Mkt	31	46	79
Mkt QMJ	27	38	77
Mkt LIQ	33	44	75
Mkt BAB	29	39	74
Mkt HMLD	25	33	72
Mkt MOM QMJ	30	41	75
Mkt IA QMJ	30	34	79
Mkt QMJ HMLD	30	37	73
Mkt MOM QMJ HMLD	30	43	75
Mkt MOM QMJ LIQ	26	51	85
Mkt CMA MOM LIQ	26	50	78
Mkt IA MOM QMJ	24	40	77
Mkt CMA LIQ BAB	19	47	82
Mkt SMB HML RMW QMJ	39	49	86
Mkt RMW CMA MOM BAB	23	45	81
Mkt CMA ROE BAB HMLD	19	45	79
Mkt RMW QMJ BAB HMLD	41	49	78
Mkt RMW CMA MOM LIQ BAB	25	43	82
Mkt RMW ROE QMJ BAB HMLD	46	55	87
Mkt HML RMW IA MOM BAB HMLD	42	57	82
Mkt HML RMW IA MOM BAB HMLD	43	50	86
Mkt HML RMW IA MOM BAB HMLD	39	62	87
Mkt HML ROE IA LIQ BAB HMLD	37	57	82

Table 1.4

better. This new method allows for reliable Bayesian work on which factors are risk factors in asset pricing models.

Chapter 2

Winners from Winners: A Tale of Risk Factors

Siddhartha Chib[1], Dashan Huang[2], Lingxiao Zhao[3], Guofu Zhou[4]

2.1 Introduction

The question of which risk factors best explain the cross-section of expected equity returns continues to draw attention on account of the large importance of this topic for theoretical and empirical finance (Cochrane, 2011). Along with the market factor, hundreds of additional risk factors have emerged (Harvey, Liu, and Zhu, 2016, Hou, Xue, and Zhang, 2017), and the set of possible such factors continues to grow. Rather than add to this list, we ask a straightforward question: could we start with the risk factor collections that have generated support in the recent literature, take

[1]Olin School of Business, Washington University in St. Louis, 1 Bookings Drive, St. Louis, MO 63130. e-mail: chib@wustl.edu.

[2]Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899. e-mail: dashanhuang@smu.edu.sg.

[3]Department of Economics, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130. e-mail: lingxiao@wustl.edu.

[4]Olin School of Business, Washington University in St. Louis, 1 Bookings Drive, St. Louis, MO 63130. e-mail: zhou@wustl.edu

the union of the factors in those collections as the pool of winners, and then find a new set of risk factors (winners from winners) that gather even more support from the data, on both statistical and financial grounds?

To answer this question, in what we call the tale of risk factors, we consider the four risk factor collections that we believe have spawned consensus support within the profession, namely those in the papers by Fama and French (1993, 2015, 2018), Hou, Xue, and Zhang (2015), Stambaugh and Yuan (2017), and Daniel, Hirshleifer, and Sun (2020). These collections, which cover the market, fundamental and behavioral factors, listed by author initials and risk factor abbreviations, are as follows[5]:

- FF6 collection: {Mkt, SMB, HML, RMW, CMA, MOM};
- HXZ collection: {Mkt, SMB, IA, ROE};
- SY collection: {Mkt, SMB, MGMT, PERF};
- DHS collection: {Mkt, PEAD, FIN}

In the first part of the analysis, the *winners model scan*, we focus on the set of these winners, and ask what collection of winners from winners emerge when each factor is allowed to play the role of a risk factor, or a non-risk factor, to produce different groupings of risk factors in a combinatorial fashion. Each grouping consists of a collection of factors that are risk factors in that grouping, and a complementary collection (the remaining factors) that are not risk factors in that grouping. We compare the resulting set of 4,095 asset pricing models, each of which is internally consistent with its assumptions of the risk factors, from a Bayesian model comparison perspective (Avramov and Chao, 2006, Barillas and Shanken, 2018, Chib and Zeng, 2019, Chib, Zeng, and Zhao, 2020).

Our first main result, from the winners model scan, is that a six-factor model, consisting of Mkt, SMB, MOM, ROE, MGMT, and PEAD from the twelve factors, gets the most support from

[5]There are slight differences in these collections in relation to the size factor, which we ignore for simplicity.

the data. This model is closely followed by a second model, a seven-factor model that has PERF as an additional risk factor, and a third model, a five-factor model, that does not have MOM as a risk factor. In terms of probabilities, these models have posterior model probabilities of around 0.13, 0.11, 0.10, respectively. Though one would have liked to witness even more decisive posterior support for the best model, this is unrealistic on a large model space composed of models constructed from factors that are a prior winners. Nevertheless, the data evidence is clear in isolating the winners from winners as the posterior probability distribution beyond the top three slumps sharply. For example, the posterior probabilities of the fourth and fifth-best models are around 0.03, and the sixth is about 0.025. The posterior probabilities of the remaining models in the model space barely register, being roughly of the same size as the prior probability of 1/4,095 = 0.00025, and even below.

Interestingly, models with the same risk factor set as FF6, HXZ, SY, and DHS do not appear in the top model set. As we demonstrate later, this relative under performance stems from a failure to clear an *internal consistency condition*. We say the risk factor set of a particular model satisfies an internal consistency condition if its risk factors can price the set of non risk factors in that model without incurring a penalty. In other words, a penalty is incurred, and the marginal likelihood suffers if the constraint that the intercepts in the conditional model of the non risk factors, which by the assumption must be all zero, is binding. If the constraints are not binding, then its marginal likelihood is significantly higher. Empirically, we find that the FF6, HXZ, SY, and DHS risk factor sets fail this internal consistency condition, while the risk factor sets of our three top factor models pass it.

We also document the performance of top models on other important statistical and economic dimensions. For one, we examine the performance of the top models in forecasting out-of-sample. Also, we compare the Sharpe ratios of the mean-variance portfolio constructed from the risk factors of the top models, and the Sharpe ratios of the portfolios from the risk factors in the FF6, HXZ, SY, and DHS collections. We find that the top models perform well in all comparisons.

In the second part of our analysis, in what we call the *winners plus genuine anomalies model scan*, we consider a more extensive model comparison of 8,388,607 asset pricing models, constructed from the twelve winners plus eleven principal components of anomalies unexplained by the winners, the *genuine anomalies*. The general question is to see if one can get an improved set of risk factors by considering models that involve the set of winners and additional factors based on the genuine anomalies, i.e., anomalies that cannot be explained by the winners. We show that from the set of 125 anomalies in Green, Hand, and Zhang (2017) and Hou, Xue, and Zhang (2017), only twenty anomalies cannot be explained by the winners. These constitute the set of genuine anomalies. For our winners plus anomalies model scan, we construct our different asset pricing models from the twelve winners and the first eleven principal components (PCs) of the genuine anomalies. In other words, each of the twelve winners, and each of the eleven PC's, is allowed to play the role of a risk-factor or a non-risk factor, leading to a model space of $2^{23}-1$ possible asset pricing models that we compare by using our Bayesian model comparison technique.

In this analysis, our tactic of reducing the 125 anomalies to the set of twenty genuine anomalies can be viewed as a *dimension-reduction strategy*. Furthermore, our idea of converting these anomalies to the space of principal components, is another element of the same strategy. Little is lost (and much is gained) by converting the genuine anomalies to PCs since, in either case, the genuine anomalies, or PCs, are portfolios of assets. What is important, however, is that we retain the identity of our winners, thus allowing us to understand whether the PC factors provide incremental information for pricing assets. And if so, whether the winners from the winners model scan continue to be risk factors in this broader space of models.

Our second main result, from the winners plus genuine anomalies scan, shows that there is much to gain by incorporating information in the genuine anomalies. For example, the Sharp ratios increase by more than 30%, which shows the benefit of incorporating information in genuine anomalies in explaining the cross-section of expected equity returns. Nonetheless, the risk factors from the winners set are the key risk factors, even though some prove to be redundant.

Our paper is part of a new wave of Bayesian approaches to risk factor selection. For instance, Kozak, Nagel, and Santosh (2020) focus on PC factors that are constructed from well-known factors and anomaly factors, and utilize interesting economic priors to isolate the relevant PCs in a classical penalized regression estimation. In contrast to their study, we retain the identity of the winners and construct PCs only from the genuine anomalies and approach the estimation from a fully Bayesian perspective. Bryzgalova, Huang, and Julliard (2019) is also part of this new wave of Bayesian work which delivers, for each factor, the marginal posterior probability that that factor is a risk factor, while our approach is concerned with the question of which collection of factors are jointly risk factors.

This paper adds broadly to the recent Bayesian literature in finance, for example, Pástor (2000), Pástor and Stambaugh (2000), Pástor and Stambaugh (2001), Avramov (2002), Ang and Timmermann (2012), and Goyal, He, and Huh (2018).

The rest of the paper is organized as follows. In Section 2, we briefly outline the methodology that we use to conduct our model comparisons. In Section 3, we consider the winners model scan, and in Section 4 the winners plus genuine anomalies model scan. Section 5 and Section 6 contain results and Section 7 concludes.

2.2 Methodology

Suppose that the potential risk factor set is denoted by $f_t : K \times 1$, where t denotes the t-th month. We now allow each factor to play the role of a risk factor (i.e., an element of the stochastic discount factor), or a non-risk factor, to produce different groupings of risk factors in a combinatorial fashion. Starting with K initial possible risk factors, there are, therefore, $J = 2^K - 1$ possible risk factor combinations (assuming that the risk-factor set cannot be empty). Each of these risk factor combinations defines a particular asset pricing model $\mathbb{M}_j$, $j = 1, ..., J$.

Consider now a specific model $\mathbb{M}_j$, $j = 1, ..., J$ consisting of the risk factors $x_{j,t} : k_{x,j} \times 1$, and the complementary set of factors (the non risk factors) $w_{j,t} : k_{x,j} \times 1$, where $K = k_{x,j} + k_{w,j}$. By definition, factors are risk factors if they are in the stochastic discount factor $M_{j,t}$. Following Hansen and Jagannathan (1991), we specify the SDF as

$$M_{j,t} = 1 - \lambda'_{x,j}\Omega_{x,j}^{-1}\left(x_{j,t} - \mathbb{E}[x_{j,t}]\right), \tag{2.2.1}$$

where $b_{x,j} \triangleq \Omega_{x,j}^{-1}\lambda_{x,j} : k_{x,j} \times 1$ are the unknown risk factor loadings, and $\Omega_{x,j} : k_{x,j} \times k_{x,j}$ is the covariance matrix of x_j. Enforcing the pricing restrictions implied by the no-arbitrage condition

$$\mathbb{E}[M_{j,t}x'_{j,t}] = 0 \quad \text{and} \quad \mathbb{E}[M_{j,t}w'_{j,t}] = 0$$

for all t, we get that $\mathbb{E}[x_{j,t}] = \lambda_{x,j}$, and $\mathbb{E}[w_{j,t}] = \Gamma_j\lambda_{x,j}$, for some matrix $\Gamma_j : k_{w,j} \times k_{x,j}$. If we assume that the joint distribution of (x_j, w_j) is Gaussian, then the latter pricing restrictions imply that under a marginal-conditional decomposition of the factors, $\mathbb{M}_j$ has the restricted reduced form given by

$$x_{j,t} = \lambda_{x,j} + \varepsilon_{x,j,t}, \tag{2.2.2}$$

$$w_{j,t} = \Gamma_j x_{j,t} + \varepsilon_{w\cdot x,j,t}, \tag{2.2.3}$$

where the errors are block independent Gaussian

$$\begin{pmatrix} \varepsilon_{x,j,t} \\ \varepsilon_{w\cdot x,j,t} \end{pmatrix} \sim \mathcal{N}_K\left(0, \begin{pmatrix} \Omega_{x,j} & 0 \\ 0 & \Omega_{w\cdot x,j} \end{pmatrix}\right), \tag{2.2.4}$$

and $\Omega_{w\cdot x,j} : k_{w,j} \times k_{w,j}$ is the covariance matrix of the conditional residuals $\varepsilon_{w\cdot x,j,t}$.

The goal of the analysis is to calculate the support for these models, $\mathbb{M}_j$, $j = 1, \ldots, J$, given the sample data on the factors. To explain how this is done, we start with the prior distributions of the parameters across models.

The parameters of model $\mathbb{M}_j$ are given by

$$\theta_j \triangleq (\lambda_{x,j}, \eta_j),$$

where $\lambda_{x,j}$ are the risk premia parameters, and $\eta_j = (\Gamma_j, \Omega_{x,j}, \Omega_{w\cdot x,j})$ are nuisance parameters. Note the key point that the dimension of these nuisance parameters equals

$$\begin{aligned} & \{k_{w,j}k_{x,j} + k_{x,j}(k_{x,j}+1)/2 + k_{w,j}(k_{w,j}+1)/2\} \\ = & \{k_{x,j}^2 + k_{w,j}^2 + 2k_{x,j}k_{w,j} + (k_{x,j}+k_{w,j})\}/2 \\ = & (K^2+K)/2, \end{aligned}$$

which is the same across models. Chib and Zeng (2019) exploit this fact to develop proper priors, and Chib, Zeng, and Zhao (2020) to develop improper priors, of $\eta_j,\ j = 1, 2, ..., J$, from a *single prior distribution*.

Let $\mathbb{M}_1$ denote the model in which all K potential risk factors are risk factors. Then, η_1 just consists of $\Omega_{x,1}$. Now let this parameter have the Jeffreys' improper prior

$$\pi(\Omega_{x,1}|\mathbb{M}_1) = c|\Omega_{x,1}|^{-\frac{K+1}{2}}. \tag{2.2.5}$$

By derivations given in Chib, Zeng, and Zhao (2020), we get that the priors of η_j, $j > 1$ are

$$\psi(\eta_j|\mathbb{M}_j) = c|\Omega_{x,j}|^{-\frac{2k_{x,j}-K+1}{2}}|\Omega_{w\cdot x,j}|^{-\frac{K+1}{2}},\ j > 1, \tag{2.2.6}$$

where c is an arbitrary constant that by construction is the same across these priors, and, hence, irrelevant in the comparison of models.

Finally, complete the prior distributions by supposing that, conditional on η_j, the priors of $\lambda_{x,j}$

are the multivariate normal distributions

$$\lambda_{x,j}|\mathbb{M}_j,\eta_j \sim \mathcal{N}_{k_{x,j}}(\lambda_{x,j,0},\kappa_j\Omega_{x,j}),$$

where $\lambda_{x,j,0}$ and κ_j are model-specific hyperparameters that are determined from a training sample.

2.2.1 Marginal Likelihoods

Assume that each model $\mathbb{M}_j \in \mathscr{M}$ has a prior model probability of $\Pr(\mathbb{M}_j)$ of being the correct model. The objective is to calculate the posterior model probability $\Pr(\mathbb{M}_j|y_{1:T})$, where $y_{1:T} = (f_1,...,f_T)$ is the estimation sample of the potential risk factors.

The key quantities for performing this prior-posterior update for the models in $\mathscr{M}$ are the marginal likelihoods, defined as

$$m_j(y_{1:T}|\mathbb{M}_j) = \int_{\Theta_j} p(y_{1:T}|\mathbb{M}_j,\theta_j)\pi(\lambda_{x,j}|\mathbb{M}_j,\eta_j)\psi(\eta_j|\mathbb{M}_j)\mathrm{d}\theta_j,$$

where Θ_j is the domain of θ_j,

$$p(y_{1:T}|\mathbb{M}_j,\theta_j) = \prod_{t=1}^{T}\mathcal{N}_{k_{x,j}}(x_{j,t}|\lambda_{x,j},\Omega_{x,j})\mathcal{N}_{k_{w,j}}(w_{j,t}|\Gamma_j x_{j,t},\Omega_{w\cdot x,j})$$

is the density of the data and $\mathcal{N}_d(\cdot|\mu,\Omega)$ is the d-dimensional multivariate normal density function with mean μ and covariance matrix Ω.

Notice that the phrase marginal likelihood encapsulates two concepts: one that it is a function that is marginalized over the parameters of model j, hence the word marginal; and second that it is the likelihood of the model parameter $\mathbb{M}_j$, hence the word likelihood.

Under our assumptions, the *log* marginal likelihoods are in closed form. Specifically,

$$\log m_1(y_{1:T}|\mathbb{M}_1) = -\frac{TK}{2}\log\pi - \frac{K}{2}\log(T\kappa_1+1) - \frac{T}{2}\log|\Psi_1| + \log\Gamma_K\left(\frac{T}{2}\right),$$

and

$$\log m_j(y_{1:T}|\mathbb{M}_j) = -\frac{Tk_{x,j}}{2}\log\pi - \frac{k_{x,j}}{2}\log(T\kappa_j+1) - \frac{\left(T+k_{x,j}-K\right)}{2}\log|\Psi_j| + \log\Gamma_{k_{x,j}}\left(\frac{T+k_{x,j}-K}{2}\right)$$
$$-\frac{(K-k_{x,j})(T-k_{x,j})}{2}\log\pi - \frac{\left(K-k_{x,j}\right)}{2}\log|W_j^*| - \frac{T}{2}\log|\Psi_j^*| + \log\Gamma_{K-L_j}\left(\frac{T}{2}\right),\ j>1\ ,$$

where we have set $c=1$ (as this choice is irrelevant, by construction), and

$$\Psi_j = \sum_{t=1}^{T}(x_{j,t}-\hat{\lambda}_{x,j})(x_{j,t}-\hat{\lambda}_{x,j})' + \frac{T}{T\kappa_j+1}\left(\hat{\lambda}_{x,j}-\lambda_{xj0}\right)\left(\hat{\lambda}_{x,j}-\lambda_{xj0}\right)'$$
$$W_j^* = \sum_{t=1}^{T}x_{j,t}x_{j,t}'\ ,\ \ \Psi_j^* = \sum_{t=1}^{T}(w_{j,t}-\hat{\Gamma}_j x_{j,t})(w_{j,t}-\hat{\Gamma}_j x_{j,t})'.$$

Note that the variables in hat in the above expressions are the least squares estimates calculated using the estimation sample, and $\Gamma_d(\cdot)$ denotes the d dimensional multivariate gamma function.

2.2.2 Model Scan Approach

We conduct a prior-posterior analysis on the model space denoted by $\mathscr{M} = \{\mathbb{M}_1, \mathbb{M}_2, \ldots, \mathbb{M}_J\}$. Assume that each model in the model space is given an uninformative and equalized prior model probability, that is, for any j

$$\Pr(\mathbb{M}_j) = 1/J. \tag{2.2.7}$$

Applying Bayes theorem to the unknown model parameter $\mathbb{M}_j$, the posterior model probability is given by

$$\Pr(\mathbb{M}_j|y_{1:T}) = \frac{m_j(y_{1:T}|\mathbb{M}_j)}{\sum_{l=1}^{J} m_l(y_{1:T}|\mathbb{M}_l)}\ , \tag{2.2.8}$$

as the model prior probabilities in the numerator and the denominator cancel out.

Both the prior and posterior probability distributions on model space acknowledge the notion of *model uncertainty*. The prior distribution on model space represents our beliefs about the models before we see the data. A discrete uniform prior is our default, but, of course, it is possible to consider other prior distributions. The posterior distribution retains model uncertainty unless the sample size is large in relation to the dimension of the model space. By this we mean that the posterior distribution on model space will not concentrate on a single model. As T becomes large, however, the asymptotic theory of the marginal likelihood (see, e.g., Chib, Shin, and Simoni (2018)), implies that the posterior model probabilities will concentrate on the true model (if it is in the set of models), or on the model that is closest to the true model in the Kullback-Leibler distance.

Regardless of the sample size, however, the end-product of our analysis is a ranking of models by posterior model probabilities (equivalently, by marginal likelihoods given that the denominator in the posterior probability calculation is just a normalization constant). We indicate these ranked models by

$$\{\mathbb{M}_{1*}, \mathbb{M}_{2*}, \ldots, \mathbb{M}_{J*}\}$$

such that

$$m_{1*}(y_{1:T}|\mathbb{M}_{1*}) > m_{2*}(y_{1:T}|\mathbb{M}_{2*}) > \cdots > m_{J*}(y_{1:T}|\mathbb{M}_{J*}).$$

This ranking provides the basis for our empirical Bayesian model comparison.

2.3 Winners Model Scan

As mentioned in the introduction, our first analysis is based on twelve factors from the studies of Fama and French (1993, 2015, 2018), Stambaugh and Yuan (2017), and Daniel, Hirshleifer, and Sun (2020). Details of these factors, {Mkt, SMB, HML, RMW, CMA, MOM, IA, ROE, MGMT, PERF, PEAD, FIN}, are given in Table 2.1. While Mkt captures the overall market risk, SMB, HML, RMW, CMA, MOM, IA, and ROE are well-known characteristic-based factors and are constructed by sorting stocks in a relatively simple way. For those remaining novel four mispricing factors, MGMT and PERF are constructed based on average rankings of eleven anomalies of stocks, and PEAD and FIN are related to investors' psychology. Although the construction and motivation of those factors are different, those twelve factors are named "winners" as they are believed and proved to have strong power in explaining the cross-section of expected equity returns.

The data on these winners are monthly, spanning the period from January 1974 to December 2018, for a total of 540 starting observations. Of these the last 12 months of data are held-out for out-of-sample prediction validation purpose. For the other 528 in-sample monthly observations, the first 10 percent is used as a training sample to construct the prior distributions of the risk premia parameters, leaving a sample size of $T = 475$ as estimation sample.

2.3.1 Two special cases

To understand how the framework is applied, consider first the model in which all twelve winners are risk factors. In this case, model $\mathbb{M}_1$ (say), the general model reduces to

$$x_{1,t} = \underbrace{\lambda_{x,1}}_{12\times 1} + \varepsilon_{x,1,t}, \;\; \varepsilon_{x,1,t} \sim \mathcal{N}_{12}(0, \Omega_{x,1}), \tag{2.3.1}$$

where $x_{1,t} = (\text{Mkt}, \text{SMB}, \text{SML}, \text{RMW}, \text{CMA}, \text{MOM}, \text{IA}, \text{ROE}, \text{MGMT}, \text{PERF}, \text{PEAD}, \text{FIN})'_t$ and, since the non-risk factor collection $w_{1,t}$ is empty, $k_{x,1} = 12$ and $\Omega_{x,1} : 12 \times 12$.

Now consider $\mathbb{M}_2$ (say) with the FF6 risk factors $x_{2,t} = (\text{Mkt}, \text{SMB}, \text{SML}, \text{RMW}, \text{CMA}, \text{MOM})'_t$. Then, we have

$$x_{2,t} = \underbrace{\lambda_{x,2}}_{6\times 1} + \varepsilon_{x,2,t}, \;\; \varepsilon_{x,2,t} \sim \mathcal{N}_{6}(0, \Omega_{x,2}),$$

while for $w_{2,t} = (\text{IA}, \text{ROE}, \text{MGMT}, \text{PERF}, \text{PEAD}, \text{FIN})_t'$ we have

$$w_{2,t} = \underbrace{\Gamma_2}_{6\times 6} x_{2,t} + \varepsilon_{w\cdot x,2,t}, \;\; \varepsilon_{w\cdot x,2,t} \sim \mathcal{N}_6(0, \Omega_{w\cdot x,2}),$$

where $k_{x,2} = 6$, $k_{w,2} = 6$, $\Omega_{x,2} : 6 \times 6$, and $\Omega_{w\cdot x,2} : 6 \times 6$. The latter model embodies the pricing restrictions that the assumed risk factors of this model price the non-risk factors $w_{2,t}$.

There are $J = 4,095$ such models in the model space $\mathscr{M}$. Our goal is to compare these J models using the model scan approach described in Section 2.

2.3.2 Winners Model Scan Results

Top Model Set

To get a clear picture of the prior-posterior update on the model space $\mathscr{M}$, we view each model as a point in that space. The prior distribution of models on that space is uniform. The posterior probabilities of the models are proportional to the product of the uniform prior and the marginal likelihoods. We can use these posterior probabilities to plot these points (or models) in that space in decreasing order. From Figure 2.1, which plots the posterior model probability of the top 220 models. We can see from the figure that the posterior model probabilities drop sharply beyond the top three models. For example, the posterior probabilities of the fourth and fifth-best models are around 0.03, and the sixth is about 0.025. The posterior probabilities of the remaining models in the model space barely register, being roughly of the same size as the prior probability of 1/4,095 = 0.00025, and even below.

In Figure 2.2 we plot these posterior model probabilities but, this time, only for the top 5 models. We see that the top three models have a joint posterior probability of 0.3407. In a sense, we can think of these models as being indistinguishable, or equivalent. To make this more precise, in the notation introduction above, let $\mathbb{M}_{1*}$ denote the highest posterior probability model. If we now let the Bayes factor of the best model against any other model be denoted by

$$B_{1j} = \frac{m_{1*}(y_{1:T}|\mathbb{M}_{1*})}{m_{j*}(y_{1:T}|\mathbb{M}_{j*})},$$

then, according to Jeffreys' scale, if $B_{1j} \leq 10^{\frac{1}{2}}$, the evidence supporting $\mathbb{M}_{1*}$ over $\mathbb{M}_{j*}$ is barely worth mentioning. Equivalently, in terms of the log Bayes factor, the indistinguishably condition above can be expressed as

$$\log B_{1j} = \log m_{1*}(y_{1:T}|\mathbb{M}_{1*}) - \log m_{j*}(y_{1:T}|\mathbb{M}_{j*}) \leq 1.15.$$

We can, therefore, refer to $\mathbb{M}_{j*}$ that is in the radius of the best model in this way as being indistinguishable from the best model.

Applying this criterion, we conclude that $\mathbb{M}_{1*}$, $\mathbb{M}_{2*}$, and $\mathbb{M}_{3*}$ constitute the top model set $\mathscr{M}_*$ in the winners scan, while $\mathbb{M}_{4*}$ and $\mathbb{M}_{5*}$ also given in Figure 2.2 are not in the top model set.

Table 2.2 shows that the six-factor model $\mathbb{M}_{1*}$ consisting of Mkt, SMB, MOM, ROE, MGMT, and PEAD as risk factors gets the most support from the data. This model is closely followed by the seven-factor model $\mathbb{M}_{2*}$ that has PERF as an additional risk factor, and the five-factor model $\mathbb{M}_{3*}$ that does not have MOM as a risk factor.

Interestingly, Mkt, SMB, ROE, MGMT, and PEAD, are present in each of the three top groupings. It appears that both fundamental and behavioral factors play an important role in pricing the cross-section of expected equity returns. It should also be noted that the top groupings feature between five and seven-factors, similar to the number of factors in most of the literature.

Besides, the ratio of the posterior model probability and the prior model probability of any given model $\mathbb{M}_j$, denoted by $\frac{\Pr(y_{1:T}|\mathbb{M}_{j*})}{\Pr(\mathbb{M}_{j*})}$, is provided in Table 2.2. That ratio reflects the information improvement of posterior over the same prior for $\mathbb{M}_j$ when data are observed. Therefore it is a good measure for evaluating the joint superiority of candidate models. For comparison, Panel

B of Table 2.2 reports the marginal likelihoods and that ratios for $\mathbb{M}_1$ and models with the same risk factor sets as CAPM, Fama and French family, HXZ, SY, and DHS. It can seen that the information improvement of each of those models is substantially smaller than that of top models. Because $\mathbb{M}_1$ is also not supported by the data, we can conclude that the twelve factors together contain information redundancies.

Parameter Updating for the Best Model of the Winners Model Scan

We now provide more details about the best model in the winners model scan $\mathbb{M}_{1*}$, which takes the form

$$\mathbb{M}_{1*}: \begin{pmatrix} \mathrm{Mkt}_t \\ \mathrm{SMB}_t \\ \mathrm{MOM}_t \\ \mathrm{ROE}_t \\ \mathrm{MGMT}_t \\ \mathrm{PEAD}_t \end{pmatrix} = \underbrace{\lambda_{x,1*}}_{6\times 1} + \varepsilon_{x,1*,t}, \;\; \varepsilon_{x,1*,t} \sim \mathcal{N}_6(0, \Omega_{x,1*}),$$

$$\begin{pmatrix} \mathrm{HML}_t \\ \mathrm{RMW}_t \\ \mathrm{CMA}_t \\ \mathrm{IA}_t \\ \mathrm{PERF}_t \\ \mathrm{FIN}_t \end{pmatrix} = \underbrace{\Gamma_{1*}}_{6\times 6} \begin{pmatrix} \mathrm{Mkt}_t \\ \mathrm{SMB}_t \\ \mathrm{MOM}_t \\ \mathrm{ROE}_t \\ \mathrm{MGMT}_t \\ \mathrm{PEAD}_t \end{pmatrix} + \varepsilon_{w\cdot x,1*,t}, \;\; \varepsilon_{w\cdot x,1*,t} \sim \mathcal{N}_6(0, \Omega_{w\cdot x,1*}).$$

In calculating the marginal likelihood of this model, which equals 14186.43, as reported earlier in Table 2.2, we used the prior on $\eta_{1*} = (\Gamma_{1*}, \Omega_{x,1*}, \Omega_{w\cdot x,1*})$ from (2.2.6) equal to $|\Omega_{x,1*}|^{-\frac{1}{2}} |\Omega_{w\cdot x,1*}|^{-\frac{13}{2}}$, and the proper prior of the risk premia parameters $\lambda_{x,1*}$ from the training sample equal to

$$\pi(\lambda_{x,1*}|\mathbb{M}_{1*},\eta_{1*}) = \mathcal{N}_6(\lambda_{x,1*}|\lambda_{x,1*,0}, 0.1915\times\Omega_{x,1*}),$$

where $\lambda_{x,1*,0} = (0.0017, 0.0130, 0.0044, 0.0041, 0.0084, 0.0085)'$.

Under our prior distributions, it is easy to confirm that the posterior distributions $\pi(\theta_j|\mathbb{M}_j, y_{1:T})$ of parameters θ_j of any given model $\mathbb{M}_j$ have the marginal-conditional forms given by

$$\pi(\Omega_{x,j}|\mathbb{M}_j, y_{1:T}) = \mathcal{W}_{k_{x,j}}^{-1}(\Omega_{x,j}|\Psi_j, T + k_{x,j} - K), \tag{2.3.2}$$

$$\pi(\lambda_{x,j}|\mathbb{M}_j, y_{1:T}, \Omega_{x,j}) = \mathcal{N}_{k_{x,j}}(\lambda_{x,j}|\lambda_{xj1}, \left(1/\kappa_j + T\right)^{-1}\Omega_{x,j}), \tag{2.3.3}$$

and

$$\pi(\Omega_{w\cdot x,j}|\mathbb{M}_j, y_{1:T}) = \mathcal{W}_{k_{w,j}}^{-1}(\Omega_{w\cdot x,j}|\Psi_j^*, T), \tag{2.3.4}$$

$$\pi(\mathrm{vec}(\Gamma_j)|\mathbb{M}_j, y_{1:T}, \Omega_{w\cdot x,j}) = \mathcal{N}_{k_{w,j}\times k_{x,j}}(\mathrm{vec}(\Gamma_j)|\mathrm{vec}(\hat{\Gamma}_j), W_j^{*-1}\otimes\Omega_{w\cdot x,j}), \tag{2.3.5}$$

where $\lambda_{x,j,1} = \frac{1}{T\kappa_j+1}\lambda_{x,j,0} + \frac{T\kappa_j}{T\kappa_j+1}\hat{\lambda}_{x,j}$ and $\mathcal{W}^{-1}(\Psi,\nu)$ denotes the inverse Wishart distribution with ν degrees of freedom and scale matrix Ψ. Thus, the posterior distribution $\pi(\theta_j|\mathbb{M}_j, y_{1:T})$ is given by the product of equations (2.3.2), (2.3.3), (2.3.4), and (2.3.5). We can apply this result to generate a large number of simulated draws, first by sampling the marginal distribution, and then by the conditional distribution given the draws from the marginal distributions. These sampled draws can be used to make marginal posterior distributions of relevant parameters, and other summaries.

Applying the above sampling procedure to $\mathbb{M}_{1*}$, we obtain the marginal posterior distributions of the risk premia parameters $\lambda_{x,1^*}$, given in Figure 2.3, and the posterior means, standard deviations and quantiles, given in Table 2.3. It is interesting that the posterior means of the risk premia parameters are similar, except for that of SMB, while the posterior standard deviations of the risk premia of Mkt and MOM are almost twice as large as the rest.

2.3.3 Why do FF6, HXZ, SY, and DHS not win?

It is crucial and meaningful to understand why models with the same risk factor set as FF6, HXZ, SY, and DHS do not appear in the top model set. The reason for this is interesting. Essentially, those models do not satisfy an internal consistency condition. We say that a particular model satisfies the internal consistency condition if its risk factors can price the set of non risk factors in that model without incurring a penalty. In other words, a penalty is incurred, and the marginal likelihood suffers if the constraint that the intercepts in the conditional model of the non risk factors, which by assumption must be all zero, is binding. If the constraints are not binding, then its marginal likelihood is significantly higher.

Consider model $\mathbb{M}_j$

$$
\begin{aligned}
x_{j,t} &= \lambda_{x,j} + \varepsilon_{x,j,t}, \\
w_{j,t} &= \Gamma_j x_{j,t} + \varepsilon_{w\cdot x,j,t},
\end{aligned}
$$

with risk factors $x_{j,t}$ and non risk factors $w_{j,t} = (w_{j,1,t}, \ldots, w_{j,k_{w,j},t})'$ with dimension $k_{w,j} \times 1$. Now for each non risk factor $w_{j,i,t}$, $i = 1, 2, \ldots, k_{w,j}$, we compare the two models,

$$
\mathbb{M}^i_{j,0} : \; w_{j,i,t} = \gamma'_{j,i} x_{j,t} + \varepsilon_{j,i,t} \tag{2.3.6}
$$

and

$$
\mathbb{M}^i_{j,1} : \; w_{j,i,t} = \alpha_{j,i} + \gamma'_{j,i} x_{j,t} + \varepsilon_{j,i,t} \tag{2.3.7}
$$

using marginal likelihoods. If the log marginal likelihood of the second model does not exceed that of the first model by more than 1.15, then, by an application of Jeffreys' rule, we can conclude that imposing the zero $\alpha_{j,i}$ pricing restriction does not result in a marginal likelihood penalty. Stated yet another way, this means that the non risk factor $w_{j,i}$ can be priced by the risk factor set x_j of that model $\mathbb{M}_j$. If this condition holds for each of the factors in w_j, we conclude that the risk factor set of that model satisfies the ICC condition.

Our analysis shows that models with the same risk factor sets as CAPM, Fama and French family, HXZ, SY, and DHS do not satisfy ICC. Specifically, the single Mkt factor cannot explain the remaining 11 non risk factors. The risk factor sets of the Fama and French family models can explain at most one non risk factor (IA). The risk factor set of HXZ can explain all of the Fama and French factors, but cannot explain MGMT and PEAD. The risk factor sets of SY and DHS models cannot explain one non risk factor, PEAD and MGMT, respectively. In contrast, the top models in $\mathscr{M}_*$ satisfy the ICC condition fully, which helps to explain why those models rank high in the winners model scan.

2.3.4 Prediction

It is worthwhile to consider how well the top models perform out-of-sample. From the Bayesian perspective, an elegant way to examine this question is by calculating the *predictive likelihood* for a set of future observations. This predictive likelihood, which like the marginal likelihood, is a number when evaluated at a particular sample of future observations, can be used to rank the predictive performance of each model in the model space.

To define the predictive likelihood, let $\pi(\theta_j|\mathbb{M}_j, y_{1:T})$ denote the posterior distributions of the parameters θ_j of $\mathbb{M}_j$, and let $y_{(T+1):(T+12)} = (f_{T+1}, ..., f_{T+12})$ denote 12 months of held-out out-of-sample data on those winners. Then, for any given model $\mathbb{M}_j$, the predictive likelihood is defined as

$$m_j\left(y_{(T+1):(T+12)}|\mathbb{M}_j, y_{1:T}\right) = \int_{\Theta_j} p(y_{(T+1):(T+12)}|\mathbb{M}_j, \theta_j)\pi(\theta_j|\mathbb{M}_j, y_{1:T})\mathrm{d}\theta_j,$$

where

$$p(y_{(T+1):(T+12)}|\mathbb{M}_j, \theta_j) = \prod_{s=1}^{12} \mathscr{N}_{k_{x,j}}(x_{j,T+s}|\lambda_{x,j}, \Omega_{x,j})\mathscr{N}_{k_{w,j}}(w_{j,T+s}|\Gamma_j x_{j,T+s}, \Omega_{w\cdot x,j})$$

is the out-of-sample density of the factors given the parameters. We can compute this integral by Monte Carlo. Taking draws $\{\theta_j^{(1)},...,\theta_j^{(G)}\}$ from $\pi(\theta_j|\mathbb{M}_j, y_{1:T})$, with G being a large integer, we calculate the predictive likelihood as the Monte Carlo average

$$m_j\left(y_{(T+1):(T+12)}|\mathbb{M}_j, y_{1:T}\right) = \frac{1}{G}\sum_{g=1}^{G} p(y_{(T+1):(T+12)}|\mathbb{M}_j, \theta_j^{(g)}).$$

Table 2.4 reports the log predictive likelihoods of the top three models as well as those competing models. We can see that the top three models also have larger predictive likelihoods, which means that they outperform the competing models on the predictive dimension.

2.4 Winners Plus Genuine Anomalies Model Scan

We now show that there are some benefits to including additional potential risk factors along with the winners. There are many additional risk factors to draw upon and the approach we describe can be used with any set of additional potential risk factors. For our analysis here we focus on the 125 anomalies in Green et al. (2017) and Hou, Xue, and Zhang (2017). These anomalies, which are re-balanced at an annual or quarterly frequency, exclude anomalies that are re-balanced at a monthly frequency, because the latter cease to be anomalies once transaction costs are taken into account (Novy-Marx and Velikov, 2016, Patton and Weller, 2020). All these portfolios are value-weighted and held for one month. Just as in the case of the winners, we have monthly observations on these anomalies spanning the period from January 1974 to December 2018, for a total of 540 observations. We partition these observations into out-of-sample and in-sample, which consists of the training sample and the estimation sample, just as in Section 3.

What we aim to show is that whether there is information in these anomalies that can be captured to produce better statistical performance (in terms of marginal likelihoods) and higher Sharpe-ratios of portfolios built from the best fitting risk-factors. In order to show this, we recognize that the winners are already carefully vetted factors and, therefore, the anomalies that are allowed to enter the pool of augmented potential risk factors must be those that cannot be priced by these winners. This point helps to limit the dimension of the model space and allows us to design a full model scan approach, as we now detail.

2.4.1 Genuine Anomalies

The model space with all 125 anomalies is 2^{137}, which is astronomically large. However, it is unnecessary to consider such a large model space because many of the anomalies can actually be priced by the winners. In fact, Harvey, Liu, and Zhu (2016) and Hou, Xue, and Zhang (2017) have cast doubt on the credibility of these anomalies and Cochrane (2011) has raised similar concerns.

The first step, therefore, is to eliminate anomalies that are not *genuine anomalies*. An anomaly is a genuine anomaly if it cannot be priced by the winners. Here is how we sort this issue out. Let z_i, $i = 1, 2, \ldots, 125$, denote the anomalies. Let $x =$ (Mkt, SMB, HML, RMW, CMA, MOM, IA, ROE, MGMT, PERF, PEAD, FIN) denote the twelve winners. Now for each anomaly z_i as the response, and x as the covariates, we estimate two models, one without an intercept and one with:

$$\mathbb{M}_0^i : \; z_{i,t} = \gamma_i' x_t + \varepsilon_{i,t}, \; \varepsilon_{i,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_i^2) \tag{2.4.1}$$

and

$$\mathbb{M}_1^i : \; z_{i,t} = \alpha_i + \gamma_i' x_t + \varepsilon_{i,t} \; \varepsilon_{i,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_i^2). \tag{2.4.2}$$

In estimating these models, the first 10 percent of the data are used as a training sample to pin down the hyperparameters of the proper prior distributions. We then use standard Bayesian Markov chain Monte Carlo methods to estimate model on the remaining 90 percent of the data, $y_{1:T}$. The log marginal likelihood of each model is computed by the Chib (1995) method. Denote these marginal likelihoods by $\log m_0^i(y_{1:T}|\mathbb{M}_0^i)$ and $\log m_1^i(y_{1:T}|\mathbb{M}_0^i)$. Then, based on the Jeffreys

(1961) scale, if the following condition holds

$$\log m_1^i(y_{1:T}|\mathbb{M}_1^i) - \log m_0^i(y_{1:T}|\mathbb{M}_0^i) > 1.15, \tag{2.4.3}$$

then x is not able to price z_i. In this case, we assert that z_i is a genuine anomaly; otherwise z_i is not a genuine anomaly.

Applying the procedure described above, twenty genuine anomalies emerge, namely, acc, age, currat, hire, lev, quick, salecash, sgr, Em, Lbp, dFin, Cop, Lfe, SA, sue, cash, OLAQ, CLAQ, TBIQ, and BLQ. Details about these anomalies are given in Table 2.5.

2.4.2 The Potential Risk Factor Set

Now instead of conducting our model scan on the original twelve winners and these twenty genuine anomalies, which leads to a model space of around four billion models ($2^{32} - 1 = 4,294,967,295$), we apply a second dimension reduction step by converting the genuine anomalies to principal components (with the rotated mean added back in), of which we then consider the first eleven that explain in total approximating 91% of the variation in the genuine anomalies, as can be seen from Table 2.6. This set, of the twelve winners and the first eleven PCs of the genuine anomalies, constitutes our potential risk factor set which we use to launch our extended risk factor analysis.

We note that this strategy of blending of winners and the PCs in this way appears to be new to the literature. By this strategy, we are able to limit the model space to a reasonable dimension (of around eight million models), while simultaneously avoiding the problem of working with twenty correlated PCs that are also quite correlated with the winners. For instance, some anomalies are related to leverage (currat, lev, quick, Lbp, and BLQ) and some are linked to sales status (salecash and sgr). Considering two groups of risk factors in this way, where some are in their original form and some are PCs, appears to be novel. It allows us to show the value of including anomalies as potential risk factors.

We also note that the idea of transforming our genuine anomalies into their corresponding principal components is similar to Kozak, Nagel, and Santosh (2020) who argued that “a relatively small set of principal component from the universe of potential characteristics-based factors can approximate the SDF quite well.” Our idea is related, but distinct, as we keep the key factors (the winners) as is, but only covert the less important (the genuine anomalies) into principal components.

We emphasize again that our approach of reducing the 125 anomalies to the set of twenty genuine anomalies is a *dimension-reduction strategy*. Furthermore, our idea of converting these anomalies to the space of principal components, is another element of that same strategy. Of course, whether as anomalies, or as PC’s, these factors are portfolios of assets. We believe that it is meaningful and useful to retain the identity of the winners, thus allowing us to understand whether the PC factors provide incremental information for pricing assets.

2.5 Winners Plus Genuine Anomalies Model Scan Results

2.5.1 Top Model Set

Starting with the potential risk factor set of dimension $\tilde{K} = 23$, twelve winners plus eleven PCs, and applying the methodology given in Section 2, we calculate the marginal likelihood of each of the $\tilde{J} = 8,388,607$ models in $\tilde{\mathscr{M}}$. Converting these marginal likelihoods into posterior model probabilities, the ratios of these posterior model probabilities and the prior model probability (assumed equal to $1/\tilde{J}$) can be calculated. The ratio, $\frac{\mathsf{Pr}(\tilde{y}_{1:T}|\tilde{\mathbb{M}}_{j*})}{\mathsf{Pr}(\tilde{\mathbb{M}}_{j*})}$, defined earlier in Section 3.2.1, makes it easier to see which models receive more support from the data. We report these ratios for the top 220 models in Figure 2.4, in which the dashed blue vertical line represents the cutoff of the top model set.

The top model set, denoted by $\tilde{\mathscr{M}}_*$ as in Section 3.2.1, can be defined in relation to the best model $\tilde{\mathbb{M}}_{1*}$. A model is in the top model set if its distance to the best model on the log marginal likelihood scale is less than 1.15. These 29 models, along with the including associated risk factor sets, log marginal likelihoods, and the ratios of posterior model probability and prior model probability are provided in Panel A of Table 2.7. The risk factors common to all these top models are Mkt, MOM, ROE, PEAD, MGMT, PC1, PC4, and PC5. We also note that the risk factors that are common to the top 3 models in the winners model scan, {Mkt, SMB, ROE, PEAD, MGMT}, are also the risk factors that are common to the top 29 models in the extended model scan except that SMB is replaced by MOM, which is also risk factors of the top 2 models of the winners scan.

Parameter Updating for the Best Model of the Winners Plus Genuine Model Scan

We now provide more details about the best model in the winners plus genuine model scan $\tilde{\mathbb{M}}_{1*}$, in which the risk factor set is given by $\tilde{x}_{1*,t} = (\text{Mkt}, \text{RMW}, \text{MOM}, \text{IA}, \text{ROE}, \text{MGMT}, \text{PEAD}, \text{FIN}, \text{PC1}, \text{PC3}, \text{PC4}, \text{PC5}, \text{PC7})_t'$ and the non risk factor set is given by $\tilde{w}_{1*,t} = (\text{SMB}, \text{HML}, \text{CMA}, \text{PERF}, \text{PC2}, \text{PC6}, \text{PC8}, \text{PC9}, \text{PC10}, \text{PC11})_t'$:

$$\begin{aligned}\tilde{\mathbb{M}}_{1*}: \ \tilde{x}_{1*,t} &= \underbrace{\tilde{\lambda}_{x,1*}}_{13\times 1} + \tilde{\varepsilon}_{x,1*,t}, \ \ \tilde{\varepsilon}_{x,1*,t} \sim \mathcal{N}_{13}(0, \tilde{\Omega}_{x,1*}), \\ \tilde{w}_{1*,t} &= \underbrace{\tilde{\Gamma}_{1*}}_{10\times 10} \tilde{x}_{1*,t} + \tilde{\varepsilon}_{w\cdot x,1*,t}, \ \ \tilde{\varepsilon}_{w\cdot x,1*,t} \sim \mathcal{N}_{10}(0, \tilde{\Omega}_{w\cdot x,1*}).\end{aligned}$$

Similar to Section 3.2, the prior and posterior statistics, i.e. prior mean, posterior mean, posterior standard deviation, posterior median, 2.5% quantile, and 97.5% quantile for the risk premia $\tilde{\lambda}_{1*}$ are provided in Table 2.8.

2.5.2 Internal Consistency Condition

Just as in Section 3.3, we can see that 27 out of 29 models in the top model set $\tilde{\mathscr{M}}_*$ satisfy the ICC condition completely. The two exceptions occur for $\mathbb{M}_{16*}$ which is unable to explain IA and $\mathbb{M}_{16*}$

which is unable to explain RMW. In constrast, models with the same risk factor sets as CAPM, Fama and French family, HXZ, SY, and DHS deviate from ICC further, leaving out 13, 12, 10, 9, 5, 5, and 4 non-risk factors unexplained. Moreover, none of models with the same risk factor sets as the top three models in the winners scan can explain PC1, PC2, and PC3.

2.5.3 Prediction

Similar to Section 3.4, it is important to consider how well the winning model performs out-of-sample and we compute the *predictive likelihood* for a set of future observations $\tilde{y}_{(T+1):(T+12)} = (\tilde{f}_{T+1},...,\tilde{f}_{T+12})$ denote 12 months of out-of-sample data on the winners and principal components. Table 2.9 reports the log predictive likelihoods for top models in $\tilde{\mathscr{M}}_*$ as well as models with the same risk factor sets as CAPM, Fama and French family models, SY and DHS. We can tell that those top models do not fail out of sample.

2.6 Economic Performance: Sharpe Ratios

Now suppose that based on the identity of the risk factors in (say) the best model $\tilde{\mathbb{M}}_{1*}$ of the winners plus genuine anomalies model scan, namely, Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, and PC7, we construct an optimal mean-variance portfolio of these risk factors together with a risk-free asset. Similarly, we construct the optimal mean-variance portfolios from the risk factors in $\tilde{\mathbb{M}}_1$, $\mathbb{M}_1$ CAPM, Fama and French family, HXZ, SY, and DHS collections, as well as the risk factors of top models of the winners model scan. This leads to the important question: how do the Sharpe ratios of those different portfolios compare?

We construct these different portfolios in the following manner. Consider model $\mathbb{M}_j$ with associated risk factors x_j. Given the data $y_{1:T}$, consider calculating the *predictive* mean of $x_{j,T+1}$

$$\mathbb{E}[x_{j,T+1}|\mathbb{M}_j,y_{1:T}] \triangleq x_{j,T+1|T} = \int x_{j,T+1}\mathscr{N}_{k_{x,j}}(x_{j,T+1}|\lambda_{x,j},\Omega_{x,j})\pi(\theta_j|\mathbb{M}_j,y_{1:T})\mathrm{d}\theta_j\mathrm{d}x_{j,T+1},$$

which by changing the order of the integration can be seen to just equal the posterior mean of $\lambda_{x,j}$

$$x_{j,T+1|T} = \hat{\lambda}_{x,j} \triangleq \int_{\Theta_j} \lambda_{x,j} \pi(\theta_j|\mathbb{M}_j, y_{1:T}) \mathrm{d}\theta_j,$$

and the *predictive* covariance of $x_{j,T+1}$

$$\Omega_{x,j,T+1|T} \triangleq \int (x_{j,T+1} - x_{j,T+1|T})(x_{j,T+1} - x_{j,T+1|T})' \mathcal{N}_{k_{x,j}}(x_{j,T+1}|\lambda_{x,j}, \Omega_{x,j}) \pi(\theta_j|\mathbb{M}_j, y_{1:T}) \mathrm{d}\theta_j \mathrm{d}x_{j,T+1},$$

which by the law of iterated expectations for covariances simplifies to the sum of the posterior mean of $\Omega_{x,j}$ and the posterior variance of $\lambda_{x,j}$:

$$\Omega_{x,j,T+1|T} = \int_{\Theta_j} \Omega_{x,j} \pi(\theta_j|\mathbb{M}_j, y_{1:T}) \mathrm{d}\theta_j + \int_{\Theta_j} (\lambda_{x,j} - \hat{\lambda}_{x,j})(\lambda_{x,j} - \hat{\lambda}_{x,j})' \pi(\theta_j|\mathbb{M}_j, y_{1:T}) \mathrm{d}\theta_j.$$

Of course, both these quantities are straightforwardly estimated from the output of the simulation of the posterior density $\pi(\theta_j|\mathbb{M}_j, y_{1:T})$. Given these predictive moments, with certain calculations, the Sharpe ratio of the optimal mean-variance portfolio of the factors in x_j plus a risk-free asset is available as

$$\mathrm{Sh}_j = \left(\hat{\lambda}_{x,j}' \Omega_{x,j,T+1|T}^{-1} \hat{\lambda}_{x,j}\right)^{\frac{1}{2}}.$$

In Table 2.10 we report the Sharpe ratios of risk-factor portfolios based on several asset pricing models. In Panel A we consider the risk factor sets of the top models in $\tilde{\mathscr{M}}_*$, in Panel B for the risk factor sets of the top models in $\mathscr{M}_*$, and in Panel C for the $\tilde{\mathbb{M}}_1$, $\mathbb{M}_1$, CAPM, Fama and French family, HXZ, SY, and DHS models.

Looking at the results in Panel B and C, we can see that the Sharpe ratios are much higher for the top models from the winners scan than those from some of the existing asset pricing models. Comparing the results in Panel A and Panel B, we see that the top models from the winners plus genuine anomalies model scan provide even higher Sharpe ratios. Taken together, if we consider the best performing DHS model as the benchmark from Panel C, we can see that the models in $\mathscr{M}_*$ have about 17% higher Sharpe ratios, and the models in $\tilde{\mathscr{M}}_*$ have about 49% higher Sharpe ratios.

Finally, in the winners model scan, all twelve winners can achieve a Sharpe ratio of 0.56, whereas investing in the seven winners of winners in $\mathbb{M}_{2*}$ gives a Sharpe ratio of 0.55. And in the winners plus genuine anomalies model scan, investing in those top risk factor sets produces a Sharpe ratio as high as 0.70 while investing in all twelve winners plus eleven PCs gives 0.71. From these close Sharpe ratios we can make two useful conclusions. First, the portfolios of risk factors from the top models perform as well as those from the complete set of risk factors or; in other words, there is some information redundancy in the potential risk factor set. Second, the marginal likelihood ranking and the Sharpe ratio ranking of models are aligned.

2.7 Conclusion

Our paper makes a contribution to the literature on Bayesian risk factor selection. Starting from the twelve distinct risk factors in Fama and French (1993, 2015, 2018), Hou, Xue, and Zhang (2015), Stambaugh and Yuan (2017), and Daniel, Hirshleifer, and Sun (2020), we construct and compare 4,095 asset pricing models and find that the top models with the highest posterior model probabilities have six risk factors, superior out-of-sample predictive performance, and higher Sharpe ratios. We show also that both fundamental and behavioral risk factors appear in our top model set, highlighting the importance of behavioral factors in pricing and investment decision making.

We also consider if we can get an improved set of risk factors by formulating models that involve the set of winners and additional factors based on the genuine anomalies, i.e., anomalies that cannot be explained by the winners. The framework we have developed, in which we reduce the 125 anomalies to the set of twenty genuine anomalies before converting these genuine anomalies to the space of principal components, allows us to understand whether the PC factors provide incremental information for pricing assets. An extensive comparison of 8,388,607 asset pricing models, constructed from the twelve winners plus eleven principal components of genuine anomalies, shows that there is much to gain by incorporating information in the genuine anomalies. For example, the Sharp ratios increase by more than 30%. Nonetheless, the risk factors from the winners set are the key risk factors, even though some prove to be redundant.

The general approach that we describe in this paper has wide applications. The idea of combining well vetted factors (the winners) with the PCs of less established factors in a model comparison setup, is likely to prove extremely useful beyond this problem to other asset categories such as bonds, currencies and commodities, and is likely to open up many interesting avenues for further research.

Table 2.1: **Winners Definitions**

Factors	Definitions
Mkt	the excess return of the market portfolio
SMB	the return spread between diversified portfolios of small size and big size stocks
HML	the return spread between diversified portfolios of high and low B/M stocks
RMW	the return spread between diversified portfolios of stocks with robust and weak profitability
CMA	the return spread between diversified portfolios of the stocks of low (conservative) and high (aggressive) investment firms
MOM	the momentum factor based on two prior returns
IA	the investment factor based on annual changes in total assets divided by lagged total assets
ROE	the profitability factor based on income before extraordinary items divided by one-quarter-lagged book equity
MGMT	the mispricing factor controlled by management
PERF	the mispricing factor related to performance
PEAD	the short-horizon behavioral factor motivated by investor inattention and evidence of short-horizon under reaction
FIN	the long-horizon behavioral factor exploiting the information in managers' decisions to issue or repurchase equity

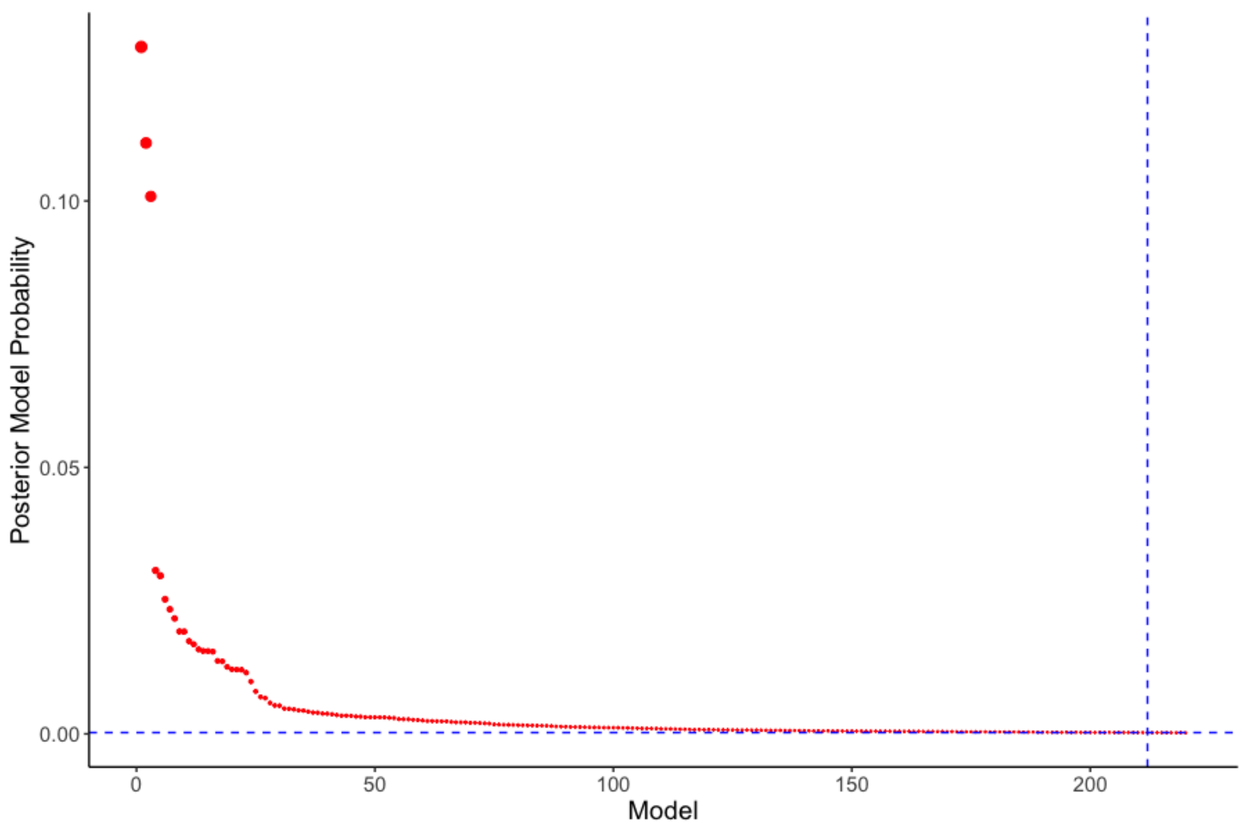

Figure 2.1: **Top 220 Models of the Winners Model Scan**

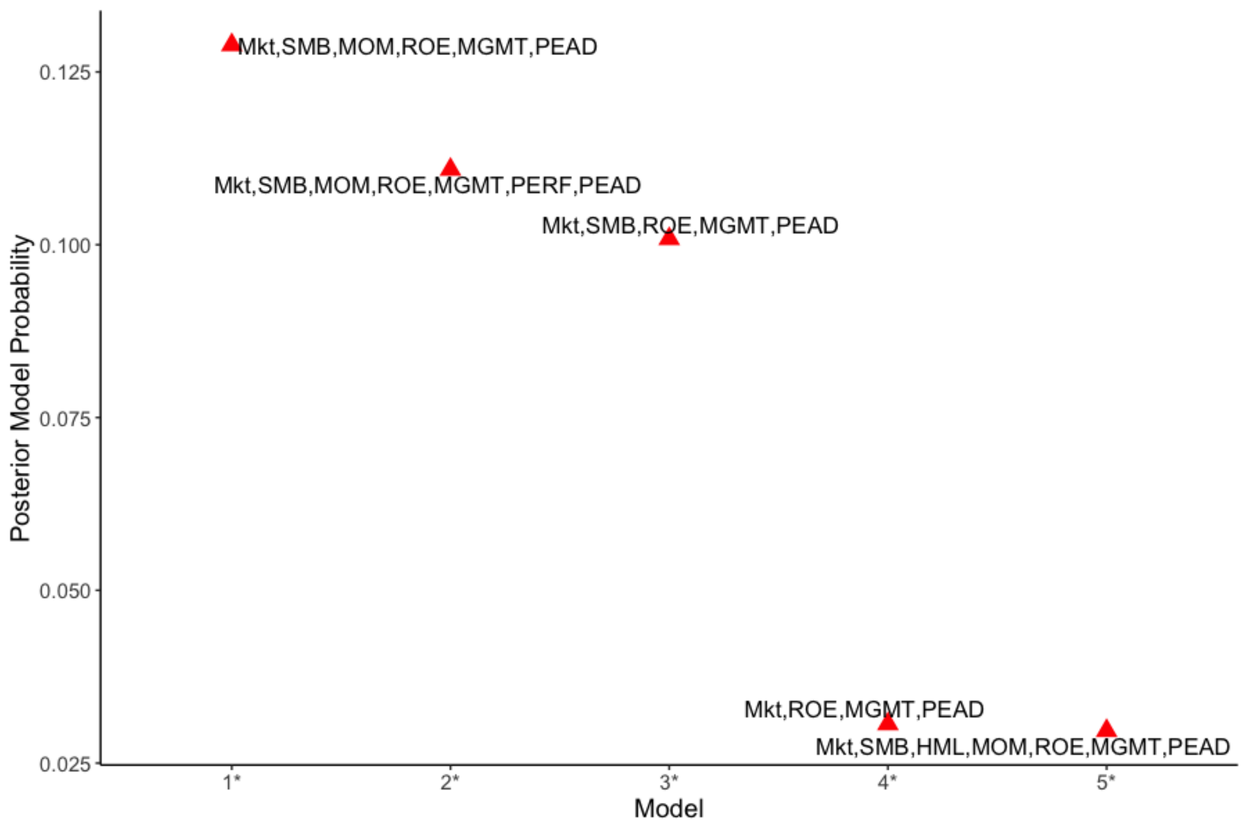

Figure 2.2: **Top 5 Models of the Winners Model Scan**

Table 2.2: **Marginal Likelihoods and the Ratios of Posterior Model Probability and Prior Model Probability of Selected Models in the Winners Model Space $\mathscr{M}$**

Results from the comparison of the $\tilde{J} = 4,095$ models. Panel A has the results for the top three models, and Panel B for $\tilde{\mathbb{M}}_1$ and models with the same risk factor sets as CAPM, FF3, FF5, FF6, HXZ, SY, and DHS.

Risk factors	$\log m_j(y_{1:T}\|\mathbb{M}_j)$	$\frac{\Pr(\mathbb{M}_j\|y_{1:T})}{\Pr(\mathbb{M}_j)}$
Panel A: Top three models		
Mkt, SMB, MOM, ROE, MGMT, PEAD	14186.43	527.89
Mkt, SMB, MOM, ROE, MGMT, PERF, PEAD	14186.28	454.09
Mkt, SMB, ROE, MGMT, PEAD	14186.18	413.01
Panel B: $\tilde{\mathbb{M}}_1$ and models with the same risk factor sets as CAPM, FF3, FF5, FF6, HXZ, SY, and DHS		
12 winners	1.66×10^{-1}	
Mkt	14140.85	8.44×10^{-18}
Mkt, SMB, HML	14140.32	4.98×10^{-18}
Mkt, SMB, HML, RMW, CMA	14152.79	1.30×10^{-12}
Mkt, SMB, HML, RMW, CMA, MOM	14154.45	6.87×10^{-12}
Mkt, SMB, IA, ROE	14164.47	1.54×10^{-7}
Mkt, SMB, MGMT, PERF	14173.32	1.07×10^{-3}
Mkt, PEAD, FIN	14178.86	2.73×10^{-1}

Table 2.3: **Prior and Posterior Statistics of the Risk Premia Parameters of the Best of the Winners Model Scan $\mathbb{M}_{1*}$**

Prior and posterior statistics, i.e. prior mean, posterior mean, posterior standard deviation, posterior median, 2.5% quantile, and 97.5% quantile for the risk premia $\lambda_{x,1*}$ of the best model $\mathbb{M}_1^*$, which has Mkt, SMB, MOM, ROE, MGMT, and PEAD as risk factors.

	Prior mean	Posterior mean	Posterior sd	Posterior median	2.5% Quantile	97.5% Quantile
Mkt	0.0017	0.0066	0.0020	0.0066	0.0026	0.0106
SMB	0.0130	0.0016	0.0013	0.0016	-0.0010	0.0043
MOM	0.0044	0.0061	0.0021	0.0061	0.0020	0.0101
ROE	0.0041	0.0058	0.0012	0.0058	0.0034	0.0081
MGMT	0.0084	0.0056	0.0013	0.0056	0.0030	0.0082
PEAD	0.0085	0.0056	0.0009	0.0056	0.0039	0.0074

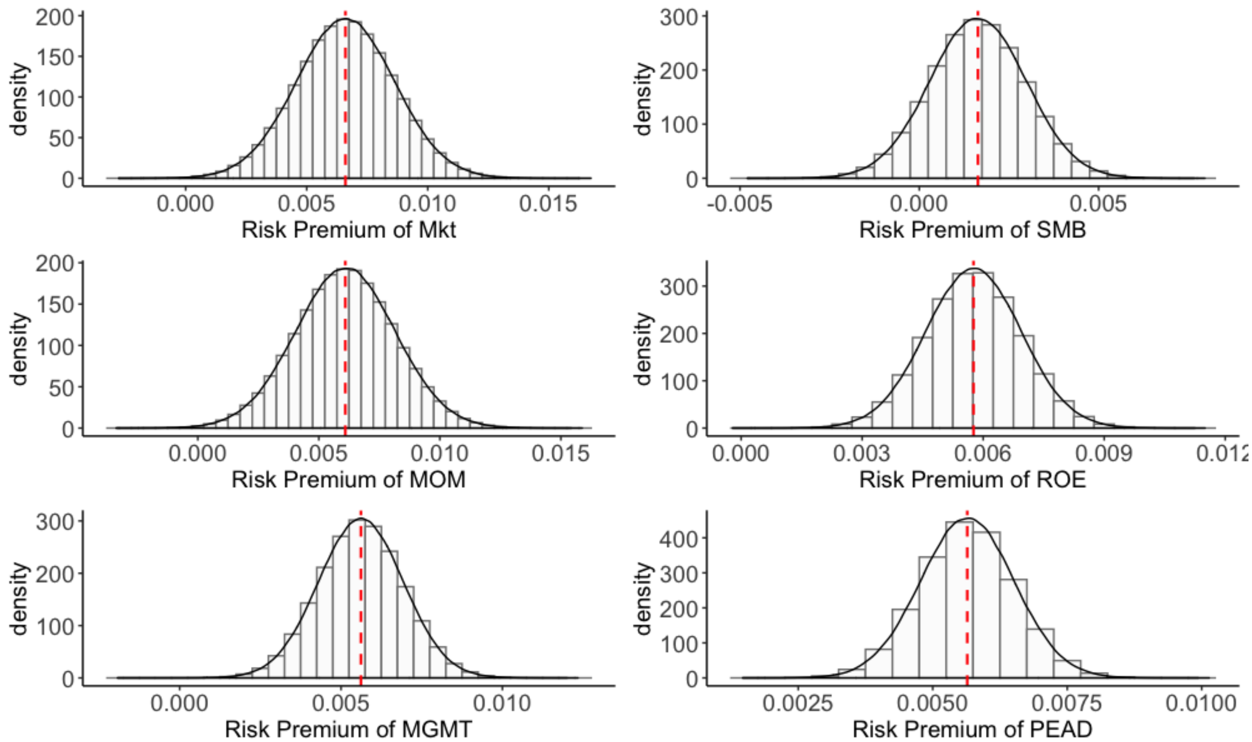

Figure 2.3: **Posterior Distributions of the Risk Premia Parameters of the Best of the Winners Model Scan** $\mathbb{M}_{1*}$

Table 2.4: **Predictive Likelihoods for the Winners Model Scan**
Predictive likelihoods of selected asset pricing models in winners model scan.

Risk factors	$\log m_j(y_{(T+1):(T+12)} \vert \mathbb{M}_j)$
Panel A: Top three models	
Mkt, SMB, MOM, ROE, MGMT, PEAD	383.48
Mkt, SMB, MOM, ROE, MGMT, PERF, PEAD	383.65
Mkt, SMB, ROE, MGMT, PEAD	383.55
Panel B: Models with the same risk factor sets as CAPM, FF3, FF5, FF6, HXZ, SY, and DHS	
Mkt	382.46
Mkt, SMB, HML	381.76
Mkt, SMB, HML, RMW, CMA	381.67
Mkt, SMB, HML, CMA, RMW, MOM	381.83
Mkt, SMB, IA, ROE	381.87
Mkt, SMB, MGMT, PERF	382.89
Mkt, FIN, PEAD	382.76

Table 2.5: **Surviving Anomalies Explanations**

Anomalies	Explanations
acc	annual income before extraordinary items minus operating cash flows divided by average total assets
age	number of years since first Compustat coverage
currat	current assets / current liabilities
hire	percent change in number of employees
lev	total liabilities divided by fiscal year-end market capitalization
quick	(current assets - inventory) / current liabilities
salecash	annual sales divided by cash and cash equivalents
sgr	annual percent change in sales (sale)
Em	enterprise value divided by operating income before depreciation (Compustat annual item OIBDP)
Lbp	leverage component of book to price
dFin	the change in net financial assets
Cop	cash-based operating profitability
Lfe	labor force efficiency
SA	SA index measuring financial constraint
sue	the high-minus-low earnings surprise
cash	cash and cash equivalents divided by average total assets
OLAQ	quarterly operating profits-to-lagged assets
CLAQ	quarterly cash-based operating profits-to-lagged assets
TBIQ	quarterly taxable income-to-book income
BLQ	quarterly book leverage

Table 2.6: **Importance of Principal Components**

	PC1	PC2	PC3	PC4	PC5
Standard deviation	0.1143	0.0785	0.0441	0.0415	0.0408
Proportion of Variance	0.3953	0.1866	0.0590	0.0521	0.0503
Cumulative Proportion	0.3953	0.5819	0.6409	0.6930	0.7433
	PC6	PC7	PC8	PC9	PC10
Standard deviation	0.0366	0.0331	0.0298	0.0282	0.0274
Proportion of Variance	0.0405	0.0332	0.0269	0.0241	0.0227
Cumulative Proportion	0.7837	0.8170	0.8438	0.8680	0.8907
	PC11	PC12	PC13	PC14	PC15
Standard deviation	0.0257	0.0231	0.0218	0.0201	0.0194
Proportion of Variance	0.0200	0.0162	0.0144	0.0122	0.0113
Cumulative Proportion	0.9106	0.9268	0.9412	0.9535	0.9648
	PC16	PC17	PC18	PC19	PC20
Standard deviation	0.0172	0.0162	0.0158	0.0142	0.0125
Proportion of Variance	0.0090	0.0079	0.0075	0.0061	0.0047
Cumulative Proportion	0.9737	0.9817	0.9892	0.9953	1.0000

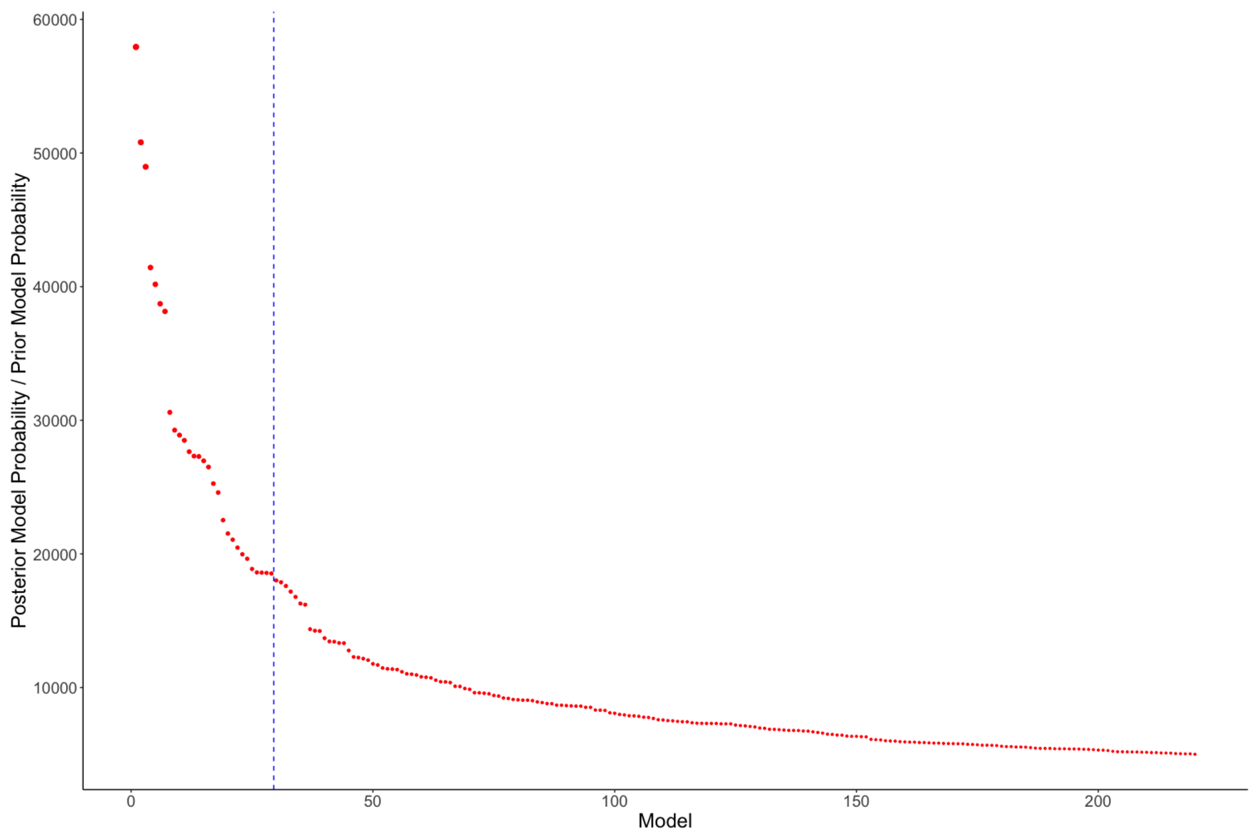

Figure 2.4: **Top 220 Models of the Winners Plus Genuine Anomalies Model Scan**

Table 2.7: **Marginal Likelihoods and Ratios of Posterior Model Probability and Prior Model Probability of Selected Models in the Winners Plus Genuine Anomalies Model Space $\tilde{\mathscr{M}}$**

Risk Factors	$\log m_j(\tilde{y}_{1:T}\|\tilde{\mathbb{M}}_j)$	$\frac{\Pr(\tilde{y}_{1:T}\|\tilde{\mathbb{M}}_j)}{\Pr(\tilde{\mathbb{M}}_j)}$
Panel A: Top models in $\tilde{\mathscr{M}}_*$		
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	24621.85	57939.28
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7	24621.72	50803.12
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	24621.68	48969.29
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5	24621.51	41429.51
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC3, PC4, PC5, PC7	24621.48	40171.69
Mkt, RMW, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	24621.44	38720.10
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5	24621.43	38145.58
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC3, PC4, PC5	24621.21	30594.32
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC4, PC5	24621.16	29262.14
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7, PC9	24621.15	28892.28
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC4, PC5, PC7	24621.14	28496.78
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC4, PC5, PC7	24621.11	27648.27
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC4, PC5	24621.10	27321.30
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	24621.09	27292.03
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5	24621.08	26966.60
Mkt, MOM, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	24621.07	26499.81
Mkt, RMW, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5	24621.02	25260.55
Mkt, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	24620.99	24590.60
Mkt, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	24620.90	22526.00
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7, PC9	24620.86	21524.20
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC9	24620.84	21061.93
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7	24620.81	20467.26
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC4, PC5, PC7	24620.78	19969.41
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7, PC9	24620.77	19636.23
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC4, PC5	24620.73	18874.75
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5	24620.71	18612.73
Mkt, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	24620.71	18597.04
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC3, PC4, PC5, PC7	24620.71	18577.15
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5	24620.71	18531.14
Panel B: $\tilde{\mathbb{M}}_1$ and models with the same risk factor sets as CAPM, FF3, FF5, FF6, HXZ, SY, and DHS		
12 winners and 11 PCs	24606.63	1.71×10^{-9}
Mkt	24560.93	2.42×10^{-29}
Mkt, SMB, HML	24560.17	1.13×10^{-29}
Mkt, SMB, HML, RMW, CMA	24572.01	1.57×10^{-24}
Mkt, SMB, HML, RMW, CMA, MOM	24573.47	6.74×10^{-24}
Mkt, SMB, IA, ROE	24583.58	1.66×10^{-19}
Mkt, SMB, MGMT, PERF	24592.21	9.25×10^{-16}
Mkt, PEAD, FIN	24597.68	2.20×10^{-13}
Panel C: Models with the same risk factor sets as the top three models in the winners model scan		
Mkt, SMB, MOM, ROE, MGMT, PEAD	24604.69	2.44×10^{-10}
Mkt, SMB, MOM, ROE, MGMT, PERF, PEAD	24604.41	1.84×10^{-10}
Mkt, SMB, ROE, MGMT, PEAD	24604.60	2.22×10^{-10}

Table 2.8: **Prior and Posterior Statistics of the Risk Premia Parameters of the Best of the Winners Plus Genuine Model Scan $\tilde{\mathbb{M}}_{1*}$**

Prior and posterior statistics, i.e. prior mean, posterior mean, posterior standard deviation, posterior median, 2.5% quantile, and 97.5% quantile for the risk premia $\tilde{\lambda}_{x,1*}$ of the best model $\mathbb{M}_1^*$, which has Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, and PC7 as risk factors.

	Prior mean	Posterior mean	Posterior sd	Posterior median	2.5% Quantile	97.5% Quantile
Mkt	0.0017	0.0066	0.0021	0.0066	0.0025	0.0106
RMW	-0.0015	0.0036	0.0011	0.0036	0.0014	0.0057
MOM	0.0044	0.0061	0.0021	0.0061	0.0020	0.0102
IA	0.0075	0.0033	0.0009	0.0033	0.0016	0.0050
ROE	0.0041	0.0058	0.0012	0.0058	0.0034	0.0081
MGMT	0.0084	0.0056	0.0013	0.0056	0.0030	0.0082
PEAD	0.0085	0.0056	0.0009	0.0056	0.0039	0.0074
FIN	0.0077	0.0070	0.0018	0.0070	0.0034	0.0106
PC1	0.0044	-0.0006	0.0056	-0.0006	-0.0117	0.0103
PC3	0.0191	0.0082	0.0020	0.0082	0.0043	0.0121
PC4	0.0221	0.0112	0.0018	0.0112	0.0076	0.0147
PC5	0.0009	0.0013	0.0019	0.0013	-0.0024	0.0051
PC7	-0.0005	-0.0011	0.0015	-0.0011	-0.0040	0.0018

Table 2.9: **Predictive Likelihoods for the Winners Plus Genuine Anomalies Model Scan**
Predictive likelihoods of selected models in the winners plus genuine anomalies model scan.

Risk factors	$\log m_j(\tilde{y}_{(T+1):(T+12)}\|\tilde{\mathbb{M}}_j)$
Panel A: Top models in $\tilde{\mathscr{M}}_*$	
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	639.27
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7	639.36
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	639.46
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5	639.01
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC3, PC4, PC5, PC7	639.86
Mkt, RMW, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	639.09
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5	639.08
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC3, PC4, PC5	639.59
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC4, PC5	639.36
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7, PC9	639.32
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC4, PC5, PC7	639.65
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC4, PC5, PC7	639.16
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC4, PC5	638.86
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	639.35
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5	639.23
Mkt, MOM, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	639.53
Mkt, RMW, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5	638.82
Mkt, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	639.48
Mkt, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	639.41
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7, PC9	639.43
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC9	639.08
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7	639.25
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC4, PC5, PC7	639.71
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7, PC9	639.19
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC4, PC5	639.42
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5	638.95
Mkt, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	639.26
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC3, PC4, PC5, PC7	639.90
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5	639.09
Panel B: $\tilde{\mathbb{M}}_1$ and models with risk factor sets same as CAPM, FF3, FF5, FF6, HXZ, SY, and DHS	
12 winners and 11 PCs	638.61
Mkt	640.03
Mkt, SMB, HML	639.35
Mkt, SMB, HML, RMW, CMA	639.27
Mkt, SMB, HML, CMA, RMW, MOM	639.43
Mkt, SMB, IA, ROE	639.48
Mkt, SMB, MGMT, PERF	640.48
Mkt, FIN, PEAD	640.35

Table 2.10: **Sharpe Ratios**
Sharpe ratios for the risk factor sets of selected asset pricing models based on $G = 100,000$.

Risk factors	Sharpe ratios
Panel A: Risk factor sets of the top models in $\tilde{\mathscr{M}}_*$	
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	0.69
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7	0.68
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	0.69
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5	0.68
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC3, PC4, PC5, PC7	0.68
Mkt, RMW, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	0.69
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5	0.67
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC3, PC4, PC5	0.67
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC4, PC5	0.66
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7, PC9	0.69
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, PC1, PC4, PC5, PC7	0.67
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC4, PC5, PC7	0.68
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC4, PC5	0.67
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	0.70
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5	0.68
Mkt, MOM, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	0.68
Mkt, RMW, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5	0.68
Mkt, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	0.67
Mkt, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	0.68
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5, PC7, PC9	0.70
Mkt, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC9	0.68
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5, PC7	0.69
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC4, PC5, PC7	0.68
Mkt, RMW, MOM, IA, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7, PC9	0.70
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC4, PC5	0.67
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC4, PC5	0.68
Mkt, CMA, MOM, ROE, MGMT, PEAD, FIN, PC1, PC3, PC4, PC5, PC7	0.68
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, PC1, PC3, PC4, PC5, PC7	0.68
Mkt, RMW, MOM, IA, ROE, MGMT, PERF, PEAD, FIN, PC1, PC3, PC4, PC5	0.69
Panel B: Risk factor sets of the top three models in the winners model scan	
Mkt, SMB, MOM, ROE, MGMT, PEAD	0.54
Mkt, SMB, MOM, ROE, MGMT, PERF, PEAD	0.55
Mkt, SMB, ROE, MGMT, PEAD	0.53
Panel C: Risk factor sets of $\tilde{\mathbb{M}}_1$, $\mathbb{M}_1$, CAPM, FF3, FF5, FF6, HXZ, SY, and DHS models	
12 winners and 11 PCs	0.71
12 winners	0.56
Mkt	0.15
Mkt, SMB, HML	0.20
Mkt, SMB, HML, RMW, CMA	0.34
Mkt, SMB, HML, CMA, RMW, MOM	0.36
Mkt, SMB, IA, ROE	0.40
Mkt, SMB, MGMT, PERF	0.45
Mkt, FIN, PEAD	0.47

Bibliography

Ang, A. and Timmermann, A. (2012), "Regime Changes and Financial Markets," *Annu. Rev. Financ. Econ.*, 4, 313–337.

Asness, C. and Frazzini, A. (2013), "The devil in HML's details," *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.

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

Avramov, D. and Chao, J. C. (2006), "An Exact Bayes Test of Asset Pricing Models with Application to International Markets," *Journal of Business*, 79, 293–324.

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

Bryzgalova, S., Huang, J., and Julliard, C. (2019), "Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models," *Available at*, SSRN.

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

Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S., Shin, M., and Simoni, A. (2018), "Bayesian Estimation and Comparison of Moment Condition Models," *Journal of the American Statistical Association*, 113, 1656–1668.

Chib, S. and Zeng, X. (2019), "Which Factors are Risk Factors in Asset Pricing? A Model Scan Framework," *Journal of Business & Economic Statistics*, 0, 1–28.

Chib, S., Zeng, X., and Zhao, L. (2020), "On Comparing Asset Pricing Models," *Journal of Finance*, 75, 551–577.

Cochrane, J. H. (2011), "Presidential Address: Discount Rates," *Journal of Finance*, 66, 1047–1108.

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

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

— (2015), "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

— (2018), "Choosing factors," *Journal of Financial Economics*, 128, 234–252.

Frazzini, A. and Pedersen, L. H. (2014), "Betting Against Beta," *Journal of Financial Economics*, 111, 1–25.

Goyal, A., He, Z. L., and Huh, S.-W. (2018), "Distance-Based Metrics: A Bayesian Solution to the Power and Extreme-Error Problems in Asset-Pricing Tests," *Swiss Finance Institute Research Paper*.

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

Hansen, L. P. and Jagannathan, R. (1991), "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

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

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

— (2017), "Replicating Anomalies," *Review of Financial Studies*, forthcoming.

Jeffreys, H. (1961), "Theory of Probability." 3rd Edn Oxford: Oxford University Press.

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

Novy-Marx, R. and Velikov, M. (2016), "A Taxonomy of Anomalies and Their Trading Costs," *Review of Financial Studies*, 29, 104–147.

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

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

Pástor, L. and Stambaugh, R. F. (2000), "Comparing asset pricing models: an investment perspective," *Journal of Financial Economics*, 56, 335–381.

— (2001), “The Equity Premium and Structural Breaks,” *Journal of Finance*, 56, 1207–1239.

Patton, A. J. and Weller, B. M. (2020), “What you see is not what you get: The costs of trading market anomalies,” *Journal of Financial Economics*, forthcoming.

Stambaugh, R. F. and Yuan, Y. (2017), “Mispricing Factors,” *Review of Financial Studies*, 30, 1270–1315.

Appendix

Appendix A

In this appendix, we describe how the 33 DGPs for generating the simulated data in the eight-factor experiment were selected (the same process is used in the twelve-factor case, and thus for brevity the details of that case are suppressed). Supposes that

$$x_t = (\text{Mkt}, f_t)'$$

consists of $\{\text{Mkt}, \text{SMB}, \text{ROE}, \text{IA}\}$. Then the DGP is given by

$$\text{Mkt}_t = \mu_m + \varepsilon_{m,t}, \quad \varepsilon_{m,t} \sim \mathcal{N}(0, \sigma_m^2) \tag{A1}$$

$$\underbrace{\begin{pmatrix} \text{SMB}_t \\ \text{ROE}_t \\ \text{IA}_t \end{pmatrix}}_{f_t:3\times 1} = \underbrace{\begin{pmatrix} \alpha_s \\ \alpha_o \\ \alpha_i \end{pmatrix}}_{\alpha:3\times 1} + \underbrace{\begin{pmatrix} \beta_{sm} \\ \beta_{em} \\ \beta_{im} \end{pmatrix}}_{\beta:3\times 1} \text{Mkt}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \underbrace{\Sigma}_{3\times 3}) \tag{A2}$$

$$\underbrace{\begin{pmatrix} \mathrm{HML}_t \\ \mathrm{RMW}_t \\ \mathrm{CMA}_t \\ \mathrm{MOM}_t \end{pmatrix}}_{f_t^*:4\times 1} = \underbrace{\begin{pmatrix} \beta_{hm}^* & \beta_{hs}^* & \beta_{he}^* & \beta_{hi}^* \\ \beta_{rm}^* & \beta_{rs}^* & \beta_{re}^* & \beta_{ri}^* \\ \beta_{cm}^* & \beta_{cs}^* & \beta_{ce}^* & \beta_{ci}^* \\ \beta_{om}^* & \beta_{os}^* & \beta_{oe}^* & \beta_{oi}^* \end{pmatrix}}_{\beta^*:4\times 4} \begin{pmatrix} \mathrm{Mkt}_t \\ \mathrm{SMB}_t \\ \mathrm{ROE}_t \\ \mathrm{IA}_t \end{pmatrix} + \varepsilon_t^*, \quad \varepsilon_t^* \sim \mathcal{N}(0, \underbrace{\Sigma^*}_{4\times 4}). \tag{A3}$$

To generate data from the DGP, we have to fix the parameters at some suitable values. A sensible choice is to fix the parameters at the maximum likelihood (ML) values to ensure that the generated data resemble the real data. A key point is that we should ensure that the generating DGP is a valid model for the purpose of generating our data. By "valid model" we mean a model in which the fitted stochastic discount factor (SDF) suggests that each assumed risk factor is statistically significant. In other words, if we let the SDF be given by

$$M_t = 1 - \lambda_x' \Omega_x^{-1}(x_t - \mu_x),$$

the fitted values of $b' = \lambda_x' \Omega_x^{-1}$ should each be significant. Otherwise, the maintained assumption that the factors $\{\mathrm{Mkt}, \mathrm{SMB}, \mathrm{ROE}, \mathrm{IA}\}$ are the risk factors would be counter to the evidence and the data generated from such a DGP would lead to misleading model comparisons. To isolate the models that we can use to generate the data, we find the ML estimates of b for each of the 128 models from monthly data on the aforementioned risk factors that run from January 1968 to December 2015 with 576 observations in total.[6] We select these DGPs by fitting each of the 128 possible models to the actual data by ML and then checking whether any component of the vector $b' = \lambda_x' \Omega_x^{-1}$ is insignificant. If any component is insignificant, the collection of factors $x_t = (\mathrm{Mkt}, f_t)'$ in that model are not used as a DGP in the simulation exercise.

For each of the 128 possible models, the ML estimates of the parameters and of b are obtained

[6]The Mkt, SMB, HML, RMW, and CMA factors are from Kenneth French's website.We thank the authors for making the data available. We also thank Lu Zhang for providing us the ME, ROE, and IA factors.

as follows. Given a particular $x_t = (\text{Mkt}, f_t)'$ and under the pricing restrictions, any one of candidate factor models takes the form

$$x_t = \Omega_x b + \eta_{x,t} \tag{A4}$$

$$f_t^* = \beta^* x_t + \varepsilon_t^*, \quad \varepsilon_t^* \sim \mathcal{N}_{K-L}(0, \Sigma^*), \tag{A5}$$

where

$$\begin{pmatrix} \eta_{x,t} \\ \varepsilon_t^* \end{pmatrix} \sim \mathcal{N}_K \left(0, \begin{pmatrix} \Omega_x & 0 \\ 0 & \Sigma^* \end{pmatrix} \right). \tag{A6}$$

Using the data from January 1968 to December 2015, we find the estimate of b by maximizing the log-likelihood function of the model implied by (A4) to (A6) and calculate the variance-covariance matrix of the estimate as the negative inverse of the Hessian matrix of the likelihood function at the ML estimate. A model is used to generate data in our simulation experiments if each element in b is significant at the 5% level.

Appendix B

In this appendix we give the proof of the Jacobian term used in Proposition 3.

Proof. By definition, the Jacobian is

$$\left| \det \left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'} \right) \right| = \left| \det \begin{pmatrix} \frac{\partial \beta_1}{\partial \beta_j'} & \frac{\partial \beta_1}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \beta_1}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \beta_1}{\partial \sigma_j'} & \frac{\partial \beta_1}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1}{\partial \beta_j'} & \frac{\partial \sigma_1}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1}{\partial \sigma_j'} & \frac{\partial \sigma_1}{\partial \sigma_j^{*\prime}} \end{pmatrix} \right|.$$

Partition β_1 and σ_1,

$$\beta_1 = \begin{pmatrix} \beta_1^f : (L_j - 1) \times 1 \\ \beta_1^{f*} : (K - L_j) \times 1 \end{pmatrix}$$

$$
\sigma_1 = \begin{pmatrix} \sigma_1^f : \frac{L_j(L_j-1)}{2} \times 1 \\ \sigma_1^{ff*} : (K-L_j)(L_j-1) \times 1 \\ \sigma_1^{f*} : \frac{(K-L_j)(K-L_j+1)}{2} \times 1 \end{pmatrix},
$$

so that the Jacobian can be rewritten as

$$
\left| \det \left(\frac{g_j^{-1}(\eta_j)}{\partial \eta_j'} \right) \right| = \left| \det \begin{pmatrix} \frac{\partial \beta_1^f}{\partial \beta_j'} & \frac{\partial \beta_1^f}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \beta_1^f}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \beta_1^f}{\partial \sigma_j'} & \frac{\partial \beta_1^f}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \beta_1^{f*}}{\partial \beta_j'} & \frac{\partial \beta_1^{f*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \beta_1^{f*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \beta_1^{f*}}{\partial \sigma_j'} & \frac{\partial \beta_1^{f*}}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1^f}{\partial \beta_j'} & \frac{\partial \sigma_1^f}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1^f}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1^f}{\partial \sigma_j'} & \frac{\partial \sigma_1^f}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1^{ff*}}{\partial \beta_j'} & \frac{\partial \sigma_1^{ff*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1^{ff*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1^{ff*}}{\partial \sigma_j'} & \frac{\partial \sigma_1^{ff*}}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1^{f*}}{\partial \beta_j'} & \frac{\partial \sigma_1^{f*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1^{f*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1^{f*}}{\partial \sigma_j'} & \frac{\partial \sigma_1^{f*}}{\partial \sigma_j^{*\prime}} \end{pmatrix} \right|.
$$

Since the partial derivatives in the first row are

$$
\frac{\partial \beta_1^f}{\partial \beta_j'} = I_{L_j-1}, \frac{\partial \beta_1^f}{\partial \beta_{j,m}^{*\prime}} = 0, \frac{\partial \beta_1^f}{\partial \beta_{j,f}^{*\prime}} = 0, \frac{\partial \beta_1^f}{\partial \sigma_j'} = 0, \frac{\partial \beta_1^f}{\partial \sigma_j^{*\prime}} = 0,
$$

we have that

$$
\left| \det \left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'} \right) \right| = \left| \det \begin{pmatrix} \frac{\partial \beta_1^{f*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \beta_1^{f*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \beta_1^{f*}}{\partial \sigma_j'} & \frac{\partial \beta_1^{f*}}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1^f}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1^f}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1^f}{\partial \sigma_j'} & \frac{\partial \sigma_1^f}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1^{ff*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1^{ff*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1^{ff*}}{\partial \sigma_j'} & \frac{\partial \sigma_1^{ff*}}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1^{f*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1^{f*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1^{f*}}{\partial \sigma_j'} & \frac{\partial \sigma_1^{f*}}{\partial \sigma_j^{*\prime}} \end{pmatrix} \right|.
$$

In addition, because

$$
\frac{\partial \sigma_1^f}{\partial \sigma_j'} = I_{\frac{L_j(L_j-1)}{2}}, \frac{\partial \sigma_1^f}{\partial \beta_{j,m}^{*\prime}} = 0, \frac{\partial \sigma_1^f}{\partial \beta_{j,f}^{*\prime}} = 0, \frac{\partial \sigma_1^f}{\partial \sigma_j^{*\prime}} = 0,
$$

we can further reduce the Jacobian to

$$\left|\det\left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'}\right)\right| = \left|\det\begin{pmatrix} \frac{\partial \beta_1^{f*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \beta_1^{f*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \beta_1^{f*}}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1^{ff*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1^{ff*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1^{ff*}}{\partial \sigma_j^{*\prime}} \\ \frac{\partial \sigma_1^{f*}}{\partial \beta_{j,m}^{*\prime}} & \frac{\partial \sigma_1^{f*}}{\partial \beta_{j,f}^{*\prime}} & \frac{\partial \sigma_1^{f*}}{\partial \sigma_j^{*\prime}} \end{pmatrix}\right|.$$

Finally, because

$$\frac{\partial \beta_1^{f*}}{\partial \beta_{j,m}^{*\prime}} = I_{K-L_j}, \frac{\partial \sigma_1^{ff*}}{\partial \beta_{j,m}^{*\prime}} = 0, \frac{\partial \sigma_1^{f*}}{\partial \beta_{j,m}^{*\prime}} = 0$$

and

$$\frac{\partial \sigma_1^{f*}}{\partial \sigma_j^{*\prime}} = I_{\frac{(K-L_j)(K-L_j+1)}{2}}, \frac{\partial \beta_1^{f*}}{\partial \sigma_j^{*\prime}} = 0, \frac{\partial \sigma_1^{ff*}}{\partial \sigma_j^{*\prime}} = 0,$$

the determinant can be evaluated as

$$\begin{aligned} \left|\det\left(\frac{\partial g_j^{-1}(\eta_j)}{\partial \eta_j'}\right)\right| &= \left|\det\left(\frac{\partial \sigma_1^{ff*}}{\partial \beta_{j,f}^{*\prime}}\right)\right| \\ &= \left|\det\left(\frac{\partial\left(\Sigma_j \otimes I_{K-L_j}\right)\beta_{j,f}^*}{\partial \beta_{j,f}^{*\prime}}\right)\right| \\ &= |\Sigma_j \otimes I_{K-L_j}| \\ &= |\Sigma_j|^{K-L_j}. \end{aligned}$$

□

Appendix C

In this appendix we provide detail for calculation of CZZ Marginal Likelihoods: General Case. We calculate marginal likelihood using the *basic marginal likelihood identity* in Chib (1995): for any θ, it holds that

$$m(y) = \frac{p(y|\theta)\pi(\theta)}{\pi(\theta|y)}, \tag{C7}$$

where the numerator is just the product of the sampling density (likelihood function) and the prior density of parameters, with all integrating constants included, and the denominator is the posterior density of θ.

We provide Bayes update results for multivariate normal regression. For $t = 1, 2, \ldots, T$,

$$\underbrace{y_t}_{d\times 1} = \underbrace{B}_{d\times k}\underbrace{x_t}_{k\times 1} + \underbrace{e_t}_{d\times 1}, \quad e_t \stackrel{\text{i.i.d.}}{\sim} N_d(0, \Sigma).$$

It can be rewritten as matrix form, i.e.

$$\underbrace{Y}_{T\times d} = \underbrace{X}_{T\times k}\underbrace{B'}_{k\times d} + \underbrace{E}_{T\times d},$$

and after taking transpose and vec we get

$$y = (X \otimes I_d)\beta + e, \quad e \sim \mathcal{N}(0, I_T \otimes \Sigma),$$

where

$$y = \text{vec}(Y') \quad \text{and} \quad \beta = \text{vec}(B).$$

Throughout this appendix, the hat denote the least squares estimators of parameters. The least squares estimator of B' is given by

$$\hat{B}' = (X'X)^{-1}X'y,$$

so

$$\hat{\beta} = \text{vec}\left(\hat{B}\right) = \left((X'X)^{-1}X' \otimes I_d\right) y.$$

Suppose that

$$\beta \sim \mathcal{N}(\beta_0, D_0^{-1}),$$

where D_0 is precision matrix. Applying Bayes update

$$\beta | y, \Sigma \sim \mathcal{N}(\beta_1, D_1^{-1}),$$

where

$$D_1 = D_0 + (X'X \otimes \Sigma^{-1}) \quad \text{and} \quad \beta_1 = D^{-1}\left(D_0\beta_0 + \left(X'X \otimes \Sigma^{-1}\right)\hat{\beta}_{LS}\right).$$

In model $\mathscr{M}_j$, $j = 2, 3, ..., \tilde{J}$, let $\tilde{f}_j$ denote the risk factors with dimension $L_j \times 1$ and let f_j^* denote the non-risk factors with dimension $(K - L_j) \times 1$. The model is given by

$$\tilde{f}_{j,t} = \tilde{\alpha}_j + \tilde{\varepsilon}_{j,t}, \quad \tilde{\varepsilon}_{j,t} \sim \mathcal{N}_{L_j}\left(0, \Sigma_j\right) \tag{C8}$$

$$f_{j,t}^* = B_{j,f}^* \tilde{f}_{j,t} + \varepsilon_{j,t}^*, \quad \varepsilon_{j,t}^* \sim \mathcal{N}_{K-L_j}\left(0, \Sigma_j^*\right) \tag{C9}$$

and its nuisance parameters are

$$\eta_j = \left(\beta_{j,f}^*, \sigma_j, \sigma_j^*\right),$$

where $\beta_{j,f}^* = \mathsf{vec}(B_{j,f}^*)$, $\sigma_j = \mathsf{vech}(\Sigma_j)$, and $\sigma_j^* = \mathsf{vech}(\Sigma_j^*)$. Under the collection of priors in Proposition 4, with the value of c set equal to 1, and $\tilde{\alpha}_j | \mathscr{M}_j \sim \mathcal{N}_{L_j}(\tilde{\alpha}_{j0}, k_j \Sigma_j)$, the marginal likelihood of model $\mathscr{M}_j$, $j = 2, 3, ..., \tilde{J}$, on the log-scale is given by

$$\log \tilde{m}(y_{n_t+1:T} | \mathscr{M}_j) = \log \tilde{m}(\tilde{f}_{j,n_t+1:T} | \mathscr{M}_j) + \log \tilde{m}(f_{j,n_t+1:T}^* | \mathscr{M}_j). \tag{C10}$$

To calculate the marginal likelihood on the log-scale for the top model, we apply the equation (C7) on the log-scale

$$\begin{aligned} \log \tilde{m}(\tilde{f}_{j,n_t+1:T} | \mathscr{M}_j) = {} & \log p(\tilde{f}_{j,n_t+1:T} | \mathscr{M}_j, \tilde{\alpha}_j, \Sigma_j) \\ & + \log \pi(\tilde{\alpha}_j | \mathscr{M}_j, \Sigma_j) + \log \tilde{\psi}(\Sigma_j | \mathscr{M}_j) \\ & - \log \pi(\tilde{\alpha}_j | \mathscr{M}_j, \Sigma_j, \tilde{f}_{j,n_t+1:T}) - \log \pi(\Sigma_j | \mathscr{M}_j, \tilde{f}_{j,n_t+1:T}), \end{aligned} \tag{C11}$$

where the first three terms are sampling density and the prior density of parameters:

$$p(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j,\tilde{\alpha}_j,\Sigma_j) = \prod_{t=n_t+1}^{T} \mathscr{N}_{L_j}(\tilde{f}_{j,t}|\tilde{\alpha}_j,\Sigma_j) \tag{C12}$$

$$\pi(\tilde{\alpha}_j|\mathscr{M}_j,\Sigma_j) = \mathscr{N}_{L_j}(\tilde{\alpha}_j|\tilde{\alpha}_{j0},k_j\Sigma_j) \tag{C13}$$

$$\tilde{\psi}(\Sigma_j|\mathscr{M}_j) = |\Sigma_j|^{-\frac{2L_j-K+1}{2}}. \tag{C14}$$

The fourth term can be obtained by applying the results of multivariate normal regression for $t = n_t+1,\ldots,T$, thus

$$\pi(\tilde{\alpha}_j|\mathscr{M}_j,\Sigma_j,\tilde{f}_{j,n_t+1:T}) = \mathscr{N}_{L_j}(\tilde{\alpha}_j|\tilde{\alpha}_{j1},D_{j1}^{-1}), \tag{C15}$$

where

$$\tilde{\alpha}_{j1} = \frac{1}{\tilde{T}k_j+1}\tilde{\alpha}_{j0} + \frac{\tilde{T}k_j}{\tilde{T}k_j+1}\hat{\tilde{\alpha}}_j \quad \text{and} \quad D_{j1} = (\frac{1}{k_j}+\tilde{T})\Sigma_j^{-1}. \tag{C16}$$

To obtain the fifth term, we calculate

$$\begin{aligned}
&p(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j,\Sigma_j)\\
=&\int p(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j,\tilde{\alpha}_j,\Sigma_j)\pi(\tilde{\alpha}_j|\mathscr{M}_j,\Sigma_j)\mathrm{d}\tilde{\alpha}_j\\
=&\int p(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j,a_j,\Sigma_j)\pi(a_j|\mathscr{M}_j,\Sigma_j)\mathrm{d}a_j,
\end{aligned}$$

where $a_j = \tilde{\alpha}_j - \tilde{\alpha}_{j0}$ and $\pi(a_j|\mathscr{M}_j,\Sigma_j) = \mathscr{N}_{L_j}(a_j|0,k_j\Sigma_j)$. Thus

$$\begin{aligned}
&p(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j,\Sigma_j)\\
=&\int (2\pi)^{-\frac{\tilde{T}L_j}{2}}|\Sigma_j|^{-\frac{\tilde{T}}{2}}e^{-\frac{1}{2}SSE}e^{-\frac{1}{2}(a_j-\hat{a}_j)'(\tilde{T}\Sigma_j^{-1})(a_j-\hat{a}_j)}(2\pi)^{-\frac{L_j}{2}}|k_j\Sigma_j|^{-\frac{1}{2}}e^{-\frac{1}{2}a_j'(k_j^{-1}\Sigma_j^{-1})a_j}\mathrm{d}a_j\\
=&|\tilde{T}k_j+1|^{\frac{\tilde{T}}{2}}(2\pi)^{-\frac{\tilde{T}L_j}{2}}|\Sigma_j|^{-\frac{\tilde{T}}{2}}e^{-\frac{1}{2}(SSE+R)}\\
&\int (2\pi)^{-\frac{L_j}{2}}|(k_j^{-1}+T)^{-1}\Sigma_j|^{-\frac{1}{2}}e^{-\frac{1}{2}(a_j-\hat{a}_j)'\left((k_j^{-1}+T)\Sigma_j^{-1}\right)(a_j-\hat{a}_j)}\mathrm{d}a_j
\end{aligned}$$

where the integral is equal to one and

$$
\begin{aligned}
SSE + R = &\sum_{t=n_t+1}^{T} (\tilde{f}_{j,t} - \tilde{\alpha}_{j0} - \hat{a}_j)' \Sigma_j^{-1} (\tilde{f}_{j,t} - \tilde{\alpha}_{j0} - \hat{a}_j) + \frac{\tilde{T} k_j^{-1}}{k_j^{-1} + \tilde{T}} \hat{a}_j' \Sigma^{-1} \hat{a}_j \\
= &\mathrm{tr}\left(\Psi_j \Sigma_j^{-1}\right)
\end{aligned}
$$

where

$$
\Psi_j = \sum_{t=n_t+1}^{T} (\tilde{f}_{j,t} - \hat{\tilde{\alpha}}_j)' (\tilde{f}_{j,t} - \hat{\tilde{\alpha}}_j) + \frac{\tilde{T}}{\tilde{T} k_j + 1} (\hat{\tilde{\alpha}}_j - \tilde{\alpha}_{j0})' (\hat{\tilde{\alpha}}_j - \tilde{\alpha}_{j0}).
$$

Combining with $\tilde{\psi}(\Sigma_j|\mathscr{M}_j)$, we achieve

$$
\begin{aligned}
\pi(\Sigma_j|\mathscr{M}_j, \tilde{f}_{j,n_t+1:T}) &\propto p(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j, \Sigma_j) \tilde{\psi}(\Sigma_j|\mathscr{M}_j) \\
&\propto |\Sigma|^{-\frac{\tilde{T} + L_j - K + L_j + 1}{2}} e^{-\frac{1}{2}\mathrm{tr}\left(\Psi_j \Sigma_j^{-1}\right)},
\end{aligned}
$$

which follows an inverse Wishart distribution with degree of freedom $\nu = \tilde{T} + L_j - K$ and the scale matrix is Ψ_j. So the fifth term is given by

$$
\pi(\Sigma_j|\mathscr{M}_j, \tilde{f}_{j,n_t+1:T}) = \mathscr{I}\mathscr{W}_{L_j}(\Sigma_j|\Psi_j, \nu). \tag{C17}
$$

Plugging equations (I.6) - (I.11) back into equation (I.5), after cancellation, we get

$$
\begin{aligned}
\log \tilde{m}(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j) = &\log(2\pi)^{-\frac{\tilde{T} L_j}{2}} + \log|\tilde{T} k_j + 1|^{\frac{\tilde{T}}{2}} - \log|\Psi_j|^{\frac{\nu}{2}} + \log 2^{\frac{\nu L_j}{2}} + \log \Gamma_{L_j}\left(\frac{\nu}{2}\right) \\
= &-\frac{(K - L_j) L_j}{2} \log 2 - \frac{\tilde{T} L_j}{2} \log \pi - \frac{L_j}{2} \log(\tilde{T} k_j + 1) \\
&- \frac{\left(\tilde{T} + L_j - K\right)}{2} \log|\Psi_j| + \log \Gamma_{L_j}\left(\frac{\tilde{T} + L_j - K}{2}\right).
\end{aligned}
$$

To calculate the marginal likelihood on the log-scale for the bottom model, we apply the equation (C7) on the log-scale

$$\begin{aligned}\log \tilde{m}(f^*_{j,n_t+1:T}|\mathscr{M}_j) = &\log p(f^*_{j,n_t+1:T}|\mathscr{M}_j, \beta^*_{j,f}, \Sigma^*_j) \\ &+ \log \pi(\beta^*_{j,f}|\mathscr{M}_j, \Sigma^*_j) + \log \tilde{\psi}(\Sigma^*_j|\mathscr{M}_j) \\ &- \log \pi(\beta^*_{j,f}|\mathscr{M}_j, \Sigma^*_j, f^*_{j,n_t+1:T}) - \log \pi(\Sigma^*_j|\mathscr{M}_j, f^*_{j,n_t+1:T}),\end{aligned} \tag{C18}$$

where the first three terms are sampling density and the prior density of parameters:

$$p(f^*_{j,n_t+1:T}|\mathscr{M}_j, \beta^*_{j,f}, \Sigma^*_j) = \prod_{t=n_t+1}^{T} \mathscr{N}_{L_j}(f^*_{j,t}|B^*_{j,f}\tilde{f}_{j,t}, \Sigma^*_j) \tag{C19}$$

$$\pi(\beta^*_{j,f}|\mathscr{M}_j, \Sigma^*_j) = 1 \tag{C20}$$

$$\tilde{\psi}(\Sigma^*_j|\mathscr{M}_j) = |\Sigma^*_j|^{-\frac{-K+1}{2}}. \tag{C21}$$

The fourth term can be obtained by applying the results of multivariate normal regression for $t = n_t+1, \ldots, T$, thus

$$\pi(\beta^*_{j,f}|\mathscr{M}_j, \Sigma^*_j, f^*_{j,n_t+1:T}) = \mathscr{N}_{K-L_j}(\beta^*_{j,f}|\hat{\beta}^*_{j,f}, D^{*-1}_{j1}), \tag{C22}$$

where

$$D^*_{j1} = W^*_j \otimes \Sigma^{*-1}_j \quad \text{and} \quad W^*_j = \sum_{t=n_t+1}^{T} \tilde{f}_{j,t}\tilde{f}'_{j,t} \tag{C23}$$

To obtain the fifth term, we calculate

$$\begin{aligned}&p(f^*_{j,n_t+1:T}|\mathscr{M}_j, \Sigma^*_j) \\ =&\int (2\pi)^{-\frac{\tilde{T}(K-L_j)}{2}} |\Sigma^*_j|^{-\frac{\tilde{T}}{2}} e^{-\frac{1}{2}SSE^*} e^{-\frac{1}{2}(\beta_{j,f}-\hat{\beta}_{j,f})'(W^*_j\otimes\Sigma^{*-1}_j)(\beta^*_{j,f}-\hat{\beta}^*_{j,f})} \mathrm{d}\beta^*_{j,f} \\ =&(2\pi)^{-\frac{(\tilde{T}-L_j)(K-L_j)}{2}} |\Sigma^*_j|^{-\frac{\tilde{T}-L_j}{2}} |W^*_j|^{-\frac{K-L_j}{2}} e^{-\frac{1}{2}SSE^*} \\ &\int (2\pi)^{-\frac{L_j(K-L_j)}{2}} |W^{*-1}_j \otimes \Sigma^*_j|^{-\frac{1}{2}} e^{-\frac{1}{2}(a_j-\hat{a}_j)'(W^*_j\otimes\Sigma^{*-1}_j)(a_j-\hat{a}_j)} \mathrm{d}a_j\end{aligned}$$

where the integral is equal to one and

$$
\begin{aligned}
SSE^* =& \sum_{t=n_t+1}^{T} (f^*_{j,t} - \hat{B}^*_{j,f}\tilde{f}_{j,t})\Sigma^{*-1}(f^*_{j,t} - \hat{B}^*_{j,f}\tilde{f}_{j,t})' \\
=& \mathrm{tr}\left(\Psi^*_j \Sigma^{*-1}_j\right)
\end{aligned}
$$

where

$$
\Psi^*_j = \sum_{t=n_t+1}^{T} (f^*_{j,t} - \hat{B}^*_{j,f}\tilde{f}_{j,t})(f^*_{j,t} - \hat{B}^*_{j,f}\tilde{f}_{j,t})'.
$$

Combining with $\tilde{\psi}(\Sigma_j|\mathscr{M}_j)$, we achieve

$$
\begin{aligned}
\pi(\Sigma_j|\mathscr{M}_j, \tilde{f}_{j,n_t+1:T}) &\propto p(\tilde{f}_{j,n_t+1:T}|\mathscr{M}_j, \Sigma_j)\tilde{\psi}(\Sigma_j|\mathscr{M}_j) \\
&\propto |\Sigma|^{-\frac{\tilde{T}-L_j+K+1}{2}} e^{-\frac{1}{2}\mathrm{tr}\left(\Psi^*_j \Sigma^{*-1}_j\right)},
\end{aligned}
$$

which follows an inverse Wishart distribution with degree of freedom $\nu^* = \tilde{T}$ and the scale matrix is Ψ^*_j. So the fifth term is given by

$$
\pi(\Sigma^*_j|\mathscr{M}_j, f^*_{j,n_t+1:T}) = \mathscr{I}\mathscr{W}_{K-L_j}(\hat{\Sigma}^*_j|\Psi^*_j, \nu^*). \tag{C24}
$$

Plugging equations (I.13) - (I.18) back into equation (I.12), after cancellation, we get

$$
\begin{aligned}
&\frac{(K-L_j)L_j}{2}\log 2 - \frac{(K-L_j)(\tilde{T}-L_j)}{2}\log\pi \\
&-\frac{(K-L_j)}{2}\log|W^*_j| - \frac{\tilde{T}}{2}\log|\Psi^*_j| + \log\Gamma_{K-L_j}\left(\frac{\tilde{T}}{2}\right).
\end{aligned}
$$