The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXV, NO. 1 • FEBRUARY 2020

On Comparing Asset Pricing Models

SIDDHARTHA CHIB, XIAMING ZENG, and LINGXIAO ZHAO*

ABSTRACT

Revisiting the framework of (Barillas, Francisco, and Jay Shanken, 2018, Comparing asset pricing models, *The Journal of Finance* 73, 715–754). 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 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 $\{\boldsymbol{\eta}_j\}$ 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.

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

*Siddhartha Chib 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 have read *The Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

DOI: 10.1111/jofi.12854

(2015), and the Carhart (1997) momentum (MOM) factor. In the second, we match 12 potential risk factors to the eight factors above as well as the Asness, Frazzini, and Pedersen (2014) quality minus junk (QMJ) factor, the Pástor 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 and Frazzini (2013). 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. 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.

I. 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 $\mathcal{M}_j$, $j = 1, 2, \ldots, J$, represents any one of the possible models. It is defined by the vector of risk factors {Mkt, $\boldsymbol{f}_j$} of size L_j and the vector of nonrisk factors $\boldsymbol{f}_j^*$ of size $(K - L_j)$. Note that $\boldsymbol{f}$ is indexed by j because what goes into $\boldsymbol{f}$ is what varies across models. Then, letting t denote a particular point in the sample, $t = 1, 2, \ldots, n$, each model in the model-space is given by

$$\boldsymbol{f}_{j,t} = \boldsymbol{\alpha}_j + \boldsymbol{\beta}_j \,\text{Mkt}_t + \boldsymbol{\varepsilon}_{j,t}, \quad \boldsymbol{\varepsilon}_{j,t} \sim \mathcal{N}_{L_j-1}(\mathbf{0}, \Sigma_j)$$

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

where an intercept vector is absent from the $\boldsymbol{f}^*_{j,t}$ model because of the pricing restrictions and the error terms $\boldsymbol{\varepsilon}_{j,t}$ and $\boldsymbol{\varepsilon}^*_{j,t}$ 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 $\boldsymbol{\beta}^*_{j,f} = \text{vec}(\boldsymbol{B}^*_{j,f})$ denote the column-vectorized form of $\boldsymbol{B}^*_{j,f}$, and $\boldsymbol{\sigma}_j = \text{vech}(\Sigma_j)$ and $\boldsymbol{\sigma}^*_j = \text{vech}(\Sigma^*_j)$ the half or unique element vectorizations of the two covariance matrices. Then the parameters of $\mathcal{M}_j$ are

$$\boldsymbol{\theta}_j = \left(\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j, \boldsymbol{\beta}^*_{j,m}, \boldsymbol{\beta}^*_{j,f}, \boldsymbol{\sigma}_j, \boldsymbol{\sigma}^*_j\right) \in \Theta_{\boldsymbol{\theta}_j},$$

of which

$$\boldsymbol{\eta}_j = \left(\boldsymbol{\beta}_j, \boldsymbol{\beta}^*_{j,m}, \boldsymbol{\beta}^*_{j,f}, \boldsymbol{\sigma}_j, \boldsymbol{\sigma}^*_j\right)$$

are the nuisance parameters of $\mathcal{M}_j$. We let $\Theta_{\boldsymbol{\theta}_j}$ and $\Theta_{\boldsymbol{\eta}_j}$ denote the parameter spaces of $\boldsymbol{\theta}_j$ and $\boldsymbol{\eta}_j$, respectively, these being obvious by context.

BS suppose that the prior density of $\boldsymbol{\theta}_j$ is given by

$$p_{BS}(\boldsymbol{\theta}_j|\mathcal{M}_j) = \pi_{BS}(\boldsymbol{\alpha}_j|\mathcal{M}_j, \boldsymbol{\eta}_j)\psi_{BS}(\boldsymbol{\eta}_j|\mathcal{M}_j), \tag{1}$$

where

$$\pi_{BS}(\boldsymbol{\alpha}_j|\mathcal{M}_j, \boldsymbol{\eta}_j) = \mathcal{N}_{L_j-1}(\boldsymbol{\alpha}_j|\mathbf{0}, k\Sigma_j) \tag{2}$$

$$\psi_{BS}(\boldsymbol{\eta}_j|\mathcal{M}_j) = |\Sigma_j|^{-L_j/2}|\Sigma^*_j|^{-(K-L_j+1)/2} \tag{3}$$

and $k > 0$ controls the spread of the prior on $\boldsymbol{\alpha}_j$. Thus, in this prior $\pi_{BS}(\boldsymbol{\alpha}_j|\mathcal{M}_j, \boldsymbol{\eta}_j)$ is a proper density and $\psi_{BS}(\boldsymbol{\eta}_j|\mathcal{M}_j)$ 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

$$\boldsymbol{y}_{1:T} = \left(\boldsymbol{f}_1, \boldsymbol{f}_1^*, \ldots, \boldsymbol{f}_T, \boldsymbol{f}_T^*\right)$$

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

$$m(\boldsymbol{y}_{1:T}|\mathcal{M}_j) \triangleq \int_{\Theta_{\eta_j}} \int_{\Theta_{\alpha_j}} p(\boldsymbol{y}_{1:T}|\mathcal{M}_j, \boldsymbol{\theta}_j)\pi(\boldsymbol{\alpha}_j|\mathcal{M}_j, \boldsymbol{\eta}_j)\psi(\boldsymbol{\eta}_j|\mathcal{M}_j)\mathrm{d}\boldsymbol{\theta}_j, \qquad (4)$$

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

$$m(\boldsymbol{y}_{1:T}|\mathcal{M}_j) = m(\boldsymbol{f}_{1:T}|\mathcal{M}_j)m\left(\boldsymbol{f}_{1:T}^*|\mathcal{M}_j\right),$$

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

II. 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 $\psi_{BS}(\boldsymbol{\eta}_j|\mathcal{M}_j)$ is an improper distribution, $c_j\psi_{BS}(\boldsymbol{\eta}_j|\mathcal{M}_j)$ 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,

$$c_j = 1, \quad j = 1, 2, \ldots, J,$$

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 $\boldsymbol{\eta}_j = (\boldsymbol{\beta}_j, \boldsymbol{\beta}_{j,m}^*, \boldsymbol{\beta}_{j,f}^*, \boldsymbol{\sigma}_j, \boldsymbol{\sigma}_j^*)$ do not vary by model, and that their parameter spaces Θ_η are also common across models. In that case,

$$m(\boldsymbol{y}_{1:T}|\mathcal{M}_j) = \int_{\Theta_\eta} \int_{\Theta_{\alpha_j}} p(\boldsymbol{y}_{1:T}|\mathcal{M}_j, \boldsymbol{\theta}_j)\pi(\boldsymbol{\alpha}_j|\mathcal{M}_j, \boldsymbol{\eta})c\psi(\boldsymbol{\eta})\mathrm{d}\boldsymbol{\theta}_j. \qquad (5)$$

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*

$$\left\{ m(\boldsymbol{y}_{1:T}|\mathcal{M}_1), \ldots, m(\boldsymbol{y}_{1:T}|\mathcal{M}_J) \right\}$$

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 $\boldsymbol{\eta}_j$ 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 $\{\boldsymbol{\eta}_j\}$ 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 $\boldsymbol{\eta}_1$ of a model that we call $\mathcal{M}_1$, and the nuisance parameters $\boldsymbol{\eta}_j$ of a generic model that we refer to as $\mathcal{M}_j$.
- Next we give the nuisance parameters $\boldsymbol{\eta}_1$ of $\mathcal{M}_1$ a Jeffreys prior.
- Then, for every other model $j > 1$, we derive the prior on $\boldsymbol{\eta}_j$ 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 $\mathcal{M}_1, \ldots, \mathcal{M}_J$. Suppose that the nuisance parameters $\boldsymbol{\eta}_1$ of model $\mathcal{M}_1$ are connected to the nuisance parameters $\boldsymbol{\eta}_j$ of $\mathcal{M}_j$ ($j > 1$) by the invertible mapping $\boldsymbol{\eta}_j = g_j(\boldsymbol{\eta}_1)$, with the inverse mapping given by*

$$\boldsymbol{\eta}_1 = g_j^{-1}(\boldsymbol{\eta}_j). \tag{6}$$

Let

$$c\psi(\boldsymbol{\eta}_1|\mathcal{M}_1)$$

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

$$\tilde{\psi}(\boldsymbol{\eta}_j|\mathcal{M}_j) = c\psi\left(g_j^{-1}(\boldsymbol{\eta}_j)|\mathcal{M}_1\right)\left|\det\left(\frac{\partial g_j^{-1}(\boldsymbol{\eta}_j)}{\partial \boldsymbol{\eta}_j'}\right)\right|, \quad j = 2, 3, \ldots, J \tag{7}$$

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

$$m(\boldsymbol{y}_{1:T}|\mathcal{M}_1) = \int_{\Theta_{\eta_1}}\int_{\Theta_{\alpha_1}} p(\boldsymbol{y}_{1:T}|\mathcal{M}_1, \boldsymbol{\theta}_1)\pi(\boldsymbol{\alpha}_1|\mathcal{M}_1, \boldsymbol{\eta}_1)c\psi(\boldsymbol{\eta}_1|\mathcal{M}_1)\mathrm{d}\boldsymbol{\theta}_1$$

and

$$\tilde{m}(\boldsymbol{y}_{1:T}|\mathcal{M}_j) = \int_{\Theta_{\eta_j}}\int_{\Theta_{\alpha_j}} p(\boldsymbol{y}_{1:T}|\mathcal{M}_j, \boldsymbol{\theta}_j)\pi(\boldsymbol{\alpha}_j|\mathcal{M}_j, \boldsymbol{\eta}_j)\tilde{\psi}(\boldsymbol{\eta}_j|\mathcal{M}_j)\mathrm{d}\boldsymbol{\theta}_j \tag{8}$$

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

$$\left\{m(\boldsymbol{y}_{1:T}|\mathcal{M}_1),\ \tilde{m}(\boldsymbol{y}_{1:T}|\mathcal{M}_2), \ldots,\ \tilde{m}(\boldsymbol{y}_{1:T}|\mathcal{M}_J)\right\}$$

are valid and comparable.

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

$$\tilde{m}(\boldsymbol{y}_{1:T}|\mathcal{M}_j) = \int_{\Theta_{\eta_j}}\int_{\Theta_{\alpha_j}} p(\boldsymbol{y}_{1:T}|\mathcal{M}_j, \boldsymbol{\theta}_j)\pi(\boldsymbol{\alpha}_j|\boldsymbol{\eta}_j)c\psi\left(g_j^{-1}(\boldsymbol{\eta}_j)|\mathcal{M}_1\right)\left|\det\left(\frac{\partial g_j^{-1}(\boldsymbol{\eta}_j)}{\partial \boldsymbol{\eta}_j'}\right)\right|\mathrm{d}(\boldsymbol{\alpha}_j, \boldsymbol{\eta}_j). \tag{9}$$

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.

III. 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 $\boldsymbol{\eta}_1$ of a model we call $\mathcal{M}_1$ and the nuisance parameters $\boldsymbol{\eta}_j$ of a generic model that we refer to as $\mathcal{M}_j$. We then derive the Jacobian of the transformation, followed by the prior density of $\boldsymbol{\eta}_j$ 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.

A. Derivation of the CZZ Priors

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

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

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

$$\boldsymbol{f}_{1,t} = \boldsymbol{\alpha}_1 + \boldsymbol{\beta}_1 \,\mathrm{Mkt}_t + \boldsymbol{\varepsilon}_{1,t}, \quad \boldsymbol{\varepsilon}_{1,t} \sim \mathcal{N}_{K-1}(\mathbf{0}, \Sigma_1) \tag{10}$$

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

$$\boldsymbol{\eta}_1 = (\boldsymbol{\beta}_1, \boldsymbol{\sigma}_1).$$

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

$$\boldsymbol{f}_{j,t} = \boldsymbol{\alpha}_j + \boldsymbol{\beta}_j \,\mathrm{Mkt}_t + \boldsymbol{\varepsilon}_{j,t}, \quad \boldsymbol{\varepsilon}_{j,t} \sim \mathcal{N}_{L_j-1}(\mathbf{0}, \Sigma_j) \tag{11}$$

$$\boldsymbol{f}^*_{j,t} = \left(\boldsymbol{\beta}^*_{j,m}\ \boldsymbol{B}^*_{j,f}\right) \begin{pmatrix} \mathrm{Mkt}_t \\ \boldsymbol{f}_{j,t} \end{pmatrix} + \boldsymbol{\varepsilon}^*_{j,t}, \quad \boldsymbol{\varepsilon}^*_{j,t} \sim \mathcal{N}_{K-L_j}\left(\mathbf{0}, \Sigma^*_j\right) \tag{12}$$

with nuisance parameters given by

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

Plugging the model in (11) into (12), we get

$$\boldsymbol{f}^*_{j,t} = \boldsymbol{B}^*_{j,f}\boldsymbol{\alpha}_j + (\boldsymbol{\beta}^*_{j,m} + \boldsymbol{B}^*_{j,f}\boldsymbol{\beta}_j)\text{Mkt}_t + (\boldsymbol{\varepsilon}^*_{j,t} + \boldsymbol{B}^*_{j,f}\boldsymbol{\varepsilon}_{j,t}). \tag{13}$$

Vectorizing equations (11) and (13), we have that

$$\begin{pmatrix} \boldsymbol{f}_{j,t} \\ \boldsymbol{f}^*_{j,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_j \\ \boldsymbol{B}^*_{j,f}\boldsymbol{\alpha}_j \end{pmatrix} + \begin{pmatrix} \boldsymbol{\beta}_j \\ \boldsymbol{\beta}^*_{j,m} + \boldsymbol{B}^*_{j,f}\boldsymbol{\beta}_j \end{pmatrix} \text{Mkt}_t + \boldsymbol{\varepsilon}_{j,t}, \tag{14}$$

where

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

Comparing the parameters of equations (10) and (14), we see that the nuisance parameters of $\mathcal{M}_1$ and $\mathcal{M}_j$, $j = 2, 3, \ldots, J-1$, are related as follows:

$$\boldsymbol{\beta}_1 = \begin{pmatrix} \boldsymbol{\beta}_j \\ \boldsymbol{\beta}^*_{j,m} + \boldsymbol{B}^*_{j,f}\boldsymbol{\beta}_j \end{pmatrix}, \tag{15}$$

$$\Sigma_1 = \begin{pmatrix} \Sigma_j & \Sigma_j \boldsymbol{B}^{*\prime}_{j,f} \\ \boldsymbol{B}^*_{j,f}\Sigma_j & \Sigma^*_j + \boldsymbol{B}^*_{j,f}\Sigma_j\boldsymbol{B}^{*\prime}_{j,f} \end{pmatrix}, \tag{16}$$

or in vech form,

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} \boldsymbol{\sigma}_j \\ (\Sigma_j \otimes \boldsymbol{I}_{K-L_j})\boldsymbol{\beta}^*_{j,f} \\ \boldsymbol{\sigma}^*_j + \text{vech}\left(\boldsymbol{B}^*_{j,f}\Sigma_j\boldsymbol{B}^{*\prime}_{j,f}\right) \end{pmatrix}. \tag{17}$$

The set of vector equations in (15) and (17) constitute the inverse map $\boldsymbol{\eta}_1 = g_j^{-1}(\boldsymbol{\eta}_j)$. The determinant of the Jacobian of this transformation can now be derived. By derivations given in Appendix B, we have that

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

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

$$c\psi(\boldsymbol{\eta}_1|\mathcal{M}_1) = c|\Sigma_1|^{-\frac{K}{2}}.$$

Then, by the rule for the determinant of a partitioned matrix applied to (16), the prior of $\boldsymbol{\eta}_j$ in model $\mathcal{M}_j$, $j = 2, 3, \ldots, J-1$, is

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

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

$$\boldsymbol{f}^*_{J,t} = \boldsymbol{\beta}^*_{J,m}\text{Mkt}_t + \boldsymbol{\varepsilon}^*_{J,t}, \quad \boldsymbol{\varepsilon}^*_{J,t} \sim \mathcal{N}_{K-1}\left(\mathbf{0}, \Sigma^*_J\right). \tag{18}$$

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

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

We have thus proved the following new result.

PROPOSITION 3: *Let the first model $\mathcal{M}_1$ in equation (10) have the improper prior on $\boldsymbol{\eta}_1$ given by*

$$c\psi(\boldsymbol{\eta}_1|\mathcal{M}_1) = c|\Sigma_1|^{-\frac{K}{2}},$$

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

$$\tilde{\psi}(\boldsymbol{\eta}_j|\mathcal{M}_j) = c|\Sigma_j|^{-\frac{2L_j-K}{2}}|\Sigma^*_j|^{-\frac{K}{2}}$$

and that of $\boldsymbol{\eta}_J$ in $\mathcal{M}_J$ given by

$$\tilde{\psi}(\boldsymbol{\eta}_J|\mathcal{M}_J) = c|\Sigma^*_J|^{-\frac{K}{2}}$$

satisfy Proposition 2.

The simplicity of this result should be noted.

B. CZZ Marginal Likelihoods

The valid marginal likelihoods for models $\mathcal{M}_1, \ldots, \mathcal{M}_J$ can now be derived. We assume that the prior of $\boldsymbol{\alpha}_j|\mathcal{M}_j, \boldsymbol{\eta}_j$ is the same as in (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.[1]

Consider the typical model $\mathcal{M}_j$ $(j \neq 1, J)$. 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:

$$\log \tilde{m}(\boldsymbol{y}_{1:T}|\mathcal{M}_j) = \log \tilde{m}(\boldsymbol{f}_{1:T}|\mathcal{M}_j) + \log \tilde{m}(\boldsymbol{f}^*_{1:T}|\mathcal{M}_j), \tag{19}$$

[1] Details are provided in the Internet Appendix which may be found in the online version of this article.

where the first term on the RHS is

$$-\frac{(K-L_j)(L_j-1)}{2}\log 2-\frac{(T-1)(L_j-1)}{2}\log\pi-\frac{(L_j-1)}{2}\log k$$
$$-\frac{(L_j-1)}{2}\log|W|-\frac{(T+L_j-K-1)}{2}\log|\Psi_j|+\log\Gamma_{L_j-1}\left(\frac{T+L_j-K-1}{2}\right), \tag{20}$$

the second term is

$$\frac{(K-L_j)(L_j-1)}{2}\log 2-\frac{(K-L_j)(T-L_j)}{2}\log\pi$$
$$-\frac{(K-L_j)}{2}\log|W_j^*|-\frac{(T-1)}{2}\log|\Psi_j^*|+\log\Gamma_{K-L_j}\left(\frac{T-1}{2}\right), \tag{21}$$

and

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

In these expressions, the hat symbol denotes the least square estimates, and $\Gamma_d(\cdot)$ denotes the d-dimensional multivariate gamma function. Finally, for $\mathcal{M}_1$ the log marginal likelihood is given by (20), and for $\mathcal{M}_J$ it is given by (21).

The computations typically take a few seconds to scan our model-space of 2,048 models in the 12-factor case.

C. 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 (2019) 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 (2019).

Let $\tilde{\boldsymbol{f}}$ denote the set of risk factors, and let $\boldsymbol{f}^*$ denote the set of nonrisk factors. The model that we describe here differs from those above because Mkt can now enter into $\tilde{\boldsymbol{f}}$ or $\boldsymbol{f}^*$. Also note that $\tilde{\boldsymbol{f}}$ 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 $\mathcal{M}_1$ all K factors are risk factors,

$$\tilde{\boldsymbol{f}}_{1,t} = \tilde{\boldsymbol{\alpha}}_1 + \tilde{\boldsymbol{\varepsilon}}_{1,t}, \quad \tilde{\boldsymbol{\varepsilon}}_{1,t} \sim \mathcal{N}_K(\mathbf{0}, \Sigma_1). \tag{22}$$

Let $\boldsymbol{\sigma}_1 = \text{vech}(\Sigma_1)$. Then, the nuisance parameters of $\mathcal{M}_1$ are simply

$$\boldsymbol{\eta}_1 = \boldsymbol{\sigma}_1.$$

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

$$\tilde{\boldsymbol{f}}_{j,t} = \tilde{\boldsymbol{\alpha}}_j + \tilde{\boldsymbol{\varepsilon}}_{j,t}, \quad \tilde{\boldsymbol{\varepsilon}}_{j,t} \sim \mathcal{N}_{L_j}\big(\mathbf{0}, \Sigma_j\big), \tag{23}$$

$$\boldsymbol{f}_{j,t}^* = \boldsymbol{B}_{j,f}^* \tilde{\boldsymbol{f}}_{j,t} + \boldsymbol{\varepsilon}_{j,t}^*, \quad \boldsymbol{\varepsilon}_{j,t}^* \sim \mathcal{N}_{K-L_j}\big(\mathbf{0}, \Sigma_j^*\big) \tag{24}$$

with nuisance parameters

$$\boldsymbol{\eta}_j = \big(\boldsymbol{\beta}_{j,f}^*, \boldsymbol{\sigma}_j, \boldsymbol{\sigma}_j^*\big),$$

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

PROPOSITION 4: *Let model $\mathcal{M}_1$ in equation (22) have the improper prior on $\boldsymbol{\eta}_1$ given by*

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

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

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

satisfy Proposition 2.

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

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

where the prior mean $\tilde{\boldsymbol{\alpha}}_{j0}$ 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

$$\tilde{\boldsymbol{\alpha}}_{j0} = n_t^{-1} \sum_{t=1}^{n_t} \tilde{\boldsymbol{f}}_{j,t}, \tag{25}$$

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 $\tilde{\boldsymbol{\alpha}}_j$, from the log of the marginal likelihood of the training sample observations $\tilde{\boldsymbol{f}}_{1:n_t}$ (conditioned on $\tilde{\boldsymbol{\alpha}}_j$ but marginalized over Σ_j):

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

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

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

The Hessian matrix (an $L_j \times L_j$ 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 $\hat{\Sigma}_{j0}$,

$$k_j = mult \times L_j^{-1}\mathrm{sum}\left(\mathrm{diag}\left(V_{j0}\right)/\mathrm{diag}\left(\hat{\Sigma}_{j0}\right)\right), \quad j = 1, 2, \ldots, \tilde{J}, \tag{26}$$

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 $\tilde{\boldsymbol{\alpha}}_j|\mathcal{M}_j \sim \mathcal{N}_{L_j}(\tilde{\boldsymbol{\alpha}}_{j0}, k_j\Sigma_j)$, the marginal likelihood of model $\mathcal{M}_j$, $j = 2, 3, \ldots, \tilde{J}$, on the log-scale is given by*

$$\log \tilde{m}(\boldsymbol{y}_{n_t+1:T}|\mathcal{M}_j) = \log \tilde{m}\left(\tilde{\boldsymbol{f}}_{n_t+1:T}|\mathcal{M}_j\right) + \log \tilde{m}\left(\boldsymbol{f}^*_{n_t+1:T}|\mathcal{M}_j\right), \tag{27}$$

where the first term on the RHS is

$$-\frac{(K-L_j)L_j}{2}\log 2 - \frac{\tilde{T}L_j}{2}\log\pi - \frac{L_j}{2}\log(\tilde{T}k_j + 1)$$
$$-\frac{\left(\tilde{T} + L_j - K\right)}{2}\log|\Psi_j| + \log\Gamma_{L_j}\left(\frac{\tilde{T} + L_j - K}{2}\right), \tag{28}$$

the second term is

$$\frac{(K-L_j)L_j}{2}\log 2 - \frac{(K-L_j)(\tilde{T}-L_j)}{2}\log\pi$$
$$-\frac{(K-L_j)}{2}\log\left|W_j^*\right| - \frac{\tilde{T}}{2}\log\left|\Psi_j^*\right| + \log\Gamma_{K-L_j}\left(\frac{\tilde{T}}{2}\right), \tag{29}$$

and the log marginal likelihood of $\mathcal{M}_1$ defined in (22) is given by (28). In these expressions, $\tilde{T} = (T - n_t)$ and

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

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

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 $\tilde{\boldsymbol{\alpha}}_j|\mathcal{M}_j \sim \mathcal{N}_{L_j}(\tilde{\boldsymbol{\alpha}}_{j0}, k_j\Sigma_j)$ produces performance gains of up to 20%, for smaller sample sizes, compared to the marginal likelihoods of Proposition 5 based on $\tilde{\boldsymbol{\alpha}}_j|\mathcal{M}_j \sim \mathcal{N}_{L_j}(\mathbf{0}, k\Sigma_j)$. 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 (25) and (26).

IV. 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, and, 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.

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

$$\underbrace{\begin{pmatrix}\text{HML}_t \\ \text{SMB}_t\end{pmatrix}}_{\boldsymbol{f}^*_t:2\times 1} = \underbrace{\begin{pmatrix}\alpha_{1,h} \\ \alpha_{1,s}\end{pmatrix}}_{\boldsymbol{\alpha}^*_1:2\times 1} + \underbrace{\begin{pmatrix}\beta_{1,hm} \\ \beta_{1,sm}\end{pmatrix}}_{\boldsymbol{\beta}^*_1:2\times 1}\text{Mkt}_t + \boldsymbol{\varepsilon}^*_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \underbrace{\begin{pmatrix}\sigma^2_{1,h} & \sigma_{1,hs} \\ \sigma_{1,hs} & \sigma^2_{1,s}\end{pmatrix}}_{\Sigma_1:2\times 2}\right). \tag{30}$$

In this case, the nuisance parameters are

$$\boldsymbol{\eta}_1 = \left(\beta_{1,hm}, \beta_{1,sm}, \sigma^2_{1,h}, \sigma_{1,hs}, \sigma^2_{1,s}\right), \tag{31}$$

which is of size five. From (3), we have that

$$\psi_{BS}(\boldsymbol{\eta}_1|\mathcal{M}_1) = \frac{1}{\left(\sigma^2_{1,h}\sigma^2_{1,s} - \sigma^2_{1,hs}\right)^{3/2}}. \tag{32}$$

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

$$\underbrace{\text{HML}_t}_{\boldsymbol{f}_t:1\times 1} = \underbrace{\alpha_{2,h}}_{\boldsymbol{\alpha}_2:1\times 1} + \underbrace{\beta_{2,hm}}_{\boldsymbol{\beta}_2:1\times 1}\text{Mkt}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(\mathbf{0}, \underbrace{\sigma^2_{2,h}}_{\Sigma_2:1\times 1}\right), \tag{33}$$

$$\underbrace{\text{SMB}_t}_{\boldsymbol{f}^*_t:1\times 1} = \underbrace{\left(\beta^*_{2,sm}\beta^*_{2,sh}\right)}_{\boldsymbol{\beta}^*_2:1\times 2}\begin{pmatrix}\text{Mkt}_t\\ \text{HML}_t\end{pmatrix} + \varepsilon^*_t, \quad \varepsilon^*_t \sim \mathcal{N}\left(\mathbf{0}, \underbrace{\sigma^{*2}_{2,s}}_{\Sigma^*_2:1\times 1}\right), \tag{34}$$

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

$$\boldsymbol{\eta}_2 = \left(\beta_{2,hm}, \beta^*_{2,sm}, \beta^*_{2,sh}, \sigma^2_{2,h}, \sigma^{*2}_{2,s}\right). \tag{35}$$

The prior density of $\boldsymbol{\eta}_2$ from (3) is

$$\psi_{BS}(\boldsymbol{\eta}_2|\mathcal{M}_2) = \frac{1}{\sigma^2_{2,h}}\frac{1}{\sigma^{*2}_{2,s}}. \tag{36}$$

The inverse mapping $g_2^{-1}(\cdot)$ in (6) can be derived by substituting the model of HML into that of SMB in model $\mathcal{M}_2$ and comparing terms with those in model $\mathcal{M}_1$. By elementary algebra, we get

$$\beta_{1,hm} = \beta_{2,hm} \tag{37}$$

$$\begin{aligned} \beta_{1,sm} &= \beta^*_{2,sm} + \beta^*_{2,sh}\beta_{2,hm} \\ \sigma^2_{1,h} &= \sigma^2_{2,h} \\ \sigma_{1,hs} &= \beta^*_{2,sh}\sigma^2_{2,h} \\ \sigma^2_{1,s} &= \beta^{*2}_{2,sh}\sigma^2_{2,h} + \sigma^{*2}_{2,s}, \end{aligned} \tag{38}$$

where

$$\left|\det\left(\frac{\partial g_j^{-1}(\boldsymbol{\eta}_j)}{\partial \boldsymbol{\eta}_2'}\right)\right| = \sigma^2_{2,h}. \tag{39}$$

We can now easily check that the BS prior of $\boldsymbol{\eta}_2$ in $\mathcal{M}_2$, given in (36), is not equal to the required prior in (7):

$$\psi_{BS}(\boldsymbol{\eta}_2|\mathcal{M}_2) = \frac{1}{\sigma_{2,h}^2}\frac{1}{\sigma_{2,s}^{*2}} \tag{40}$$

$$\neq \tilde{\psi}(\boldsymbol{\eta}_2|\mathcal{M}_2) = \psi_{BS}(g_j^{-1}(\boldsymbol{\eta}_j)|\mathcal{M}_1)\sigma_{2,h}^2 \tag{41}$$

$$= \frac{1}{\left(\sigma_{2,h}^2\sigma_{2,s}^{*2}\right)^{\frac{3}{2}}}\sigma_{2,h}^2 = \frac{1}{\sigma_{2,h}\sigma_{2,s}^{*3}}. \tag{42}$$

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

B. 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 $\mathcal{P}_1 = \{\text{Mkt, HML, SMB}\}$, and in the second they are ordered as $\mathcal{P}_2 = \{\text{Mkt, SMB, HML}\}$. Under $\mathcal{P}_1$, three nested models can be shown to arise by suitably restricting the parameters of the model

$$\text{HML}_t = a + b\text{Mkt}_t + e, \tag{43}$$

$$\text{SMB}_t = c + d\text{Mkt}_t + g\text{HML}_t + u. \tag{44}$$

For instance, the model $\mathcal{M}_1|\mathcal{P}_1$ (Mkt, HML, SMB are risk factors) arises by setting $a \neq 0$ and $c \neq 0$, $\mathcal{M}_2|\mathcal{P}_1$ (Mkt, HML are risk factors) arises by setting $a \neq 0$ and $c = 0$, and $\mathcal{M}_3|\mathcal{P}_1$ (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 $\mathcal{P}_2$, three nested models can be shown to arise by suitably restricting the parameters of the model

$$\text{SMB}_t = a' + b'\text{Mkt}_t + e', \tag{45}$$

$$\text{HML}_t = c' + d'\text{Mkt}_t + g'\text{SMB}_t + u'. \tag{46}$$

For instance, the model $\mathcal{M}_1|\mathcal{P}_2$ (Mkt, SMB, HML are risk factors) arises by setting $a' \neq 0$ and $c' \neq 0$, $\mathcal{M}_2|\mathcal{P}_2$ (Mkt, SMB are risk factors) arises by setting $a' \neq 0$ and $c = 0$, and $\mathcal{M}_3|\mathcal{P}_2$ (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

Table I

Performance of the BS Method with Eight Potential Risk Factors

This table reports the performance of the BS method in simulation experiments 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, $k = (\text{Sh}^2_{\text{max}} - \text{Sh}(\text{Mkt})^2)/7$, where Sh refers to the Sharpe ratio and $\text{Sh}_{\text{max}} = 3 \times \text{Sh}(\text{Mkt})$.

	Barillas and Shanken (2018)				
Risk Factors in the True Model	$T =$ 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

be used in the improper prior because this situation corresponds to that of Proposition 1. Notice also that $\mathcal{M}_1|\mathcal{P}_1$ and $\mathcal{M}_1|\mathcal{P}_2$ are the same model, as are $\mathcal{M}_3|\mathcal{P}_1$ and $\mathcal{M}_3|\mathcal{P}_2$. Thus, according to Proposition 1, one can replace c_2 with c_1 in calculating the marginal likelihoods of $\mathcal{M}_1|\mathcal{P}_2$ and $\mathcal{M}_3|\mathcal{P}_2$.

Table II

Performance of the BS Method with Twelve Potential Risk Factors

This table reports the performance of the BS method in simulation experiments with twelve potential risk factors. The model-space consists of $J = 2{,}048$ models.

	Barillas and Shanken (2018)				
Risk Factors in the True Model	$T =$ 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

(Continued)

Table II
Continued

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 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

The problem arises in this method in comparing the two distinct models $\mathcal{M}_2|\mathcal{P}_1$ and $\mathcal{M}_2|\mathcal{P}_2$, 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 $\mathcal{M}_2|\mathcal{P}_2$ 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 $\mathcal{M}_2|\mathcal{P}_1$ and $\mathcal{M}_2|\mathcal{P}_2$. 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.

V. 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

Table III

Performance of the CZZ Method with Eight Potential Risk Factors

This table reports the performance of the CZZ method in simulation experiments 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, $k = (\text{Sh}^2_{\text{max}} - \text{Sh}(\text{Mkt})^2)/7$, where Sh refers to the Sharpe ratio and $\text{Sh}_{\text{max}} = 3 \times \text{Sh}(\text{Mkt})$.

	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

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

Table IV

Performance of the CZZ Method with Twelve Potential Risk Factors

This table reports the performance of the CZZ method in simulation experiments 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

(Continued)

Table IV
Continued

	CZZ Method		
Risk Factors in the True Model	$T = 600$	$T = 1{,}200$	$T = 12{,}000$
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

$K = 12$ 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.

A. *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 $\boldsymbol{f}_t$ (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 I, 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 (2) is given by

$$k = \left(\mathrm{Sh}^2_{\max} - \mathrm{Sh}(\mathrm{Mkt})^2\right)/7, \tag{47}$$

where Sh(Mkt) is the Sharpe ratio of the simulated Mkt factor, $\text{Sh}_{\max} = \tau \times$ Sh(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 1,000 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 12 potential risk factors.

B. *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 12 potential risk factors. The overall experiment and implementation are similar to those for the eight-factor experiment.

We match our 12 factors to the eight factors in the first experiment, as well as the Asness, Frazzini, and Pedersen (2014) quality minus junk (QMJ) factor, the Pástor 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 and Frazzini (2013). We now have $2^{11} = 2{,}048$ possible models depending on the assumption made about the collection of factors that go into $\boldsymbol{f}_t$. As in our first experiment, the parameters of the DGP are fixed at the maximum likelihood (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 upon 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 II 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.

VI. 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 III and IV, for the eight-factor and 12-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.

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

Initial submission: May 2, 2018; Accepted: March 27, 2019
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

Appendix A

In this appendix, we describe how the 33 data-generating processes (DGPs) for generating the simulated data in the eight-factor experiment were selected (the same process is used in the 12-factor case, and thus for brevity the details of that case are suppressed). Supposes that

$$\boldsymbol{x}_t = (\text{Mkt},\ \boldsymbol{f}_t)'$$

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

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

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

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

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

$$M_t = 1 - \boldsymbol{\lambda}_x' \Omega_x^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_x),$$

the fitted values of $\boldsymbol{b}' = \boldsymbol{\lambda}_x' \Omega_x^{-1}$ should each be significant. Otherwise, the maintained assumption that the factors {Mkt, SMB, ROE, 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 $\boldsymbol{b}$ 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.[2] 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 $\boldsymbol{b}' = \boldsymbol{\lambda}_x' \Omega_x^{-1}$ is insignificant. If any component is insignificant, the collection of factors $\boldsymbol{x}_t = (\text{Mkt}, \boldsymbol{f}_t)'$ 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 $\boldsymbol{b}$ are obtained as follows. Given a particular $\boldsymbol{x}_t = (\text{Mkt}, \boldsymbol{f}_t)'$ and under the pricing restrictions, any one of candidate factor models takes the form

$$\boldsymbol{x}_t = \Omega_x \boldsymbol{b} + \boldsymbol{\eta}_{x,t}, \tag{A4}$$

$$\boldsymbol{f}_t^* = \boldsymbol{\beta}^* \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t^*, \quad \boldsymbol{\varepsilon}_t^* \sim \mathcal{N}_{K-L}(\mathbf{0}, \Sigma^*), \tag{A5}$$

where

$$\begin{pmatrix} \boldsymbol{\eta}_{x,t} \\ \boldsymbol{\varepsilon}_t^* \end{pmatrix} \sim \mathcal{N}_K\left(\mathbf{0}, \begin{pmatrix} \Omega_x & \mathbf{0} \\ \mathbf{0} & \Sigma^* \end{pmatrix}\right). \tag{A6}$$

Using the data from January 1968 to December 2015, we find the estimate of $\boldsymbol{b}$ 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 $\boldsymbol{b}$ is significant at the 5% level.

[2] 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.

Appendix B

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

PROOF: By definition, the Jacobian is

$$\left|\det\left(\frac{\partial g_j^{-1}(\boldsymbol{\eta}_j)}{\partial \boldsymbol{\eta}_j'}\right)\right| = \left|\det\begin{pmatrix} \frac{\partial \boldsymbol{\beta}_1}{\partial \boldsymbol{\beta}_j'} & \frac{\partial \boldsymbol{\beta}_1}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\beta}_1}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1}{\partial \boldsymbol{\beta}_j'} & \frac{\partial \boldsymbol{\sigma}_1}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\sigma}_1}{\partial \boldsymbol{\sigma}_j^{*\prime}} \end{pmatrix}\right|.$$

Partition $\boldsymbol{\beta}_1$ and $\boldsymbol{\sigma}_1$,

$$\boldsymbol{\beta}_1 = \begin{pmatrix} \boldsymbol{\beta}_1^f : (L_j - 1) \times 1 \\ \boldsymbol{\beta}_1^{f*} : (K - L_j) \times 1 \end{pmatrix},$$

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} \boldsymbol{\sigma}_1^f : \frac{L_j(L_j-1)}{2} \times 1 \\ \boldsymbol{\sigma}_1^{ff*} : (K - L_j)(L_j - 1) \times 1 \\ \boldsymbol{\sigma}_1^{f*} : \frac{(K-L_j)(K-L_j+1)}{2} \times 1 \end{pmatrix},$$

so that the Jacobian can be rewritten as

$$\left|\det\left(\frac{g_j^{-1}(\boldsymbol{\eta}_j)}{\partial \boldsymbol{\eta}_j'}\right)\right| = \left|\det\begin{pmatrix} \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\beta}_j'} & \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\beta}_j'} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\beta}_j'} & \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_j'} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\beta}_j'} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \end{pmatrix}\right|.$$

Since the partial derivatives in the first row are

$$\frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\beta}_j'} = \boldsymbol{I}_{L_j-1}, \quad \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} = \mathbf{0}, \quad \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} = \mathbf{0}, \quad \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\sigma}_j'} = \mathbf{0}, \quad \frac{\partial \boldsymbol{\beta}_1^f}{\partial \boldsymbol{\sigma}_j^{*\prime}} = \mathbf{0},$$

we have that

$$\left|\det\left(\frac{\partial g_j^{-1}(\boldsymbol{\eta}_j)}{\partial \boldsymbol{\eta}_j'}\right)\right| = \left|\det\begin{pmatrix} \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\sigma}_j'} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \end{pmatrix}\right|.$$

In addition, because

$$\frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\sigma}_j'} = \boldsymbol{I}_{\frac{L_j(L_j-1)}{2}}, \quad \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} = \mathbf{0}, \quad \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} = \mathbf{0}, \quad \frac{\partial \boldsymbol{\sigma}_1^f}{\partial \boldsymbol{\sigma}_j^{*\prime}} = \mathbf{0},$$

we can further reduce the Jacobian to

$$\left|\det\left(\frac{\partial g_j^{-1}(\boldsymbol{\eta}_j)}{\partial \boldsymbol{\eta}_j'}\right)\right| = \left|\det\begin{pmatrix} \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \\ \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}} & \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} \end{pmatrix}\right|.$$

Finally, because

$$\frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} = \boldsymbol{I}_{K-L_j}, \quad \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} = \mathbf{0}, \quad \frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\beta}_{j,m}^{*\prime}} = \mathbf{0}$$

and

$$\frac{\partial \boldsymbol{\sigma}_1^{f*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} = \boldsymbol{I}_{\frac{(K-L_j)(K-L_j+1)}{2}}, \quad \frac{\partial \boldsymbol{\beta}_1^{f*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} = \mathbf{0}, \quad \frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\sigma}_j^{*\prime}} = \mathbf{0},$$

the determinant can be evaluated as

$$\begin{aligned}\left|\det\left(\frac{\partial g_j^{-1}(\boldsymbol{\eta}_j)}{\partial \boldsymbol{\eta}_j'}\right)\right| &= \left|\det\left(\frac{\partial \boldsymbol{\sigma}_1^{ff*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}}\right)\right| \\ &= \left|\det \frac{\partial\left(\Sigma_j \otimes \boldsymbol{I}_{K-L_j}\right)\boldsymbol{\beta}_{j,f}^{*}}{\partial \boldsymbol{\beta}_{j,f}^{*\prime}}\right| \\ &= |\Sigma_j \otimes \boldsymbol{I}_{K-L_j}| \\ &= |\Sigma_j|^{K-L_j}.\end{aligned}$$

■

REFERENCES

Asness, Clifford, and Andrea Frazzini, 2013, The devil in HML's details, *Journal of Portfolio Management* 39, 49–68.

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen, 2014, Quality minus junk, Working paper.

Barillas, Francisco, and Jay Shanken, 2018, Comparing asset pricing models, *Journal of Finance* 73, 715–754.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chib, Siddhartha, 1995, Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association* 90, 1313–1321.

Chib, Siddhartha, and Xiaming Zeng, 2019, Which factors are risk factors in asset pricing? A model scan framework, *Journal of Business & Economic Statistics*.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Pástor, Lubos, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication code.