Comparing Priors for Comparing Asset Pricing Models

FRANCISCO BARILLAS and JAY SHANKEN*

ABSTRACT

We analyze the properties of two prior specifications for the betas and residual (co)variance parameters in the Bayesian procedure for comparing asset pricing models initiated by Barillas and Shanken (2018). We state the key underlying results that led to the alternative prior proposed by Chib, Zeng and Zhao (2020, CZZ) and show that the implied posterior model probabilities are invariant to reparameterization of the models. Simulation evidence indicates that the new prior is preferred, though the dramatic performance differential reported by CZZ is not observed.

* Barillas is from the University of New South Wales School of Banking and Finance, and Shanken is from the Goizueta Business School at Emory University and the National Bureau of Economic Research. Corresponding author: Jay Shanken, Goizueta Business School, Emory University, 1300 Clifton Road, Atlanta, Georgia, 30322, USA; telephone: (404)727-4772; fax: (404)727-5238. E-mail: jay.shanken@emory.edu. Thanks to Rohit Allena for helpful comments.

In Barillas and Shanken (2018, BS), we develop a Bayesian procedure for comparing asset pricing models with traded factors. In this short paper, we discuss a modification of the prior and prove that it satisfies a desirable form of prior consistency across models, as well as invariance to reparametrization. The modified approach retains most of the framework in BS, including the joint treatment of included/excluded model factors and the form of the informative prior for alphas. The difference, which drives consistency and invariance, corresponds to the prior for the so-called "nuisance parameters" – the betas and residual (co)variance parameters.

The key insight that underlies the modified prior is our observation that there is a natural invertible mapping between the nuisance parameters for each pair of models, as explained in Section I below. Prior consistency then leads to the sensible requirement that the nuisance priors for all models be the same after the parameter spaces have been transformed such that they are identical. This condition is violated in BS. Invariance goes further in requiring that the posterior model probabilities be independent of the choice of parametric representation of the models. Our exploration of invariance was prompted by a question from a respected colleague as to whether the modified nuisance priors are truly noninformative for *all* models.

The explicit invertible mapping between the nuisance parameters for different models, together with our related observations about the properties of the corresponding induced priors under reparametrization, serves as the foundation for the work of Chib, Zeng, and Zhao (2019, CZZ). We communicated these thoughts to CZZ in response to an earlier version of their paper - which, for the record, contained no hint of this transformation or the induced prior,[1] – arguing that the issues they raised about differing parameter spaces and prior support were not a problem. CZZ then used these ideas in developing the modified prior and derived a useful expression for this prior. Given the respective contributions of each of us to the analysis that has emerged, hereafter we refer to the modified prior as the BS-CZZ prior. As we show in Section III, the resulting likelihood measure for each model can be obtained from a simple variation of the moment-generating approach of BS.

Whether the technical properties of the nuisance priors considered here will have a material impact in applications is not clear. We, therefore, examine some frequentist properties of the two Bayesian prior approaches. This involves repeated sampling of factor returns simulated under the null that a given pricing model holds. CZZ examine such simulation evidence. However, as we discuss in Section IV, the models that they examine likely suffer from selection bias, given the nature of their data-based model screening procedure. Moreover, although the sole metric that they focus on yields dramatic results favoring the BS-CZZ prior, it provides rather limited information. We therefore explore a broader set of null models and more informative metrics, including a simple economic measure with an investment orientation.

This paper is organized as follows. Section I describes the statistical framework of BS. Section II then analyzes the invariance of model comparison under the BS-CZZ prior. Section III discusses the calculation of marginal likelihoods under the BS-CZZ prior, and Section IV provides simulation evidence under the BS and BS-CZZ priors. Section V concludes.

I. The Statistical Framework of BS

[1] CZZ allude to an argument made by a "reader" (we revealed that one of us was a reviewer of the paper), but do not acknowledge the central role that the ideas in that argument (the explicit mapping and induced prior under reparametrization) played in the development of their paper. Also see footnote 6 below.

Bayesian model comparison ultimately reduces to a comparison of model marginal likelihoods – the result of averaging the likelihood function over the prior for the given model, M. In contrast to earlier work in this area, a key feature of the BS approach involves conditioning the posterior analysis of each model on all of the factor data. Along with some simplifying assumptions on the prior, this yields a marginal likelihood of the form

$$ML = ML_U(f|Mkt) \times ML_R(f^*|Mkt, f), \tag{1}$$

where f denotes the nonmarket factors included in the model, f^* denotes the excluded factors,[2] $ML_U(f|Mkt)$ is the unrestricted ML for the multivariate linear regression of f on the market factor and a constant,

$$f = \alpha + \beta Mkt + \varepsilon, \tag{2}$$

and $ML_R(f^*|Mkt, f)$ is the restricted ML for the regression of f^* on the market and f (constant excluded),

$$f^* = \beta^*[Mkt, f]' + \varepsilon^*. \tag{3}$$

In this way, the model restriction on the intercepts ($\alpha^* = 0$) is imposed in (3), while the expected returns of the model's factors are unconstrained in (2).[3] The market excess return, *Mkt*, is included in all models.

In BS, the prior belief about the model's factor alphas in (2) is taken to be informative, reflecting the researcher's view about the potential magnitude of the investment Sharpe ratio. The prior for the "nuisance parameters" is a product of Jeffreys (1961) priors, with independence between the parameters of (2) and (3). This is an "improper" prior in that the integral over all values of the parameters is not finite (a proper prior integrates to one). Nonetheless, improper priors often give rise to well-defined posterior beliefs and are routinely used in Bayesian analysis. However, the use of improper priors in model comparison raises additional issues that we discussed in BS and return to now.

Consider improper priors $p(\theta_1)$ and $p(\theta_2)$ for models M_1 and M_2 respectively. Since the priors do not integrate to one, without further consideration we could just as well replace these priors with, for example, 10 x $p(\theta_1)$ and 2 x $p(\theta_2)$, increasing the posterior odds in favor of M_1 by a factor of 5 = 10/2. The question then arises as to when is it appropriate to use the same proportionality constant for each model, as we did. Such conditions are discussed by Berger and Pericchi (2001), who consider the common case in which parameters are the same across models, but also consider other scenarios in which the parameters are "essentially similar." We used the latter to motivate our analysis with different nuisance parameters under each model, but below we show that the BS improper prior approach is equivalent to (produces the same posterior probabilities as) one in which the nuisance parameters are actually *common* across models. This follows from the invertible mapping mentioned earlier, which we now describe.

The all-factors model M_1 involves regressing all of the non-market factors on *Mkt* and a constant. The market betas, β_1, of these factors and the corresponding residual covariance matrix, Var_1, together constitute the nuisance parameters θ_1 for M_1. Let θ_2 be the set of nuisance parameters $\{\beta, \beta^*, var(\varepsilon)$ and $var(\varepsilon^*)\}$ for the model M with factors f and excluded factors f^*, as in (2) and (3).

[2] Consistent with Barillas and Shanken (2017), test asset returns give rise to an additional *ML* term that is the same for all models and thus drops out of the model comparison.

[3] Since no factors are excluded in the all-factors model, the *ML* for this model consists only of the first term, ML_U.

In BS, we note that the number of parameters in θ_1 and θ_2 is the same for all models, a condition emphasized by Berger and Pericchi (2001). Here we take this an important step further. Substituting the expression for the included factors *f* from (2) into (3) and collecting terms gives the regression of f^* on *Mkt*. Together with the regression of *f* on *Mkt* in (2), we have θ_1 expressed in terms of θ_2. This is our mapping from θ_2 to θ_1, and some additional algebra shows that the mapping is invertible.

For example, suppose that we start with the three factors of Fama and French (1993): *Mkt*, *HML*, and *SMB*.[4] Consider the two-factor pricing model {*Mkt, HML*}, where *f* is the included factor *HML* and f^* is the excluded factor *SMB*. In this case, (2) and (3) are $HML = \alpha + \beta Mkt + \varepsilon$ and $SMB = \beta^*[Mkt, HML]' + \varepsilon^*$, where $\beta^* = (\beta_M^*, \beta_H^*)$. The nuisance parameters are $\theta_2 = (\beta, \beta_M^*, \beta_H^*, \sigma_\varepsilon^2, \sigma_\varepsilon^{*2})$. Now substituting the expression for *HML* from the first regression into the second yields

$$SMB = \beta_M^* Mkt + \beta_H^*(\alpha + \beta Mkt + \varepsilon) + \varepsilon^* = \beta_H^*\alpha + (\beta_M^* + \beta_H^*\beta)Mkt + (\varepsilon^* + \beta_H^*\varepsilon),$$

where ε and ε^* are uncorrelated with Mkt and with each other. Hence, this equation is the regression of SMB on Mkt and a constant, and so $\theta_1 = (\beta_1, Var_1)$, where $\beta_1 = (\beta, \beta_M^* + \beta_H^*\beta)$ is the 1x2 vector of simple regression coefficients for *HML* and *SMB* on *Mkt* and Var_1 is the 2x2 residual covariance matrix with diagonal elements σ_ε^2 and $\sigma_\varepsilon^{*2} + \beta_H^{*2}\sigma_\varepsilon^2$ and off-diagonal element $\beta_H^*\sigma_\varepsilon^2$. This is the mapping from θ_2 to θ_1. The following result makes use of such a transformation and will be applied to derive our main conclusions.

LEMMA: *Let M be a factor pricing model, as in* (2) *and* (3), *with nuisance parameters* θ *and let* $\varphi = g(\theta)$ *be a one to one mapping such that the inverse is differentiable with Jacobian* $J_g^{-1}(\theta)$. *Then, by a change of variables, the induced prior for* φ *is*

$$p(g^{-1}(\varphi))|det(J_{g^{-1}}(\varphi))|, \tag{4}$$

where $p(\theta)$ *is the prior for* θ. *With* $\pi(\alpha \mid \theta)$ *the (conditional) prior for* α, *the ML for M is unaffected by the reparametrization, that is, the ML is the same with* $p(\theta)$ *or the induced prior for* φ. *In addition, if* $p(\theta)$ *is an improper prior with proportionality constant c, then the induced prior inherits this proportionality constant through the first term,* $p(g^{-1}(\varphi))$.

Proof: First note that the likelihood for (α, φ) *is* $L(\alpha, g^{-1}(\varphi) \mid M, F)$, where *F* is the factor return data. Also, the prior for α under the reparametrization is $\pi(\alpha \mid g^{-1}(\varphi))$. Then the *ML* under the induced prior for φ *is*

$$\iint \pi(\alpha \mid g^{-1}(\varphi))p(g^{-1}(\varphi))L(\alpha, g^{-1}(\varphi) \mid M, F)|det(J_{g^{-1}}(\varphi))|d(\alpha, \varphi). \tag{5}$$

Substituting $\theta = g^{-1}(\varphi)$ and $d(\alpha, \theta) = |det(J_g^{-1}(\varphi))|\, d(\alpha, \varphi)$, this integral reduces to

$$\iint \pi(\alpha \mid \theta)p(\theta)L(\alpha, \theta \mid M, F)d(\alpha, \theta), \tag{6}$$

the *ML* under the original parametrization. Q.E.D.

Let us consider the BS prior approach in light of this result. Without loss of generality, suppose we have two models, M_1 and M_2, with nuisance parameters θ_1 and θ_2, respectively, linked by an invertible mapping, $g(\theta_2) = \theta_1$. Let $p(\theta_1)$ be the BS prior for θ_1. By the lemma (with θ_2 as θ and θ_1 as φ), we can replace θ_2 and its BS prior, $p(\theta_2)$, with θ_1 and the prior induced by $p(\theta_2)$ without

[4] This is precisely the example that we supplied to CZZ in a review of their initial draft.

changing ML_2. Of course, ML_1 is also unchanged, because we have not modified the parametrization of M_1. Therefore, the posterior model probabilities will be unaffected. Note that after the transformation by g, we have the same nuisance parameters, θ_1, for each model and the same prior support.[5] However, there is nothing constraining the induced prior for θ_1 and $p(\theta_1)$ to be the same, a weakness of the original approach that has become evident through this transformation argument. As a result, the original prior specification may favor certain models in unanticipated and unintended ways. Fortunately, a slight twist on this argument motivates a modified approach that essentially amounts to imposing the *same* nuisance prior under each model.

II. The Modified Prior, Consistency, and Invariance of the Model Comparison

Again, we start with models M_1 and M_2 and a prior $p(\theta_1)$ for θ_1. However, rather than map θ_2 to θ_1, we now let g denote the reverse mapping of θ_1 to $\theta_2 = g(\theta_1)$ and we take the prior for θ_2 to be the induced prior based on $p(\theta_1)$ and g. By the lemma (now with θ_1 as θ and θ_2 as φ), the ML_2 obtained using the induced prior for θ_2 is the same as that based on the θ_1 parametrization and $p(\theta_1)$. Moreover, if $p(\theta_1)$ is improper with a proportionality constant c then, as observed in the lemma, the induced prior for θ_2 under M_2 will inherit the same constant, which will then cancel out in the comparison of *ML*s. This modification is therefore equivalent to requiring that the nuisance prior be the same for all models, a sensible consistency condition. This is the prior approach used by CZZ, although they do not mention this interpretation.

While we focus on the case of an improper nuisance prior in our previous work, the main points made here are quite general and driven by the invertible mapping between nuisance parametrizations. The lemma holds for informative nuisance priors as well, and thus the strategy of starting with a prior for one "reference model" and employing the equivalent induced priors for all other models can therefore be applied in this case too. A previous paper by Chib and Zeng (2019) uses a proper prior and notes that "the model-by-model prior must be proper." But as noted above and in our review of the first draft of what became the CZZ paper, the induced prior inherits the proportionality constant under a change of variables and so an improper prior *can* be used, as CZZ proceeded to do. Thus, our invertible mapping and observations about the induced prior led to their modification of our initial approach.

As the three-factor example of the previous section demonstrates, there is more than one way to statistically represent the space of nuisance parameters. Therefore, one may wonder whether the model comparison would be affected if we started with a different parametrization for M_1 or M_2. The following proposition shows that the model comparison would not change.

PROPOSITION: *Suppose the prior for M_1 is $\pi(\alpha_1 \mid \theta_1)p(\theta_1)$. The prior for any other model, M_2, is defined as $\pi(\alpha_2 \mid \theta_2)$ times the induced nuisance prior based on the M_1 prior and the mapping $\theta_2 = g(\theta_1)$.*[6] *i) Let $\theta_{1r} = h(\theta_1)$ be an alternative reparametrization of M_1 with the parametrization of M_2 unchanged. The prior for θ_{1r} is that induced by h and $p(\theta_1)$. Then the induced prior for M_2, the MLs, and the posterior model probabilities are unaffected. ii) If the parametrization of M_1 is fixed and $\theta_{2r} = q(\theta_2)$ is an alternative reparametrization of M_2, then the MLs and the posterior model*

[5] In the first draft of CZZ, they emphasized that these conditions fail to hold in the original BS parametrization. The argument above, an important part of what we provided to the authors in our response, shows that this was not the problem. Our mistake at the time was not immediately recognizing that there was a problem nonetheless.

[6] For the CAPM, there are no alphas and so the π term is simply one. Note that the g used here is the inverse of the g used at the end of the previous section.

probabilities are again unaffected. iii) If the prior for M_1 (whatever the parametrization) is taken to be the Jeffreys prior, then the MLs and the posterior model probabilities do not depend on the initial model parametrizations.

Proof: See Appendix A.

Jeffreys rule is a method for coming up with a prior that involves taking the square root of the determinant of the information matrix. This rule guarantees that the posterior analysis of a given model does not depend on the initial parametrization of that model, a property often associated with the notion of a noninformative prior.[7] The upshot of our proposition is that the entire model comparison analysis is invariant to the initial parametrization of the reference model M_1, as well as the parametrizations of the other models, provided that we start with a Jeffreys prior for M_1. Thus, we extend Jeffreys invariance for a single model to the comparison of a set of models in this context. Next, we show that the BS-CZZ marginal likelihoods can be obtained by a straightforward modification of the method used by BS. We then compare the performance of the BS and invariant BS-CZZ priors in simulations.

III. Calculation of Marginal Likelihoods under the BS-CZZ Prior

Let L be the number of factors, Mkt and f, in model M and let K be the total number of factors, Mkt, f, and f^*, in (2) and (3), where f^* denotes the factors excluded from M. The residual covariance matrix of dimension L-1 for regression (2) of f on Mkt and a constant is Σ and that of dimension K-L for regression (3) of f^* on Mkt and f is Σ^*. Whereas the nuisance prior under BS was $det(\Sigma)^{-L/2}det(\Sigma^*)^{-(K-L+1)/2}$, the modification derived by CZZ changes the exponents in this expression, resulting in the prior

$$P(B, \Sigma) = det(\Sigma)^{-(2L-K)/2}det(\Sigma^*)^{-K/2}. \tag{7}$$

Importantly, this prior is a product of functions of the parameters in (2) and (3). Therefore, the nuisance priors for (2) and (3) are independent and the resulting marginal likelihood for M is a product of restricted and unrestricted MLs, as in BS:

$$ML = ML_U(f \mid Mkt) \times ML_R(f^* \mid Mkt, f) \tag{8}$$

with

$$ML_R(f^* \mid Mkt, f) = \left(\frac{1}{2\pi}\right)^{\frac{(K-L)(T-L)}{2}} 2^{(K-L)v1/2}\Gamma_{(K-L)}(v1/2)det(S^R)^{-v1/2}det(F'F)^{-(K-L)/2}$$

$$ML_U(f \mid Mkt) = \left(\frac{1}{2\pi}\right)^{\frac{(L-1)(T-1)}{2}} 2^{(L-1)v2/2}\Gamma_{(L-1)}(v2/2)det(S)^{-v2/2}(Mkt'Mkt)^{-(L-1)/2}Q$$

$$Q = \left[1 + \frac{a}{a+k}(W/T)\right]^{-v2/2}(1 + k/a)^{-(L-1)},$$

where $v1 = T - 1, v2 = T-(K-L) - 1$, $a = 1 + \widehat{Sh}(Mkt)^2$, and $W = T\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}/a$, with maximum likelihood estimates indicated by the hat symbol. The terms S^R and S denote the residual

[7] See https://en.wikipedia.org/wiki/Jeffreys_prior.

cross-product matrices for the regressions, restricted (no regression constant) and unrestricted, respectively, and $F = (Mkt, f)$.[8] See Appendix B.

IV. Simulation Evidence

In their simulation analysis, CZZ report what they call "% correct," the proportion of simulation iterations for which the null model receives the highest posterior probability. We focus on the sample size T = 600 with eight factors, which is close to that often used in practice.[9] CZZ find that the BS procedure never identifies the correct model in 100 iterations for each of 33 data-generating processes (DGPs) corresponding to different null models based on the eight factors. Indeed, this remains true at much larger sample sizes. In contrast, the BS-CZZ prior delivers % correct measures ranging from 42% to 65%. Given the consistency and invariance properties analyzed in the previous section, it makes sense that the new prior would perform better. But we now argue that these levels of performance and the extreme contrast between results for the two priors may be misleading. We investigate this issue from two perspectives – the first relates to the limited metric employed by CZZ, and the second to potential selection bias in their choice of models to simulate.

The so-called % correct measure is based on an overly simple binary summary outcome: either the highest probability is obtained for the null model or not, which seems contrary to the spirit of the Bayesian perspective, with its continuous measure of degree of belief. Note that it is possible for a procedure to be "correct" in this sense with only the slightest deviation from equal-weighted model probabilities. For example, with 128 models (all models of the eight factors, which include the market factor), equal probabilities would be 0.78%. A procedure that obtains a posterior probability of 1% for the null model, with all other model probabilities a bit lower, would be scored "correct" in that case. Thus, this metric provides minimal information about the relative performance of the procedures. This is easily remedied by keeping track of the average posterior probability assigned to the null model across simulations. However, while it provides useful information, this measure has its own limitations from an economic perspective.

In particular, the Bayesian procedure can yield a modest probability for the null model but place substantial weight on models that are closely related to the null, with similar Sharpe ratios (expected excess return over standard deviation). For example, a related model might include different versions of what we call "categorical factors" in BS (such as the CMA or IA investment factors), or it might leave out a factor that is included in the true model but contributes little to its overall Sharpe ratio. Distinguishing between such a scenario and one in which the procedure tends to exclude economically important factors is essential.

Of course, if all we cared about was maximizing the true Sharpe ratio, we would simply include all factors in the model, ignoring the desirability of parsimony. In practice, however, estimation error in a model's tangency portfolio weights will degrade the portfolio performance. In particular, if we include factors that are redundant, these factors will receive some weight in the estimated tangency portfolio, whereas the true optimal weight is zero. Ideally, a procedure would do a good job of figuring out when to exclude such factors from a model.[10]

---

[8] For simplicity, we let the factor names, *Mkt* or *f*, do double duty and also refer to the time series for that factor.
[9] As in CZZ, the prior Sharpe multiple is taken to be three.
[10] Parsimony may also help obtain more efficient estimates of the cost of capital from a model by avoiding noisy estimation of the contribution that exposure to an irrelevant factor makes to expected returns.

One simple metric that reflects these considerations is the true Sharpe ratio of the *estimated* tangency portfolio for the null model. More specifically, we calculate the weighted average of the true Sharpe ratios for all models, with weights equal to the model posterior probabilities. Following Jobson and Korkie (1980), we estimate the portfolio weights for each model as the unbiased estimate of the inverse of the covariance matrix times the sample mean vector of the model factors, normalized so that the weights sum to one. The data are simulated for a null model M with factors Mkt and f, and excluded factors f^*, from a multivariate normal distribution with mean and covariance matrix taken to be the maximum likelihood estimates of (2) and (3) based on the original data. The Sharpe ratio under the null for a given model (which may be the null or any other model) is then based on the estimated portfolio weights for the model and these "true" moments.

Thus, for each procedure, we report an average across null experiments and 100 simulation iterations for each null of a scaled value for this posterior-weighted Sharpe ratio. The scaling involves dividing by the maximum Sharpe ratio under the null assuming knowledge of the true tangency weights. Thus, the scaled values are all less than one and indicate the fraction of the Sharpe ratio of the true tangency portfolio that is realized using the given Bayesian procedure. We refer to this as the Sharpe ratio (SR) proportion. As additional descriptive information to assess the extent to which a procedure is parsimonious, we provide the posterior-weighted average of the number of model factors for each procedure and the actual number of factors in the null model. Again, these are averaged across simulation iterations and all nulls considered.

We also explore the effect of restricting attention to models that contain at most one version of each categorical factor, a plausible a priori condition that is imposed in BS. In keeping with CZZ, we focus here on frequentist properties of the Bayesian methods. It would be of interest to evaluate the Bayesian predictive distribution of returns for each prior and calculate the certainty-equivalent returns, given an assumption about utility. Doing so is beyond the scope of this short paper but will be explored in future work.[11]

Another aspect of the CZZ simulation analysis raises doubts about taking it at face value. In deciding which models to treat as null model DGPs, CZZ conduct a pre-screening to ensure that each assumed pricing factor is statistically significant. This sort of "model mining" is likely to impart biases in the choice of null models – in addition to large point estimates of factor means, Linnainmaa and Roberts (2018) suggest that unusually low volatility and low factor correlations will often be obtained. To determine whether this has a material impact on the conclusions derived from the simulations, we evaluate the performance of the Bayesian methodologies over *all* models that can be formed from the given factors. That some factor pricing results are not statistically significant in the original data may better reflect the ambiguities that are inherent in actual model comparison,[12] although inevitably some characteristics of the true DGP will be under- and others overstated. In any event, this would provide some additional perspective on the performance of the two prior approaches.

Table 1 reports various measures for the BS and BS-CZZ priors. As a warm-up exercise with a relatively small number of factors and no categorical factors, we start with the four factors of Hou, Xue, and Zhang (2015). Panel A gives the results averaged across both the eight models based on these factors and the 128 models based on the eight factors used by CZZ. With four

---

[11] Also see related work by Pastor and Stambaugh (2000).

[12] Of course, failure to reject a null does not mean it is true, particularly given the volatility of returns.

factors, the performance of the two priors is similar, but the new version has a slight edge. For example, the SR proportions are 95.4% and 95.8%, respectively. With eight factors, the gap widens, although the SR proportions still differ by less than one percentage point. The % correct and Prob(null) measures are substantially reduced for both priors with eight factors. While the % correct for BS-CZZ is about five times that for BS, the more relevant Prob(null) is a little more than two times that for BS.

Table I

Simulations: BS and BS-CZZ prior

The table reports simulation results under both the BS and the BS-CZZ priors. Panel A gives the results averaged across the eight models based on the four factors of Hou, Xue, and Zhang (2015) as well as for the 128 models based on the eight factors used by CZZ. Panel B reports the results for the eight-factor analysis that excludes models that have two versions of the same categorical factor. % correct denotes the proportion of times the procedure gives the null model the highest posterior model probability. P(null model) is the average model probability given to the null model. SR prop is the fraction of the population Sharpe ratio under the null achieved by the procedure described in the text that incorporates sample uncertainty in estimating tangency portfolio weights. Factors mean is the posterior-weighted average of the number of model factors for each prior and Factors null is the average number of factors in the null model.

Panel A. All Models

Prior	BS	BS-CZZ	BS	BS-CZZ
Number Factors	4	4	8	8
% correct	67.9%	77.0%	4.0%	20.5%
P(null model)	51.9%	55.7%	3.2%	7.2%
SR prop	95.4%	95.8%	94.6%	95.5%
Factors mean	2.77	2.79	5.28	4.97
Factors null	2.50	2.50	4.50	4.50

Panel B. Models with Categorical Restriction

Prior	BS	BS-CZZ
Number Factors	8	8
% correct	17.3%	33.3%
P(null model)	13.5%	15.5%
SR prop	91.8%	94.1%
Factors mean	3.52	4.29
Factors null	3.83	3.83

In Panel B we turn to the eight-factor analysis, which excludes models that have two versions of the same categorical factor. Here, % correct for BS is about half that of BS-CZZ. However, consistent with our earlier discussion of the limitations of this metric, the probabilities for the null model are quite close, at 13.5% versus 15.5%. In addition, while the BS-CZZ prior leads to inclusion of 0.8 more factors than BS on average, the SR proportion is 94.1% using this modified prior, as compared to 91.8% for BS. These results are far from the extreme performance of the BS method that CZZ report for the % correct metric and the pre-screened models.

V. Conclusion

The desirable nuisance-prior consistency and invariance properties that we establish for model comparison with the BS-CZZ prior formalize the analytical sense in which this modified prior is preferred to the original BS prior. However, while this suggests that the new approach will perform better, it is not clear what to anticipate in terms of quantitative magnitudes. In order to shed light on the latter question, we conduct simulation analysis. The results suggest that, while our various metrics consistently favor the new method, the divergence between the two prior approaches is far less than the CZZ simulations suggest. Not surprisingly, the ability to identify the "correct" model with the new prior declines considerably in the eight-factor analysis with a broader set of DGPs. Thus, our results may provide a more realistic indication of the "power" of using the BS-CZZ prior. Nonetheless, the findings with the economically oriented SR proportion metric are encouraging and suggest that the models favored using both prior approaches tend to be closely related to the true models, but with the edge going to the BS-CZZ prior.

We can also report that if we use the BS-CZZ prior with the 10 factors explored in BS, the model with the highest probability when the categorical restriction is imposed is the same six-factor model identified in BS, namely, {Mkt ROE HML^m UMD SMB IA}. Also, as in BS, the more timely versions of the value and profitability factors (HML^m and ROE), together with momentum (UMD) and the market factor, are included in the top seven models, which garner most of the posterior probability. Thus, the main empirical conclusions continue to hold.

Appendix A: Proof of the Proposition

i) A direct application of the lemma with θ_1 and θ_{1r} in the place of θ and φ, respectively, implies that ML_1 is unchanged. Now consider the effect of the reparametrization of M_1 on the ML for M_2 with parameters θ_2. Note that $\theta_2 = g(\theta_1) = g(h^{-1}(\theta_{1r}))$ and so the new induced prior for θ_2 is the induced prior based on the mapping $g \circ h^{-1}$ and the prior for θ_{1r}. But the prior for θ_{1r} is itself induced by the mapping h and the prior $p(\theta_1)$,

$$p(h^{-1}(\theta_{1r}))|J_{h^{-1}}(\theta_{1r}))| \,, \tag{A1}$$

as in (4). Using (4) again, to get the induced prior for θ_2, this time we first evaluate the prior for θ_{1r} in (A1) at $(g \circ h^{-1})^{-1} = h \circ g^{-1}$ applied to θ_2 and then multiply by the corresponding Jacobian term for $h \circ g^{-1}$. The expression based on (A1) is

$$p(h^{-1}(h \circ g^{-1}(\theta_2)))|det(J_{h^{-1}}(h \circ g^{-1}(\theta_2))) = p(g^{-1}(\theta_2))|\det(J_h(g^{-1}(\theta_2)))^{-1}|, \tag{A2}$$

where we use the inverse function theorem to evaluate $J_{h^{-1}}$.[13] We still need to multiply (A2) by the Jacobian term $|det(J_{h \circ g^{-1}}(\theta_2))|$ to finish off the induced prior for θ_2. By the chain rule, $J_{h \circ g^{-1}}(\theta_2)$ is $J_h(g^{-1}(\theta_2))J_{g^{-1}}(\theta_2)$.[14] Therefore, cancelling the J_h term with its inverse, we are left with $p(g^{-1}(\theta_2))$ times $|det(J_{g^{-1}}(\theta_2)|$. Thus, the new induced prior for θ_2 is the original prior induced by $p(\theta_1)$ and the mapping $\theta_2 = g(\theta_1)$. The rest of i) follows.

ii) ML_1 is unchanged by assumption. As to M_2, we can write $\theta_{2r} = q(\theta_2) = q(g(\theta_1))$. Therefore, by the lemma with $q \circ g$ in the role of the mapping, ML_2 is the same with $p(\theta_1)$ or the induced prior for θ_{2r}. Similarly, ML_2 is the same with $p(\theta_1)$ or the induced prior for θ_2. Thus, ML_2 is the same with either induced prior (for θ_2 or for θ_{2r}).

iii) If we start with parameters θ_{1r} instead of θ_1, then the Jeffreys prior for θ_{1r} will, by the usual invariance property, be the prior induced by $p(\theta_1)$ and the mapping h. Thus, part (i) of the proposition applies and ii) applies in general. Q.E.D.

Appendix B: Calculating the BS-CZZ Marginal Likelihoods

Here, we discuss the changes to the appendices in Barillas and Shanken (2018) that are needed to accommodate the BS-CZZ prior. We start with the internet Appendix to BS, with N the number of left-hand-side factors, K the number of right-hand-side factors, and Σ the residual covariance matrix. As in BS, we first consider the restricted case (no constant in the regressions) and then the unrestricted case. The nuisance prior is now assumed to be

$$P(B, \Sigma) = det(\Sigma)^{-z/2}, \tag{B1}$$

generalizing $z = N+1$ in BS. The likelihood function is unchanged.

It follows that the degrees of freedom in the inverted Wishart density that emerges in the marginal likelihood derivation are $v = T\text{-}K\text{-}N\text{-}1 + z$, which equals $T\text{-}K$ when $z = N+1$. As a result,

$$ML_R = \left(\frac{1}{2\pi}\right)^{\frac{N(T-K)}{2}} 2^{Nv/2}\Gamma_N(v/2)det(S^R)^{-v/2}det(F'F)^{-N/2}. \tag{B2}$$

For the unrestricted marginal likelihood, the change in degrees of freedom from $T\text{-}K$ to v also affects the calculation of Q in (B2) of Appendix B to BS. This gives

[13] See https://en.wikipedia.org/wiki/Inverse_function_theorem.

[14] See https://en.wikipedia.org/wiki/Chain_rule.

$$ML_U = \left(\frac{1}{2\pi}\right)^{\frac{N(T-K)}{2}} 2^{Nv/2}\Gamma_N(v/2)det(S)^{-v/2}det(F'F)^{-N/2}Q, \tag{B3}$$

$$Q = \left[1 + \frac{a}{a+k}\left(\frac{W}{T}\right)\right]^{-v/2}\left(1+\frac{k}{a}\right)^{-N/2},$$

where $v = T\text{-}K\text{-}(N\text{+}1)\text{+}z$, $a = 1 + \widehat{Sh}(Mkt)^2$, and $W = T\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}/a$.

The likelihood ML_R in (8) is now obtained by letting *K-L* play the role of *N* and *L* the role of *K* in (B2) above. The prior input *z* is chosen to give the exponent for Σ^* in (7), that is, *z* = *K*. The corresponding v is *T-L-(K-L+1)+z = T-K-1+z. = T-1.* The likelihood ML_U is obtained by substituting *L-1* for *N, 1* for *K* and *Mkt* for *F* in (B3) above. Then *z* is chosen to give the exponent for Σ in (7), i.e., *z = 2L-K*. The corresponding v is *T-1-(L-1+1)+*z = *T-1+L-K.* We have confirmed that (8) gives identical *MLs* to those obtained using the CZZ method.

REFERENCES

Barillas, F., and J. Shanken, 2017, Which alpha? *Review of Financial Studies* 30, 1316-1338.

Barillas, F., and J. Shanken, 2018, Comparing asset pricing models, *Journal of Finance* 73, 715-754.

Berger, James, and Luis Pericchi, 2001, Objective Bayesian methods for model selection: Introduction and comparison, IMS Lecture Notes: Monograph Series 38, 137–193.

Chib, S., and X. Zeng, 2019, Which factors are risk factors in asset pricing? A model scan framework, *Journal of Business & Economic Statistics* 38, 771-783.

Chib, S., X. Zeng, and L. Zhao, 2020, On comparing asset pricing models, *Journal of Finance* 75, 551-577.

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

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

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

Jobson, J. D., and B.M. Korkie, 1980, Estimation of Markowitz efficient portfolios, *Journal of the American Statistical Association* 75, 544-554.

Linnainmaa, Juhani, T., and Michael R. Roberts, 2018, The history of the cross-section of stock returns, *Review of Financial Studies* 31, 2606–2649.

Pastor, Lubos, and Robert F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* 56, 335–381.