Which Factors are Risk Factors in Asset Pricing? A Model Scan Framework

Siddhartha CHIB

Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Brookings Drive, St. Louis, MO 63130 (*chib@wustl.edu*)

Xiaming ZENG

Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Brookings Drive, St. Louis, MO 63130 (*zengxiaming@wustl.edu*)

A key question for understanding the cross-section of expected returns of equities is the following: which factors, from a given collection of factors, are risk factors, equivalently, which factors are in the stochastic discount factor (SDF)? Though the SDF is unobserved, assumptions about which factors (from the available set of factors) are in the SDF restricts the joint distribution of factors in specific ways, as a consequence of the economic theory of asset pricing. A different starting collection of factors that go into the SDF leads to a different set of restrictions on the joint distribution of factors. The conditional distribution of equity returns has the same restricted form, regardless of what is assumed about the factors in the SDF, as long as the factors are traded, and hence the distribution of asset returns is irrelevant for isolating the risk-factors. The restricted factors models are distinct (nonnested) and do not arise by omitting or including a variable from a full model, thus precluding analysis by standard statistical variable selection methods, such as those based on the lasso and its variants. Instead, we develop what we call a *Bayesian model scan* strategy in which each factor is allowed to enter or not enter the SDF and the resulting restricted models (of which there are 114,674 in our empirical study) are simultaneously confronted with the data. We use a Student-t distribution for the factors, and model-specific independent Student-t distribution for the location parameters, a training sample to fix prior locations, and a creative way to arrive at the joint distribution of several other model-specific parameters from a single prior distribution. This allows our method to be essentially a scaleable and tuned-black-box method that can be applied across our large model space with little to no user-intervention. The model marginal likelihoods, and implied posterior model probabilities, are compared with the prior probability of 1/114,674 of each model to find the best-supported model, and thus the factors most likely to be in the SDF. We provide detailed simulation evidence about the high finite-sample accuracy of the method. Our empirical study with 13 leading factors reveals that the highest marginal likelihood model is a Student-t distributed factor model with 5 degrees of freedom and 8 risk factors.

KEY WORDS: Bayes inference; Marginal likelihood; MCMC sampling; Metropolis–Hastings; Pricing kernel; Stochastic discount factor.

1. INTRODUCTION

A fundamental goal of theoretical and empirical finance is to explain and measure the risk premium (the difference between the expected return of the asset and the risk-free return) for the cross-section of assets in financial markets. According to the factor theory of asset pricing, financial assets earn a risk-premium because the returns on those assets are systematically related to the underlying pricing factors. Such factors are called risk factors and represent sources of systematic risk affecting all assets. An important element of the factor theory is that the same set of risk factors affect the risk premium of all assets, and that these different assets earn different risk premiums because of different covariances (also called exposures) to the underlying risk factors. The original such risk factor is the so-called market portfolio proposed in the 1960s. Knowledge of that risk factor led to the prescription of market index investing (one just holds a version of the market portfolio rather than the individual assets), which is a key principle that underpins the investment/wealth management industry.

According to the theory of asset pricing, a risk-factor is any variable that is in the stochastic discount factor (SDF), or pricing kernel. In other words, a factor is a risk factor if it is in the SDF, and it is not a risk factor if it not in the SDF (Cochrane 2009). The question of establishing the identity of such risk factors has generated a large literature in finance (see, e.g., Fama and French 2017; Harvey and Liu 2017; Pukthuanthong, Roll, and Subrahmanyam 2017) though the question of which factors in total are in the SDF is far from settled. The importance of settling this question is that, first, it is central for evaluating portfolio performance and, second, it has a bearing on investment strategies of both individuals and professional fund-managers. If one determines, for example, the identity of the relevant risk factors, then from an investment perspective it is sometimes beneficial to hold the risk factors in an investment

portfolio, assuming that the factors are investable, rather than the individual assets, a principle referred to as factor investing. This is because the risk of the individual assets is composed of the risk arising from the covariance with the returns of the risk factors plus an idiosyncratic risk which, in principle, can be diversified away. One can avoid having to find the appropriate diversification strategy by just holding the relevant risk factors (Ang 2014). Isolating such risk factors, therefore, has enormous practical implications.

In this article, we provide a Bayesian approach for isolating the risk factors, namely the factors that are in the SDF. In particular, if we denote the SDF at time t by M_t, and let $\boldsymbol{f}_t : d \times 1$ denote the factors in contention for being in the SDF, then the goal is to find the collection of factors $\boldsymbol{x}_t : d_x \times 1$ $(1 \leq d_x \leq d)$ from $\boldsymbol{f}_t$ that are in the SDF and the complementary set of factors $\boldsymbol{w}_t : d_w \times d_w$ $(d_x + d_w = d)$ that are not in the SDF. For any given decomposition of the factors in this way, the SDF-based pricing condition implies that when the joint distribution of the factors is written in terms of a marginal distribution of the x-factors and a conditional distribution of the w-factors given the x-factors, the latter conditional distribution must have zero intercepts (assuming that all factors are traded). A different starting collection of factors that go into $\boldsymbol{x}_t$ leads to a different M_t, which leads to a different set of restrictions on the factor model. This means that we can identify the x-factors in the SDF by simultaneously confronting each of the resulting restricted factor models with the data. The conditional distribution of equity returns has the same restricted form, regardless of what is assumed about the factors in the SDF, as long as the factors are traded, and hence the distribution of asset returns is irrelevant for isolating the risk-factors. It should be noted that the restricted factor models are distinct (nonnested) and do not arise by omitting or including a variable from a full model, thus precluding analysis by standard statistical variable selection methods, such as those based on the lasso and its variants.

In our empirical analysis, we consider a set of 13 potential risk factors and suppose that the joint distribution of the factors is Student-t with unknown degrees of freedom ν_f. Under this collection of factors, and a grid consisting of 14 degrees of freedom, our universe of models consists of $(2^{13} - 1) \times 14 = 114,674$ possible models. We estimate each of these models by tuned Bayesian methods and compare these models in terms of marginal likelihoods. There is a powerful theoretical basis to model comparisons via marginal likelihoods (see, e.g., Chib, Shin, and Simoni 2018). This theory shows that, as the sample size goes to infinity, the model picked according to the highest value of the marginal likelihood is either the true model or the model that is closest to the true model in the Kullback–Leibler information sense. Of course, ultimately, we are interested in finite-sample comparisons. With this in mind, much of our approach is devoted to careful consideration of the prior distribution on the parameters of each of our 114,674 models. Our formulation of this prior distribution, which must be proper, can act as a template for other similar large-scale model comparison problems. Clearly, it is important that the prior distributions across models should not steer the marginal likelihood to favor one or the other models. It also must be set

up in an automatic way in order to allow the comparison of such a large collection of models to proceed in a black-box fashion with little or no user-intervention.

An important feature of this prior is that starting with a prior on the covariance matrix of the factors, Ω, it is possible to construct the implied prior on several key parameters of the model, for any choice of the x and w-factors. This key feature, which we explain in Section 2, allows us to by-pass the model-by-model prior specification of a large number of parameters in the universe of possible models. In addition, we set up the hyperparameters of our prior distribution from a training sample approach, where the training sample refers to a sample of observations prior to the estimation sample. The use of a training sample serves to bring a measure of objectivity to this prior that is difficult to ensure otherwise. We show that the prior we work with satisfies all the criteria enumerated in the preceding paragraph, namely it is a black-box prior that requires little to no user-intervention, is objective and does not predispose the choice of models in contradiction to the sample evidence. We demonstrate the last of these features by the sample the prior method of Chib and Ergashev (2009) under which the factor data are simulated for each model by sampling the parameters from the prior of that model and then sampling the factors of that model given the parameters. If these sampled distributions are the same (or approximately the same) across models, then the implication is that the effect of the model-specific priors is similar, and differences in the marginal likelihoods reflect the ability of the different models in describing the observed data. We calculate the marginal likelihoods as a by-product of the Markov chain Monte Carlo (MCMC) simulation output by the method of Chib (1995) and its Metropolis–Hastings version in Chib and Jeliazkov (2001).

We thoroughly document the effectiveness of our framework to pick the true model from the universe of possible models in carefully designed simulation studies that mimic the real-world features of the factors that are typically employed in this setting. In this simulation, we involve 11 factors and use our framework to calculate the marginal likelihood of all possible models that cover a grid of $\nu_f = 4, 8, 16, 32,$ and ∞ degrees of freedom, for samples of size $T = 600, 1200,$ and $2400,$ and for 100 replications of the data for each sample size. The results show that our method is capable of locating the true model accurately. This accuracy increases with sample size as per the asymptotic theory of the marginal likelihood.

Finally, we provide detailed results from the application of our method to the actual data on 13 common risk factors. These are the Fama and French (1993, 2015) factors: excess market return (MKT), size (SMB), value (HML), profitability (RMW), and investment (CMA); the Hou, Xue, and Zhang (2015) q-factors: size (ME), profitability (ROE), and investment (IA); the Carhart (1997) momentum (MOM) factor; 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 (HMLD) factor proposed by Asness and Frazzini (2013). In this real-data analysis, we use a training sample that runs from January 1968 to December 1972 to form our prior distribution and an estimation sample

that runs from January 1973 to December 2015 to estimate and compare 114,674 factor models. This analysis reveals that the highest marginal likelihood model is a Student-t distributed factor model with 5 degrees of freedom in which MKT, SMB, RMW, ROE, MOM, QMJ, BAB, and HMLD are the x-factors and HML, CMA, ME, IA, and LIQ are the w-factors.

Our article adds to the literature on the use of Bayesian techniques in finance (see Ruppert 2010). For example, Shanken (1987) and Harvey and Zhou (1990) used the Bayesian framework to test the efficiency of a given portfolio, while Avramov and Chao (2006) used Bayes factors to compare conditional linear factor models with prespecified traded factors. More recently, Barillas and Shanken (2018) answer the same question as in this article by fitting factor models and comparing marginal likelihoods but, in their framework, both the likelihood and prior components of the Bayesian model are questionable. First, the MKT factor is assumed to be a risk-factor in all models, whereas in our formulation the status of the MKT factor is not preestablished in this way. Second, Barillas and Shanken (2018) assume that the joint factor distribution is multivariate normal, an assumption that is at odds with the data. In our empirical study, we show that the multivariate Student-t distribution is better supported than the multivariate normal distribution, and that the best model under the normality assumption is quite different from the best-supported Student-t model. Third, and critically, Barillas and Shanken (2018) use off-the-shelf Jeffreys' priors on nuisance parameters that violate a required *across-models* change-of-variable formula, thus rendering the marginal likelihoods noncomparable. In extensive simulation experiments, Chib, Zeng, and Zhao (2018) show that, for each of 500 true DGP's and 100 replications of the data from each of those DGP's, in a model space of 2048 models, the Barillas and Shanken (2018) method does not detect the true DGP even once, even for samples of size 12,000 (corresponding to a thousand years of monthly data).

The remainder of the article proceeds as follows. In Section 2, we describe the model, data and prior. In Section 3, we give the MCMC algorithm for sampling the posterior distribution and the algorithm for calculating the marginal likelihood. Section 4 deals with the results from our simulation studies and Section 5 gives the results from the analysis of the real data. Section 6 concludes the article.

2. MODEL, DATA, AND PRIOR

2.1. Model

In our framework, each factor can be either present in the SDF (in which case it is called the x-factor) or absent from the SDF (in which case it is the w-factor). The collection of factors that are in the SDF are denoted by $x_t : d_x \times 1$ and the complementary collection of factors by $w_t : d_w \times 1$. Now consider the joint model of factors and asset returns. Writing this joint model as a marginal model of the factors times the conditional model of the asset returns given the factors it follows that the latter conditional model depends on both x_t (the pricing factors) and the w-factors. Thus, the conditional distribution of asset returns is the same, regardless of the identity of the risk factors. Therefore, this conditional model

does not help to provide evidence about whether a particular factor is in the SDF. Henceforth, therefore, it is enough to focus our attention on the factor model. Barillas and Shanken (2017) made the same point but by a different argument.

Factors: Suppose now that the factors follow the Student-t distribution

$$f_t \overset{\text{iid}}{\sim} \mathcal{St}_d(\mu, \Omega, \nu_f) , \ t \geq 1 \tag{2.1}$$

where μ : $d \times 1$ (the mean vector), Ω : $d \times d$ (the positive-definite dispersion matrix) and $\nu_f > 2$ (the degrees of freedom) are unknown parameters. The Student-t assumption is important in the current case because the factor data tend to display fatter tails than those of the Gaussian distribution (Fama 1965; Affleck-Graves and Mcdonald 1989; Richardson and Smith 1993; Zhou 1993; Dufour, Khalaf, and Beaulieu 2003). Making use of the well-known representation of the Student-t distribution as a Gamma-scale mixture of normal distributions, we have that

$$f_t | \tau_{f,t} \sim \mathcal{N}_d(\mu, \tau_{f,t}^{-1} \Omega), \tag{2.2}$$

$$\tau_{f,t} \overset{\text{iid}}{\sim} \mathcal{G}\left(\frac{\nu_f}{2}, \frac{\nu_f}{2}\right), \tag{2.3}$$

where the scale $\tau_{f,t} > 0$ is latent. Now for a particular model in the universe of possible models (constructed by letting a collection of factors take the role of the x-factors and the complementary set take the role of the w-factors), let us suppose that the factors are rearranged as $f_t = (x_t, w_t)$, and the parameters are partitioned accordingly as

$$\mu = \begin{pmatrix} \mu_x \\ \mu_w \end{pmatrix}$$

and

$$\Omega = \begin{pmatrix} \Omega_x & \Omega_{xw} \\ (d_x \times d_x) & \\ \Omega'_{xw} & \Omega_w \\ (d_w \times d_x) & (d_w \times d_w) \end{pmatrix} .$$

Let

$$\Gamma = \Omega'_{xw}\Omega_x^{-1} : d_w \times d_x$$

denote the matrix of regression coefficients in the regression of the w-factors on the x-factors. Conditioned on the scale $\tau_{f,t}$, write the factor model as the marginal distribution of the x-factors and the conditional distribution of the w-factors given the x-factors,

$$x_t = \mu_x + \eta_{x,t}, \tag{2.4}$$

$$w_t = \mu_w + \Gamma(x_t - \mu_x) + \eta_{w \cdot x,t}, \tag{2.5}$$

where

$$\begin{pmatrix} \eta_{x,t} \\ \eta_{w \cdot x,t} \end{pmatrix} \bigg| \tau_{f,t} \sim \mathcal{N}_d \left(0, \tau_{f,t}^{-1} \begin{pmatrix} \Omega_x & 0 \\ 0 & \Omega_{w \cdot x} \end{pmatrix} \right) \tag{2.6}$$

and

$$\Omega_{w \cdot x} = \Omega_w - \Omega'_{xw}\Omega_x^{-1}\Omega_{xw} : d_w \times d_w.$$

A remark about notation. Although it would be more precise to label x_t, w_t and the parameters by a model subscript, we refrain from doing so to avoid notational clutter. It should be

understood, however, that every aspect of this factor model is model-specific.

SDF: Now suppose that the SDF M_t, defined following Hansen and Jagannathan (1991), is given by

$$M_t = 1 - \frac{v_f - 2}{v_f} \boldsymbol{\lambda}_x' \Omega_x^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_x), \qquad (2.7)$$

where $\boldsymbol{\lambda}_x$ is a vector of unknown coefficients, and each element in the vector $\Omega_x^{-1} \boldsymbol{\lambda}_x$ (the market price of factor risks) is nonzero by virtue of the definition of $\boldsymbol{x}_t$. The SDF-based economic pricing theory dictates that (excess returns or return spreads) $\boldsymbol{f}_t$ must satisfy the vector pricing restrictions

$$\mathbb{E}(M_t \boldsymbol{f}_t) = \boldsymbol{0}, \qquad (2.8)$$

where the expectation is with respect to the joint distribution of $\boldsymbol{f}_t$. By direct calculation it can be seen that these pricing restrictions require that

$$\boldsymbol{\mu}_x = \boldsymbol{\lambda}_x, \qquad (2.9)$$
$$\boldsymbol{\mu}_w = \Gamma \boldsymbol{\lambda}_x, \qquad (2.10)$$

or, in other words, that the marginal and conditional distributions of the factors in Equations (2.4)–(2.6) take the restricted form

$$\boldsymbol{x}_t = \boldsymbol{\lambda}_x + \boldsymbol{\eta}_{x,t}, \qquad (2.11)$$
$$\boldsymbol{w}_t = \Gamma \boldsymbol{x}_t + \boldsymbol{\eta}_{w \cdot x,t}, \qquad (2.12)$$

where $(\boldsymbol{\eta}_{x,t}, \boldsymbol{\eta}_{w \cdot x,t}) | \tau_{f,t}$ are distributed as in Equation (2.6). This shows that assumptions about which factors are in the SDF restricts the joint distribution of factors in specific ways. A different starting collection of x-factors leads to a different final model of the factors. Our idea, therefore, is to simultaneously confront each of these possible restricted factor models with the data, and compare them in terms of the Bayesian marginal likelihood criterion. The best-supported pricing factors are the x-factors in the model with the largest marginal likelihood. It may be noted that simultaneous comparison in this way avoids postmodel selection biases (in the frequentist sense), because we enumerate all the models before the analysis of the data.

We explain our procedure with the help of the following running example.

Example 1. (q-factor model) Suppose the x-factors are $\{MKT, ME, ROE, IA\}$ as in the q-factor model. Then $\boldsymbol{x}_t = (MKT_t, ME_t, ROE_t, IA_t)$ with length $d_x = 4$ and $\boldsymbol{w}_t = (SMB_t, HML_t, RMW_t, CMA_t, MOM_t, QMJ_t, LIQ_t, BAB_t, HMLD_t)$ with length $d_w = 9$. In this case,

$$\begin{pmatrix} MKT_t \\ ME_t \\ ROE_t \\ IA_t \end{pmatrix} = \underbrace{\boldsymbol{\lambda}_x}_{4 \times 1} + \boldsymbol{\eta}_{x,t} \qquad (2.13)$$

$$\begin{pmatrix} SMB_t \\ HML_t \\ RMW_t \\ CMA_t \\ MOM_t \\ QMJ_t \\ LIQ_t \\ BAB_t \\ HMLD_t \end{pmatrix} = \underbrace{\Gamma}_{9 \times 4} \begin{pmatrix} MKT_t \\ ME_t \\ ROE_t \\ IA_t \end{pmatrix} + \boldsymbol{\eta}_{w \cdot x,t} \qquad (2.14)$$

with $\Omega_x : 4 \times 4$ and $\Omega_{w \cdot x} : 9 \times 9$.

2.2. Data

In our analysis, we involve 13 prominent factors that are often considered in the literature. The sample of the 13 factors we use here is from January 1968 to December 2015, with 576 observations in total.[1] Table 1 reports the sample correlation among the factor returns. As we see from Table 1, most of the factors have significant nonzero correlations.

In our analysis, we use an initial portion of these data, which runs from January 1968 to December 1972 to form our prior distribution, and the subsequent portion, which runs from January 1973 to December 2015, to estimate and compare all our 114,674 factor models.

2.3. Prior Specification

As mentioned earlier, the model-by-model prior must be proper, that is, with integral over the parameter space equal to one. In addition, the prior distributions should, in some sense, be comparable across the different models so that the differences in marginal likelihoods are not merely caused by the differences in the priors. Finally, it is important that the formulation of these model-specific priors should require little to no user-intervention.

We now discuss how we specify such a model-specific prior. The parameters of a particular restricted factor model are given by

$$\boldsymbol{\theta} = (\boldsymbol{\lambda}_x, \Omega_x, \Omega_{w \cdot x}, \boldsymbol{\gamma}),$$

where $\boldsymbol{\gamma} = \text{vec}(\Gamma) : q \times 1$ and $q = d_w \times d_x$. These parameters, in general, lie in a high-dimensional parameter space. For example, in Example 1, the total number of parameters in the model is 95. If the prior is not formulated sensibly, it is possible that the prior may concentrate in regions of the parameter space that are in conflict with the model (likelihood function).

We develop our model-specific prior by using an initial portion of the data (the training sample) to arrive at the hyperparameters of each prior distribution. In our application, this training sample runs from January 1968 to December 1972, with 60 observations in total. From our experience, the specific period used as a training sample is rather inconsequential because of the fact that the training sample is used primarily for locating the mean of the prior distribution; the spread of the prior distribution is largely a user-specified hyperparameter to ensure adequate uncertainty about the choice of location. Of course, the training sample is employed solely for the construction of the model-specific prior and is not used in the model estimation procedure.

Prior of $\boldsymbol{\lambda}_x$: Based on the training sample, we represent our prior knowledge on each element of $\boldsymbol{\lambda}_x$, model-by-model, by a product of Student-t distributions. We use Student-t distributions, instead of (say) normal distributions, to allow a particular component to shrink separately of the other components. Also, the thick tails of the Student-t distribution (under our small degrees of freedom) is helpful in avoiding prior-likelihood conflicts. Let $\boldsymbol{\lambda}_x = (\lambda_1, \ldots, \lambda_{d_x})$. Then, our prior density is

---

Table 1. Correlation of candidate risk factors, monthly observations from January 1968 to December 2015.

	MKT	SMB	HML	RMW	CMA	ME	ROE	IA	MOM	QMJ	LIQ	BAB
SMB	0.27**											
HML	−0.31**	−0.12*										
RMW	−0.21**	−0.38**	0.11*									
CMA	−0.40**	−0.09	**0.71****	−0.06								
ME	0.26**	**0.97****	−0.08	−0.37**	−0.05							
ROE	−0.20**	−0.39**	−0.10	0.67**	−0.08	−0.32**						
IA	−0.38**	−0.19**	0.69**	0.06	**0.91****	−0.14**	0.04					
MOM	−0.14*	−0.05	−0.16**	0.09	0.01	−0.01	0.50**	0.03				
QMJ	−0.53**	−0.52**	0.02	**0.76****	0.08	−0.50**	0.69**	0.15**	0.26**			
LIQ	−0.07	−0.03	0.02	0.02	0.02	−0.03	−0.05	0.02	−0.01	0.04		
BAB	−0.09	−0.02	0.39**	0.25**	0.32**	0.01	0.26**	0.34**	0.19**	0.19**	0.06	
HMLD	−0.13*	−0.01	**0.77****	−0.07	0.51**	−0.01	−0.45**	0.50**	−0.65**	−0.22**	0.06	0.13*

NOTE: The top five correlations are in bold. Significance levels: 0.001 (**), and 0.01 (*).

given by

$$\pi(\boldsymbol{\lambda}_x) = \prod_{i=1}^{d_x} \mathcal{S}t(\lambda_i|\nu, \lambda_{0,i}, s_\lambda), \qquad (2.15)$$

where $\mathcal{S}t(\cdot|\nu, \lambda_{0,i}, s_\lambda)$ denotes the Student-t density. For computational convenience, we represent this prior density in terms of a scale mixture of normal densities as

$$\pi(\boldsymbol{\lambda}_x|\boldsymbol{\tau}_\lambda) = \prod_{i=1}^{d_x} \mathcal{N}(\lambda_i|\lambda_{0,i}, \tau_{\lambda,i}^{-1} s_\lambda), \qquad (2.16)$$

$$\pi(\boldsymbol{\tau}_\lambda) = \prod_{i=1}^{d_x} \mathcal{G}\left(\tau_{\lambda,i}\Big|\frac{\nu}{2}, \frac{\nu}{2}\right), \qquad (2.17)$$

where $\mathcal{N}(\cdot|\lambda_{0,i}, \tau_{\lambda,i}^{-1} s_\lambda)$ and $\mathcal{G}\left(\cdot|\frac{\nu}{2}, \frac{\nu}{2}\right)$ are normal and Gamma densities, respectively, $\lambda_{0,i}$ is the prior mean, s_λ is the dispersion, ν is the number of degrees of freedom and $\boldsymbol{\tau}_\lambda = (\tau_{\lambda,1}, \ldots, \tau_{\lambda,d_x})$ are the latent scale random-variables. In fixing the hyperparameters, we set $\boldsymbol{\lambda}_0 = (\lambda_{0,1}, \ldots, \lambda_{0,d_x})$ to be the sample average of x-factors in the training sample, s_λ to equal 0.0025 and ν to equal 2.1 (a choice that ensures that the distribution has a finite variance plus thick-tails), which implies that the prior standard deviation of each component of $\boldsymbol{\lambda}_x$ is about 0.2291. This is a generously dispersed distribution that implies a prior range of (2)(0.2291), or about 45%, on either side of $\lambda_{0,i}$, where the latter quantity is typically about .1%.

Prior of $(\Omega_x, \Omega_{w\cdot x}, \boldsymbol{\gamma})$: For any choice of the x and w-factors, we set up the prior on these parameters from a *single* inverse Wishart prior on Ω

$$\Omega \sim \mathcal{IW}_d(\rho_0, Q_0), \qquad (2.18)$$

where ρ_0 denotes the degrees of freedom of this distribution and the scale matrix Q_0 is derived from the prior mean Ω_0 on Ω as

$$Q_0 = (\rho_0 - d - 1)\Omega_0. \qquad (2.19)$$

We set ρ_0 as $d + 6$, which leads to $Q_0 = 5 \times \Omega_0$, and set Ω_0 as $\frac{\nu_f - 2}{\nu_f}$ multiplied by the sample variance-covariance matrix of

the factors in the training sample. If we partition Q_0 as

$$Q_0 = \begin{pmatrix} \underset{d_x \times d_x}{Q_{x,0}} & Q'_{g,0} \\ \underset{d_w \times d_x}{Q_{g,0}} & \underset{d_w \times d_w}{Q_{w,0}} \end{pmatrix}, \qquad (2.20)$$

then from properties of the inverse Wishart distribution it follows that the implied joint density on Ω_x, $\Omega_{w\cdot x}$, and $\boldsymbol{\gamma}$ is

$$\pi(\Omega_x, \Omega_{w\cdot x}, \boldsymbol{\gamma}) = \pi(\Omega_x)\pi(\Omega_{w\cdot x})\pi(\boldsymbol{\gamma}|\Omega_{w\cdot x}), \qquad (2.21)$$

where (on letting $\mathcal{IW}_d(\cdot|\rho, Q)$ stand for the d-dimensional inverse Wishart density with ρ degrees of freedom, and scale matrix Q)

$$\pi(\Omega_x) = \mathcal{IW}_{d_x}(\Omega_x|\rho_0 - d_w, Q_{x,0}), \qquad (2.22)$$

$$\pi(\Omega_{w\cdot x}) = \mathcal{IW}_{d_w}(\Omega_{w\cdot x}|\rho_0, Q_{w\cdot x,0}), \qquad (2.23)$$

$$\pi(\boldsymbol{\gamma}|\Omega_{w\cdot x}) = \mathcal{N}_q(\boldsymbol{\gamma}|\boldsymbol{\gamma}_0 = \text{vec}(Q_{g,0}Q_{x,0}^{-1}), B_{\gamma,0} = Q_{x,0}^{-1} \otimes \Omega_{w\cdot x}), \qquad (2.24)$$

where $Q_{w\cdot x,0} = Q_{w,0} - Q_{g,0}Q_{x,0}^{-1}Q'_{g,0}$, "vec" is the column vectorization operator, and $\otimes$ is the Kronecker product operator.

Thus, our prior density on $(\Omega_x, \Omega_{w\cdot x}, \boldsymbol{\gamma})$, for all the possible model specifications, is computed from the same pair of hyperparameters, (ρ_0, Ω_0). Naturally, the elements of Ω_0 have to be rearranged depending on the order of the x and w factors in a particular model. Apart from this qualification, what is important is that we only need to determine (ρ_0, Ω_0) once, regardless of the number of factor models that are being compared. This is not only computationally convenient but maintaining a comparable prior across models in this way is important in ensuring that differences in marginal likelihoods are due to the data evidence, and not due to differences in the prior on these parameters.

We note that our prior of $\boldsymbol{\gamma}$ is conditional on $\Omega_{w\cdot x}$. Because of this dependence, a Metropolis–Hastings step (Chib and Greenberg 1995) is needed in sampling the conditional posterior distribution of $\Omega_{w\cdot x}$. Nonetheless, as we show later, the M-H step can be implemented efficiently with an almost costless proposal distribution.

Table 2. Summary for the simulated factor distribution implied by the prior from different restricted factor models.

| | $\mathcal{M}_1$ | | $\mathcal{M}_2$ | | $\mathcal{M}_3$ | | $\mathcal{M}_4$ | |
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
MKT	−0.3322	0.3098	−0.2204	0.2235	−0.2157	0.2267	−0.2212	0.2249
SMB	−0.2287	0.2153	−0.2257	0.2229	−0.2124	0.2161	−0.2148	0.2140
HML	−0.1053	0.1032	−0.2199	0.2293	−0.1910	0.1986	−0.2106	0.2161
RMW	−0.0981	0.1037	−0.1510	0.1487	−0.2109	0.2067	−0.2044	0.2108
CMA	−0.0964	0.1018	−0.2240	0.2391	−0.2068	0.2185	−0.1984	0.2068
ME	−0.1912	0.1793	−0.2147	0.2116	−0.2157	0.2183	−0.2412	0.2441
ROE	−0.1600	0.1706	−0.2056	0.2104	−0.2130	0.2122	−0.2059	0.2150
IA	−0.0868	0.0924	−0.2009	0.2150	−0.2005	0.2104	−0.2055	0.2135
MOM	−0.1761	0.1882	−0.2933	0.3006	−0.5102	0.5137	−0.2099	0.2226
QMJ	−0.2022	0.2148	−0.1423	0.1459	−0.1974	0.2029	−0.2085	0.2133
LIQ	−0.1332	0.1405	−0.2135	0.2056	−0.3794	0.3820	−0.4623	0.4573
BAB	−0.0986	0.1031	−0.2480	0.2696	−0.3402	0.3573	−0.3983	0.4083
HMLD	−0.1270	0.1220	−0.2577	0.2665	−0.2110	0.2151	−0.2378	0.2364

NOTES: In $\mathcal{M}_1$, x-factor is QMJ. $\mathcal{M}_2$ is the q-factor model in Example 1. In $\mathcal{M}_3$, x-factors are MKT, SMB, RMW, CMA, and HMLD. In $\mathcal{M}_4$, x-factors are MKT, SMB, HML, RMW, ROE, IA, MOM, and QMJ. All the models have $v_f = 5$. The summaries are based on 100,000 draws. "Lower" and "upper" refer to the 0.025 and 0.975 quantiles of the simulated draws.

Given these assumptions, our prior density of $\boldsymbol{\theta}$ is now given by the expression

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\lambda}_x)\pi(\Omega_x)\pi(\Omega_{w \cdot x})\pi(\boldsymbol{\gamma}|\Omega_{w \cdot x}). \quad (2.25)$$

This prior is a key part of our procedure. As we document later in the article, careful testing in simulation experiments shows that our marginal likelihood procedure, based on this prior, is rather effective in finding the underlying true model. Whether another prior can perform as well is not relevant for our purposes. What matters is that we have found a prior construction that works well for this problem.

To illustrate the details of the prior set up, we continue our previous example assuming $v_f = 4$.

Example 1. (continued) The parameter list in the model is $\boldsymbol{\theta} = (\boldsymbol{\lambda}_x : 4 \times 1, \boldsymbol{\gamma} : 36 \times 1, \Omega_x : 4 \times 4, \Omega_{w \cdot x} : 9 \times 9)$. Under our prior setup, the prior mean on $\boldsymbol{\lambda}_x$ is the sample average of the factors MKT, ME, ROE, and IA in the training sample

$$\boldsymbol{\lambda}_0 = \begin{pmatrix} 0.0020 \\ -0.0002 \\ 0.0043 \\ 0.0067 \end{pmatrix}. \quad (2.26)$$

The prior mean on Ω is $\frac{v_f - 2}{v_f} = .5$ multiplied by the sample variance-covariance matrix of the factors MKT, ME, ROE, IA, SMB, HML, RMW, CMA, MOM, QMJ, LIQ, BAB, and HMLD (in that order) in the training sample.

To see what this prior implies about the (a priori) factor distribution, we use the sample the prior method of Chib and Ergashev (2009) and generate 100,000 draws from the prior distribution. Given each generated draw of the parameters, we simulate the 13 factors from the model. To compare the simulated factor distribution from the q-factor model with those of other models, we repeat the simulation for three alternative models. In the first alternative model x-factor is QMJ. In the second alternative model, x-factors are MKT, SMB, HMLD, RMW, and CMA. In the third alternative model, x-factors are

MKT, SMB, HML, RMW, ROE, IA, MOM, and QMJ. All the alternative models have $v_f = 4$. Table 2 reports the simulated quantiles from these models.

In general, the implied factor distributions from these models are comparable. There are small discrepancies across the models for some factors. However, these discrepancies are driven by the pricing restrictions instead of the priors. To see that, we simulate the factor distributions from the unrestricted versions of $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$, given by the equations from Equations (2.4)–(2.6). For the unrestricted factor model, the priors on $(\Gamma, \Omega_x, \Omega_{w \cdot x})$ are imposed in the same way as in the restricted models; the priors on $(\boldsymbol{\mu}_x, \boldsymbol{\mu}_w)$ are imposed as those on $\boldsymbol{\lambda}_x$. Table 3 reports the simulated quantiles from the unrestricted factor models. As expected, the simulated quantiles are the same across the different models.

3. POSTERIOR COMPUTATIONS

3.1. MCMC simulation algorithm

We now discuss the sampling of the posterior distribution by MCMC methods. In this simulation, we take advantage of the scale mixture of normals representation of the Student-t distribution and focus on the (augmented) posterior distribution given by

$$\pi(\boldsymbol{\theta}, \boldsymbol{\tau}_f, \boldsymbol{\tau}_\lambda | \boldsymbol{f}_{1:T}) \propto p(\boldsymbol{x}_{1:T} | \boldsymbol{\lambda}_x, \Omega_x, \boldsymbol{\tau}_f) p(\boldsymbol{w}_{1:T} | \boldsymbol{x}_{1:T}, \boldsymbol{\gamma}, \Omega_{w \cdot x}, \boldsymbol{\tau}_f)$$
$$\times \pi(\boldsymbol{\tau}_f)\pi(\boldsymbol{\tau}_\lambda)\pi(\boldsymbol{\lambda}_x | \boldsymbol{\tau}_\lambda)\pi(\Omega_x)\pi(\Omega_{w \cdot x})\pi(\boldsymbol{\gamma}|\Omega_{w \cdot x}), \quad (3.27)$$

where $\boldsymbol{f}_{1:T} = (\boldsymbol{f}_1, \dots, \boldsymbol{f}_T)$ denotes the sample data on the factors $\boldsymbol{f}_t$, for t running from January 1973 to December 2015, and $\boldsymbol{\tau}_f = (\tau_{f,1}, \dots, \tau_{f,T})$ are the latent scales for all T time periods with distribution given by

$$\pi(\boldsymbol{\tau}_f) = \prod_{t=1}^{T} \mathcal{G}\left(\tau_{f,t} \middle| \frac{v_f}{2}, \frac{v_f}{2}\right).$$

We sample this distribution by MCMC methods (for background on these methods, see Gelfand and Smith 1990; Tier-

Table 3. Summary for the simulated factor distribution implied by the prior from different unrestricted factor models.

	$\mathcal{M}_1$		$\mathcal{M}_2$		$\mathcal{M}_3$		$\mathcal{M}_4$	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
MKT	−0.2195	0.2213	−0.2217	0.2237	−0.2219	0.2226	−0.2212	0.2283
SMB	−0.2152	0.2148	−0.2168	0.2115	−0.2194	0.2112	−0.2147	0.2101
HML	−0.2069	0.2166	−0.2074	0.2139	−0.2069	0.2167	−0.2055	0.2144
RMW	−0.2040	0.2110	−0.2077	0.2086	−0.2060	0.2097	−0.2075	0.2068
CMA	−0.2001	0.2143	−0.2030	0.2149	−0.2052	0.2122	−0.2090	0.2140
ME	−0.2154	0.2103	−0.2101	0.2116	−0.2121	0.2142	−0.2130	0.2103
ROE	−0.2048	0.2123	−0.2058	0.2145	−0.2056	0.2152	−0.2052	0.2133
IA	−0.2032	0.2148	−0.2018	0.2128	−0.1994	0.2150	−0.2030	0.2195
MOM	−0.2123	0.2164	−0.2110	0.2157	−0.2127	0.2197	−0.2105	0.2235
QMJ	−0.2029	0.2156	−0.2030	0.2140	−0.2037	0.2148	−0.2014	0.2121
LIQ	−0.2079	0.2120	−0.2115	0.2145	−0.2131	0.2108	−0.2112	0.2162
BAB	−0.1994	0.2202	−0.2010	0.2234	−0.2050	0.2148	−0.2031	0.2188
HMLD	−0.2088	0.2147	−0.2085	0.2146	−0.2104	0.2171	−0.2136	0.2164

NOTES: In $\mathcal{M}_1$, x-factor is QMJ. $\mathcal{M}_2$ is the q-factor model in Example 1. In $\mathcal{M}_3$, x-factors are MKT, SMB, RMW, CMA, and HMLD. In $\mathcal{M}_4$, x-factors are MKT, SMB, HML, RMW, ROE, IA, MOM, and QMJ. All the models have $\nu_f = 5$. The summaries are based on 100,000 draws. "Lower" and "upper" refer to the 0.025 and 0.975 quantiles of the simulated draws.

ney 1994; Chib and Greenberg 1995; Chib 2001; Robert and Casella 2005), by iterating on the following steps.

Step 1: Sample each element $\tau_{f,t}$ of $\boldsymbol{\tau}_f$ from the Gamma densities

$$\pi(\tau_{f,t}|\boldsymbol{f}_t, \boldsymbol{\theta})$$
$$= \mathcal{G}\left(\tau_{f,t}\left|\frac{\nu_f + 1}{2}, \frac{\nu_f + \boldsymbol{\eta}_t' \text{blockdiag}(\Omega_x^{-1}, \Omega_{w \cdot x}^{-1})\boldsymbol{\eta}_t}{2}\right.\right),$$
$$t = 1, \ldots, T, \qquad (3.28)$$

where $\boldsymbol{\eta}_t' = \left((\boldsymbol{x}_t - \boldsymbol{\lambda}_x)', (\boldsymbol{w}_t - \Gamma\boldsymbol{x}_t)'\right)'$.

Step 2: Sample each element of $\boldsymbol{\tau}_\lambda$ from the Gamma densities

$$\pi(\tau_{\lambda,i}|\boldsymbol{f}_{1:T}, \boldsymbol{\lambda}_x, \boldsymbol{\gamma}) = \mathcal{G}\left(\tau_{\lambda,i}\left|\frac{\nu + 1}{2}, \frac{\nu + (\lambda_i - \lambda_{0,i})^2/s_\lambda}{2}\right.\right),$$
$$i = 1, \ldots, d_x. \qquad (3.29)$$

Step 3: Sample Ω_x from the density $\pi(\Omega_x|\boldsymbol{x}_{1:T}, \boldsymbol{\lambda}_x, \boldsymbol{\tau}_f)$, which can be derived by writing the data on the x-factors in matrix form as

$$\underbrace{X_{\tau_f}}_{T \times d_x} = \underbrace{\boldsymbol{1}_{\tau_f}}_{T \times 1} \boldsymbol{\lambda}_x' + \underbrace{E_x}_{T \times d_x}, \qquad (3.30)$$

where

$$X_{\tau_f} = \begin{pmatrix} \sqrt{\tau_{f,1}}\boldsymbol{x}_1' \\ \vdots \\ \sqrt{\tau_{f,T}}\boldsymbol{x}_T' \end{pmatrix}, \quad \boldsymbol{1}_{\tau_f} = \begin{pmatrix} \sqrt{\tau_{f,1}} \\ \vdots \\ \sqrt{\tau_{f,T}} \end{pmatrix} \qquad (3.31)$$

and the error matrix E_x follows the matrix normal distribution

$$\text{vec}(E_x) \sim \mathcal{N}_{T \times d_x}(\boldsymbol{0}, \Omega_x \otimes I_T).$$

On combining with the inverse Wishart prior of Ω_x in Equation (2.22), it can be seen that

$$\pi(\Omega_x|\boldsymbol{x}_{1:T}, \boldsymbol{\lambda}_x, \boldsymbol{\tau}_f) = \mathcal{IW}_{d_x}(\Omega_x|\rho_1 - d_w, Q_{x,1}), \qquad (3.32)$$

where

$$Q_{x,1} = Q_{x,0} + (X_{\tau_f} - \boldsymbol{1}_{\tau_f}\boldsymbol{\lambda}_x')'(X_{\tau_f} - \boldsymbol{1}_{\tau_f}\boldsymbol{\lambda}_x'), \qquad (3.33)$$

and $\rho_1 = \rho_0 + T$.

Step 4: Sample $\boldsymbol{\lambda}_x$ from the density $\pi(\boldsymbol{\lambda}_x|\boldsymbol{x}_{1:T}, \Omega_x, \boldsymbol{\tau}_f, \boldsymbol{\tau}_\lambda)$, which can be derived by starting with the vectorized form of the model in (3.30):

$$\text{vec}(X_{\tau_f}) = (I_{d_x} \otimes \boldsymbol{1}_{\tau_f})\boldsymbol{\lambda}_x + \text{vec}(E_x). \qquad (3.34)$$

Combining with the conditional normal prior of $\boldsymbol{\lambda}_x$ in Equation (2.16) and applying standard Bayesian updates we get that

$$\pi(\boldsymbol{\lambda}_x|\boldsymbol{x}_{1:T}, \Omega_x, \boldsymbol{\tau}_f, \boldsymbol{\tau}_\lambda) = \mathcal{N}_{d_x}(\hat{\boldsymbol{\lambda}}_x, B_{\lambda,T}), \qquad (3.35)$$

where

$$B_{\lambda,T} = (B_{\lambda,0}^{-1} + (\boldsymbol{1}_{\tau_f}'\boldsymbol{1}_{\tau_f}) \otimes \Omega_x^{-1})^{-1} \qquad (3.36)$$
$$\hat{\boldsymbol{\lambda}}_x = B_{\lambda,T}(B_{\lambda,0}^{-1}\boldsymbol{\lambda}_0 + ((\boldsymbol{1}_{\tau_f}'\boldsymbol{1}_{\tau_f}) \otimes \Omega_x^{-1})\hat{\boldsymbol{\lambda}}_{LS}), \qquad (3.37)$$

and

$$B_{\lambda,0} = \text{diag}(s_\lambda\tau_{\lambda,1}^{-1}, \ldots, s_\lambda\tau_{\lambda,d_x}^{-1}), \qquad (3.38)$$
$$\hat{\boldsymbol{\lambda}}_{LS} = ((\boldsymbol{1}_{\tau_f}'\boldsymbol{1}_{\tau_f})^{-1}\boldsymbol{1}_T'X)'. \qquad (3.39)$$

Step 5: Sample $\Omega_{w \cdot x}$ from the density

$$\pi(\Omega_{w \cdot x}|\boldsymbol{w}_{1:T}, \boldsymbol{x}_{1:T}, \boldsymbol{\gamma}, \boldsymbol{\tau}_f) \propto p(\boldsymbol{w}_{1:T}|\boldsymbol{x}_{1:T}, \boldsymbol{\gamma}, \Omega_{w \cdot x}, \boldsymbol{\tau}_f)\pi(\Omega_{w \cdot x})$$
$$\times \pi(\boldsymbol{\gamma}|\Omega_{w \cdot x}),$$

which is not in closed form because of the third term. Following (Chib and Greenberg 1995), however, we note that the first two terms on the right-hand side combine nicely and can be used as a proposal distribution in a Metropolis–Hastings (M-H) step. The M-H probability of move then just involves the third term. To combine those first two terms, express the matrix form of the data on the w-factors as

$$\underbrace{W_{\tau_f}}_{T \times d_w} = X_{\tau_f}\Gamma' + \underbrace{E_{w \cdot x}}_{T \times d_w}, \qquad (3.40)$$

where

$$W_{\tau_f} = \begin{pmatrix} \sqrt{\tau_{f,1}}w'_1 \\ \vdots \\ \sqrt{\tau_{f,T}}w'_T \end{pmatrix} \qquad (3.41)$$

and

$$\mathrm{vec}(E'_{w\cdot x}) \sim \mathcal{N}_{T \times d_w}(\mathbf{0}, I_T \otimes \Omega_{w\cdot x}).$$

Multiplying with the inverse Wishart prior of $\Omega_{w\cdot x}$ in Equation (2.23), the proposal distribution becomes

$$\mathcal{IW}_{d_w}(\rho_1, Q_{w\cdot x,1}), \qquad (3.42)$$

where

$$Q_{w\cdot x,1} = Q_{w\cdot x,0} + (W_{\tau_f} - X_{\tau_f}\Gamma')'(W_{\tau_f} - X_{\tau_f}\Gamma'). \qquad (3.43)$$

Then, in the M-H step, we propose $\Omega_{w\cdot x}$ from the latter distribution, which we accept with probability given by

$$\alpha(\Omega_{w\cdot x}, \Omega_{w\cdot x} | w_{1:T}, x_{1:T}, \gamma, \tau_f)$$
$$= \min\left\{1, \frac{\mathcal{N}_q(\gamma | \mathrm{vec}(Q_{g,0}Q_{x,0}^{-1}), Q_{x,0}^{-1} \otimes \Omega_{w\cdot x})}{\mathcal{N}_q(\gamma | \mathrm{vec}(Q_{g,0}Q_{x,0}^{-1}), Q_{x,0}^{-1} \otimes \Omega_{w\cdot x})}\right\}. \qquad (3.44)$$

We stay at the current value $\Omega_{w\cdot x}$ if the proposed value is not accepted.

Step 6: *Sample* γ *from the distribution* $p(\gamma | w_{1:T}, x_{1:T}, \Omega_{w\cdot x}, \tau_f)$, which can be derived starting with the row-vectorized form of the model in Equation (3.40),

$$\mathrm{vec}(W'_{\tau_f}) = (X_{\tau_f} \otimes I_{d_w})\gamma + \mathrm{vec}(E'_{w\cdot x}). \qquad (3.45)$$

Combining with the conditional normal prior of γ in Equation (2.24) and applying standard Bayesian updates we get that

$$\pi(\gamma | w_{1:T}, x_{1:T}, \Omega_{w\cdot x}, \tau_f) = \mathcal{N}_q(\hat{\gamma}, B_{\gamma,T}). \qquad (3.46)$$

where

$$B_{\gamma,T} = (B_{\gamma,0}^{-1} + (X'_{\tau_f}X_{\tau_f}) \otimes \Omega_{w\cdot x}^{-1})^{-1}, \qquad (3.47)$$

$$\hat{\gamma} = B_{\gamma,T}(B_{\gamma,0}^{-1}\gamma_0 + ((X'_{\tau_f}X_{\tau_f}) \otimes \Omega_{w\cdot x}^{-1})\hat{\gamma}_{\mathrm{LS}}), \qquad (3.48)$$

and

$$\hat{\gamma}_{\mathrm{LS}} = \mathrm{vec}(((X'_{\tau_f}X_{\tau_f})^{-1}X'_{\tau_f}W_{\tau_f})'). \qquad (3.49)$$

We repeat these sampling steps a large number of times, say 80,000 times and discard the burn-in draws, say the first 40,000. The remaining 40,000 draws, ignoring the draws on the latent scales, are samples from the posterior distribution $\pi(\theta | f_{1:T})$.

Example 1. (continued) Applying our fitting algorithm to the q-factor model (with $\nu_f = 4$), the marginal posterior distributions of $\lambda_x, \gamma, \Omega_x, \Omega_{w\cdot x}$ are summarized in Table 4. The M-H acceptance rate is around 82% and the inefficiency factor, the ratio of the numerical variance of the mean to the variance of the mean assuming independent draws are small (indicating that the draws, although serially correlated, are essentially independent). The example showcases the simulation efficiency of our MCMC algorithm.

3.2. Marginal Likelihood Computation

Our main task in the empirical study is not just estimating a single model but also comparing multiple models. In Bayesian analysis, we use the posterior odds of models given the observed data to compare alternative models. The posterior odds reflect how data favor each competing model. Let $p_j = \Pr(\mathcal{M}_j)$ denote the prior probability for model $\mathcal{M}_j$. The posterior odds between model $\mathcal{M}_i$ and $\mathcal{M}_k$ are

$$\frac{\Pr(\mathcal{M}_i | f_{1:T})}{\Pr(\mathcal{M}_k | f_{1:T})} = \frac{p_i}{p_k}\frac{m(f_{1:T} | \mathcal{M}_i)}{m(f_{1:T} | \mathcal{M}_k)}, \qquad (3.50)$$

where the marginal likelihood is defined as the integral of the likelihood function with respect to the prior density,

$$m(f_{1:T} | \mathcal{M}_j) = \int_{\Theta_j} p(f_{1:T} | \theta, \mathcal{M}_j)\pi(\theta | \mathcal{M}_j)\,\mathrm{d}\theta \qquad (3.51)$$

and the integration is over the parameter space

$$\Theta = \mathbb{R}^{d_x} \times \mathbb{R}^q \times \{\text{set of } d_x \times d_x \text{ pd matrices}\}$$
$$\times \{\text{set of } d_w \times d_w \text{ pd matrices}\},$$

where "pd" stands for positive-definite. If the prior odds on the models are one, which we are going to assume here, the posterior odds are equal to the ratio of marginal likelihoods. Therefore, the central piece of model performance is the marginal likelihood. Marginal likelihoods automatically penalize models based on complexity, ensuring that more complex models will not rank higher merely because the more flexible model is capable of fitting the noise in the data. Moreover, model selection based on the comparison of Bayes factors or marginal likelihoods has attractive asymptotic properties. If the true model is among the candidates under consideration, the highest marginal likelihood model will select it with probability approaching one in the limit; if it is not among the candidates (the likely case), the highest marginal likelihood model will select the model that is closest to the true model in the Kullback–Leibler distance (see, e.g., Chib, Shin, and Simoni 2018). We lean on this latter property to help us ascertain, from the comparison of all subset models, which group of factors are better supported by the data as being in the SDF.

To estimate the marginal likelihood of each contending model, we employ the Chib (1995) method which starts with the convenient expression of the log-marginal likelihood

$$\ln m(f_{1:T} | \mathcal{M}_j) = \ln \pi(\theta^* | \mathcal{M}_j) + \ln p(f_{1:T} | \mathcal{M}_j, \theta^*)$$
$$- \ln \pi(\theta^* | \mathcal{M}_j, f_{1:T}), \qquad (3.52)$$

where $\theta^* = (\lambda_x^*, \gamma^*, \Omega_x^*, \Omega_{w\cdot x}^*)$ is some chosen point, say the posterior mean. In this expression, the prior and likelihood ordinates can be found analytically. As for the third term, the posterior ordinate, suppress the model index, and use a marginal-conditional decomposition to write

$$\ln \pi(\theta^* | f_{1:T}) = \ln \pi(\Omega_{w\cdot x}^* | f_{1:T}) + \ln \pi(\lambda_x^*, \gamma^* | f_{1:T}, \Omega_{w\cdot x}^*)$$
$$+ \ln \pi(\Omega_x^* | f_{1:T}, \lambda_x^*, \gamma^*, \Omega_{w\cdot x}^*), \qquad (3.53)$$

Table 4. Posterior summary for selected parameters in Example 1 (q-factor model).

Param.	Mean	sd	Lower	Median	Upper	Ineff.
λ_1	6.93×10^{-3}	1.65×10^{-3}	6.91×10^{-3}	3.72×10^{-3}	1.02×10^{-2}	1.28
λ_2	2.51×10^{-3}	1.05×10^{-3}	2.51×10^{-3}	4.19×10^{-4}	4.57×10^{-3}	1.26
λ_3	6.91×10^{-3}	8.00×10^{-4}	6.91×10^{-3}	5.32×10^{-3}	8.44×10^{-3}	1.28
λ_4	3.70×10^{-3}	6.62×10^{-4}	3.70×10^{-3}	2.38×10^{-3}	4.98×10^{-3}	1.28
$\Gamma_{1,1}$	-1.06×10^{-2}	6.89×10^{-3}	-1.07×10^{-2}	-2.39×10^{-2}	2.92×10^{-3}	1.23
$\Gamma_{2,1}$	-8.81×10^{-2}	2.10×10^{-2}	-8.79×10^{-2}	-1.30×10^{-1}	-4.72×10^{-2}	1.22
$\Gamma_{3,1}$	-4.37×10^{-2}	1.49×10^{-2}	-4.37×10^{-2}	-7.28×10^{-2}	-1.45×10^{-2}	1.20
$\Gamma_{4,1}$	-3.01×10^{-2}	7.58×10^{-3}	-3.01×10^{-2}	-4.51×10^{-2}	-1.55×10^{-2}	1.22
$\Gamma_{5,1}$	-2.02×10^{-3}	3.30×10^{-2}	-1.86×10^{-3}	-6.81×10^{-2}	6.16×10^{-2}	1.22
$\Gamma_{6,1}$	-1.73×10^{-1}	1.50×10^{-2}	-1.73×10^{-1}	-2.03×10^{-1}	-1.44×10^{-1}	1.23
$\Gamma_{7,1}$	-2.39×10^{-2}	3.89×10^{-2}	-2.39×10^{-2}	-1.01×10^{-1}	5.27×10^{-2}	1.28
$\Gamma_{8,1}$	9.08×10^{-2}	3.13×10^{-2}	9.08×10^{-2}	2.89×10^{-2}	1.52×10^{-1}	1.33
$\Gamma_{9,1}$	-4.47×10^{-2}	2.35×10^{-2}	-4.45×10^{-2}	-9.08×10^{-2}	1.10×10^{-3}	1.21
$\Omega_{x,1,1}$	1.16×10^{-3}	8.29×10^{-5}	1.16×10^{-3}	1.01×10^{-3}	1.33×10^{-3}	2.54
$\Omega_{x,2,2}$	4.81×10^{-4}	3.48×10^{-5}	4.80×10^{-4}	4.17×10^{-4}	5.52×10^{-4}	2.40
$\Omega_{x,3,3}$	2.73×10^{-4}	1.95×10^{-5}	2.72×10^{-4}	2.37×10^{-4}	3.14×10^{-4}	2.54
$\Omega_{x,4,4}$	1.92×10^{-4}	1.35×10^{-5}	1.91×10^{-4}	1.66×10^{-4}	2.20×10^{-4}	2.47
$\Omega_{w \cdot x,1,1}$	2.37×10^{-5}	1.68×10^{-6}	2.36×10^{-5}	2.05×10^{-5}	2.71×10^{-5}	3.38
$\Omega_{w \cdot x,2,2}$	2.26×10^{-4}	1.60×10^{-5}	2.25×10^{-4}	1.96×10^{-4}	2.59×10^{-4}	2.96
$\Omega_{w \cdot x,3,3}$	1.11×10^{-4}	7.86×10^{-6}	1.11×10^{-4}	9.64×10^{-5}	1.27×10^{-4}	3.10
$\Omega_{w \cdot x,4,4}$	2.90×10^{-5}	2.04×10^{-6}	2.89×10^{-5}	2.53×10^{-5}	3.32×10^{-5}	3.51
$\Omega_{w \cdot x,5,5}$	5.36×10^{-4}	3.82×10^{-5}	5.35×10^{-4}	4.66×10^{-4}	6.17×10^{-4}	3.39
$\Omega_{w \cdot x,6,6}$	1.13×10^{-4}	8.09×10^{-6}	1.13×10^{-4}	9.79×10^{-5}	1.30×10^{-4}	3.54
$\Omega_{w \cdot x,7,7}$	7.39×10^{-4}	5.35×10^{-5}	7.37×10^{-4}	6.43×10^{-4}	8.49×10^{-4}	3.21
$\Omega_{w \cdot x,8,8}$	4.81×10^{-4}	3.41×10^{-5}	4.80×10^{-4}	4.16×10^{-4}	5.50×10^{-4}	3.36
$\Omega_{w \cdot x,9,9}$	2.78×10^{-4}	1.99×10^{-5}	2.77×10^{-4}	2.42×10^{-4}	3.20×10^{-4}	3.43

NOTES: The summaries are based on 10,000 MCMC draws beyond a burn-in of 40,000. "Lower" and "upper" refer to the 0.025 and 0.975 quantiles of the simulated draws, respectively, and "ineff" to the inefficiency factor, the ratio of the numerical variance of the mean to the variance of the mean assuming independent draws.

Now appealing to the approach of Chib and Jeliazkov (2001) we have that

$$
\pi(\Omega^*_{w \cdot x}|f_{1:T})
$$
$$
= \frac{\mathbb{E}_1\{\alpha(\Omega_{w \cdot x}, \Omega^*_{w \cdot x}|f_{1:T}, \lambda_x, \gamma, \tau_f)\mathcal{IW}_{d_w}(\Omega^*_{w \cdot x}|\rho_1, Q_{w \cdot x,1})\}}{\mathbb{E}_2\{\alpha(\Omega^*_{w \cdot x}, \Omega_{w \cdot x}|f_{1:T}, \lambda_x, \gamma, \tau_f)\}}.
$$
(3.54)

where $\mathbb{E}_1$ denotes the expectation with respect to the posterior distribution $\pi(\theta|f_{1:T})$, and $\mathbb{E}_2$ denotes the expectation with respect to the distribution

$$
\pi(\lambda_x, \gamma|f_{1:T}, \Omega^*_{w \cdot x}) \times \mathcal{IW}_{d_w}(\Omega_{w \cdot x}|\rho_1, Q_{w \cdot x,1}) \quad (3.55)
$$

We can calculate the former expectation by Monte Carlo with the draws on θ from the full MCMC run. For the latter expectation, we perform a reduced MCMC run in which $\Omega_{w \cdot x}$ is fixed at $\Omega^*_{w \cdot x}$ and the remaining blocks of parameters are sampled as before. This reduced MCMC run gives rise to the draws

$$
\left\{\lambda_x^{(j)}, \gamma^{(j)}, \Omega_x^{(j)}, \tau_\lambda^{(j)}, \tau_f^{(j)}\right\}. \quad (3.56)
$$

For each of these draws, we sample $\Omega_{w \cdot x}^{(j)}$ from the proposal distribution

$$
\mathcal{IW}_{d_w}(\rho_1, Q_{w \cdot x,1}^{(j)}), \quad (3.57)
$$

where $Q_{w \cdot x,1}^{(j)}$ are conditional posterior quantities computed at $(\gamma^{(j)}, \tau_f^{(j)})$.

The second ordinate, $\pi(\lambda_x^*, \gamma^*|f_{1:T}, \Omega^*_{w \cdot x})$, is estimated from the output of the previous reduced run as

$$
\hat{\pi}(\lambda_x^*, \gamma^*|f_{1:T}, \Omega^*_{w \cdot x})
$$
$$
= \frac{1}{J}\sum_{j=1}^{J}\mathcal{N}_{d_x}(\lambda_x^*|\hat{\lambda}_x^{(j)}, B_{\lambda,T}^{(j)})\mathcal{N}_q(\gamma^*|\hat{\gamma}^{(j)}, B_{\gamma,T}^{(j)}), \quad (3.58)
$$

where $\hat{\lambda}_x^{(j)}$, $B_{\lambda,T}^{(j)}$, $\hat{\gamma}^{(j)}$ and $B_{\gamma,T}^{(j)}$ are computed conditional on $(\Omega_x^{(j)}, \Omega^*_{w \cdot x}, \tau_\lambda^{(j)}, \tau_f^{(j)})$.

Finally, $\pi(\Omega_x^*|f_{1:T}, \lambda_x^*, \gamma^*, \Omega^*_{w \cdot x})$ is from the output of another reduced MCMC run. Fixing $(\lambda_x, \gamma, \Omega_{w \cdot x})$ at $(\lambda_x^*, \gamma^*, \Omega^*_{w \cdot x})$, we sample the remaining blocks of parameters. If we let

$$
\left\{\Omega_x^{(l)}, \tau_f^{(l)}\right\}_{l=1}^{L}
$$

denote the draws in this second reduced run then our estimate of the final ordinate is given by

$$
\hat{\pi}(\Omega_x^*|f_{1:T}, \lambda_x^*, \gamma^*, \Omega^*_{w \cdot x}) = \frac{1}{L}\sum_{l=1}^{L}\mathcal{IW}_{d_x}(\Omega_x^*|\rho_1 - d_w, Q_{x,1}^{(l)}),
$$
(3.59)

Table 5. Log-marginal likelihoods of six common models.

q_5	FF5$_5$	CAPM_5	q_∞	FF5$_\infty$	CAPM_∞
17,569.61	17,538.61	17,511.86	16,809.67	16,790.97	16,782.65

where $Q_{x,1}^{(l)}$ is computed conditional on $(\lambda_x^*, \tau_f^{(l)})$.

Though the marginal likelihood in our framework does not have a closed form, the numerical computations follow in a straightforward manner. In addition, an R package for implementing all the computations is available on request.

Example 1. (continued) Consider the q-factor model (with $v_f = 5$). Let θ^* denote the average of 10,000 MCMC draws beyond a burn-in of 40,000. Let the number of iterations in both reduced runs be 10,000. The estimated log-marginal likelihood is 17,555.31. Suppose we want to compare this q-factor model with five alternative models: the CAPM, in which the only x-factor is MKT, with $v_f = 5$ and ∞, the Fama-French five-factor model, in which x-factors are $\{MKT, SMB, HML, RMW, CMA\}$, with $v_f = 5$ and ∞, and the q-factor model with $v_f = \infty$. The log-marginal likelihood for these models are given in Table 5.

4.　SIMULATION EXPERIMENTS

In our framework, a different starting collection of factors that go into x_t leads to a different set of restrictions on the factor model and, hence, a different model. Simultaneously, confronting each of these possible restricted factor models with the data, and comparing them in terms of the Bayesian marginal likelihood criterion, provides the means to identify the x-factors in the SDF. In this section, we use simulations to assess how well the framework performs in identifying the x-factors. The simulation studies are designed to mimic the real-world features of the factors that are typically employed in this setting. We suppose that there are 11 factors and use our framework to calculate the marginal likelihood of all possible models that cover a grid of $v_f = 4$, 8, 16, 32, and ∞ degrees of freedom, for samples of size $T = 600$, 1200, and 2400, and for 100 replications of the data for each sample size. We could use a finer grid of v_f values (as we do in our real data example) but in general it is not possible to discriminate between close degrees of freedom, unless the sample size is large. The results show that our method is capable of locating the true model with remarkable accuracy. This accuracy increases with sample size as per the asymptotic theory of the marginal likelihood.

4.1.　Simulation Design

Our simulations proceed as follows.

1. We assume an eleven-factor world in which the factors are matched to the following 11 factors, MKT, SMB, HML, RMW, CMA, ME, ROE, IA, MOM, QMJ, and LIQ.
2. We then make an assumption about the true pricing factors. In this setting, for a given factor distributional form, there

Table 6. Maximum likelihood estimates of b in $M_t = 1 - \frac{v_f - 2}{v_f} b'(x_t - \mu_x)$, with standard errors in parentheses.

Factor	MKT	ME	ROE	IA
$\hat{b}$	9.61	6.34	28.66	30.64
	(1.69)	(2.43)	(3.45)	(4.22)

NOTES: The estimation is for the factor model that has MKT, ME, ROE, IA as the x-factors and SMB, HML, RMW, CMA, MOM, QMJ, and LIQ as the w-factors, and has $v_f = 4$, based on the monthly data from January 1968 to December 2015.

are $2^{11} - 1 = 2047$ possible true models depending on the assumption made about the collection of factors that go into x_t. For instance, suppose that in the true model x_t consists of MKT, ME, ROE, and IA and the factor distribution is a student-t with $v_f = 4$. Our aim is to generate data from this model and to then confront the $(2^{11} - 1) \times 5 = 10,235$ possible models over the grid of $v_f = 4$, 8, 16, 32, and ∞ to these data to see if the marginal likelihoods correctly pick this true model. Of course, in order to generate data from this true model we need 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 on the factors resemble the real data. A key point is that we need to ensure that this is a valid model for the purpose of generating our data. By valid model, we mean a model in which the fitted SDF suggests that each assumed x-factors is statistically significant. In other words, we need to check that in the expression of $M_t = 1 - \frac{v_f - 2}{v_f} \lambda_x' \Omega_x^{-1}(x_t - \mu_x)$, the fitted values of $b = \Omega_x^{-1} \lambda_x$ are each significant. Otherwise, the maintained assumption that the factors $\{MKT, ME, ROE, IA\}$ are the pricing factors would counter to the evidence and data generated from such a data-generating process (DGP) would lead to misleading model comparisons. In the following example, we show the maintained assumption is valid for generating the simulated data.

Example 2. Consider the factor model that has MKT, ME, ROE, and IA as the x-factors and SMB, HML, RMW, CMA, MOM, QMJ, and LIQ as w-factors. The factor distribution is assumed to be a student-t with $v_f = 4$. To check whether the model is valid for generating the simulated data, we find the estimate of b by maximizing the log-likelihood function and calculate the variance-covariance matrix of the estimate as the negative inverse of the Hessian matrix of the likelihood function at the maximum likelihood estimate. The estimates and the associated standard errors are reported in Table 6. As we can see from the table, each element of b is significant at the level of 0.05.

We select three models that are valid for generating the simulated data. The composition of their x_t are listed in the first column of Table 7 and their factor distributional forms are listed in the second column of the same table.

3. Then, we simulate data from each of these three true models. As mentioned above, the true parameter values are those obtained from the maximum likelihood fitting of the models

Table 7. Percentage of times the true model is selected, from within an enumerated model space of 10,235 models, over 100 simulated datasets. There are 11 factors in the candidate set.

x-factors	v_f	$T = 600$	$T = 1200$	$T = 2400$
MKT, ME, ROE, IA	3.5	52 (5322)	67 (6857)	72 (7369)
MKT	∞	70 (7165)	96 (9826)	97 (9928)
MKT, HML	4	78 (7983)	85 (8700)	95 (9723)

NOTES: The first and second columns are the x-factors and the degrees of freedom used in the true model. Given each true model, we consider three time-series sizes, $T = 600$ (column 3), $T = 1200$ (column 4) and $T = 2400$ (column 5). For each dataset, we select the model that has the highest marginal likelihood. The numbers in parentheses are the ratio of the selection percentage divided by 1/10,235, the prior probability of each model.

to the monthly data on the eleven factors which runs from January 1968 to December 2015.

Example 2. (continued) The ML estimates of the model over the actual data are

$$\hat{\lambda}_x = \begin{pmatrix} 0.0065 \\ 0.0021 \\ 0.0069 \\ 0.0038 \end{pmatrix},$$

$$\hat{\Omega}_x = \begin{pmatrix} 12.76 & 2.00 & -0.64 & -1.68 \\ 2.00 & 5.21 & -0.67 & -0.38 \\ -0.64 & -0.67 & 3.07 & -0.28 \\ -1.68 & -0.38 & -0.28 & 2.10 \end{pmatrix} \times 10^{-4},$$

(4.60)

$$\hat{\Gamma} = \begin{pmatrix} -0.00 & 0.96 & -0.10 & -0.07 \\ -0.08 & -0.01 & -0.27 & 0.86 \\ -0.04 & -0.11 & 0.50 & -0.15 \\ -0.03 & -0.00 & -0.16 & 0.93 \\ 0.00 & 0.13 & 0.85 & 0.26 \\ -0.18 & -0.20 & 0.56 & -0.07 \\ -0.04 & -0.06 & 0.11 & 0.07 \end{pmatrix},$$

(4.61)

$$\hat{\Omega}_{w \cdot x} = \begin{pmatrix} 0.26 & -0.14 & 0.12 & -0.02 & 0.08 & 0.09 & 0.03 \\ -0.14 & 2.38 & 0.12 & 0.26 & -0.21 & -0.33 & 0.23 \\ 0.12 & 0.12 & 1.16 & -0.01 & -0.47 & 0.62 & 0.25 \\ -0.02 & 0.26 & -0.01 & 0.32 & 0.03 & 0.03 & 0.00 \\ 0.08 & -0.21 & -0.47 & 0.03 & 5.73 & -0.22 & 0.26 \\ 0.09 & -0.33 & 0.62 & 0.03 & -0.22 & 1.19 & 0.22 \\ 0.03 & 0.23 & 0.25 & 0.00 & 0.26 & 0.22 & 7.72 \end{pmatrix}$$
$$\times 10^{-4}.$$

(4.62)

Given these parameter values, we then simulate the factors for each t from the assumed model.

4. Given a particular DGP and a particular sample from that DGP, we calculate the marginal likelihood of the 10,235 possible factor models on a grid of 4, 8, 16, 32, and ∞ degrees of freedom, and see if the highest marginal likelihood corresponds to the true model that generated the data.
5. We repeat this exercise 100 times for sample sizes of $T = 600$, 1200, and 2400 and for each of the three true DGP's.

4.2. Results

Note that for the first true DGP, the candidate models exclude the true degrees of freedom. The true number of degrees of freedom is 3.5 whereas the candidate degrees of freedom are 4, 8, 16, 32, and ∞. In Table 7, we report the percentages of times in 100 replications that the model with the correct x-factors and the closest degrees of freedom (which is 4) to the true degrees of freedom is selected for sample sizes of size $T = 600$, 1200, and 2400 based on the marginal likelihood. For the other two designs, Table 7 reports the percentage of times in 100 replications of data that the true model is selected by the marginal likelihood comparison for each of the three samples sizes. The results show that our method locates the true model (a needle in the haystack of possible models) with remarkable accuracy. This accuracy increases with sample size as per the asymptotic theory of the marginal likelihood. The results also confirm the ability of our framework to find the model closest to the true model (when the true model is not in the pool of possible models).

5. EMPIRICAL RESULTS

We now turn to the application of our framework to the actual data on 13 common risk factors. These are the Fama and French (1993, 2015) factors: excess market return (MKT), size (SMB), value (HML), profitability (RMW) and investment (CMA); the Hou, Xue, and Zhang (2015) q-factors: size (ME), profitability (ROE) and investment (IA); the Carhart (1997) momentum (MOM) factor; 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 (HMLD) factor proposed by Asness and Frazzini (2013). In this real-data analysis, we use a training sample that runs from January 1968 to December 1972 to form our prior distribution, and the period from January 1973 to December 2015 as the estimation sample. In fitting the different Student-t models, we use a grid of 14 values of v_f, composed of 11 values from 2.5 to 8 in increments of 0.5, and the values 16, 32, and inf. This leads to a collection of $(2^{13} - 1) \times 14 = 114,674$ models under contention. Each of these 114,674 contending models is estimated from 40,000 draws of the MCMC algorithm, collected after an initial burn-in period of 40,000 iterations. The subsequent reduced runs for the marginal likelihood estimation are of length 10,000.

Our analysis shows that the model with the highest marginal likelihood, which is listed in the second row of Table 8, is a Student-t distributed factor model with 5 degrees of freedom. It has a log-marginal likelihood value of 17,613.06 and includes eight x-factors. Three of the Fama-French 5 factors, namely the market excess return MKT, the size factor SMB, and the profitability factor RMW are in this most preferred model. The profitability factor ROE from the q-factor model is also in this best model. Finally, the other x-factors in this model are the momentum factor MOM, the quality factor QMJ, the betting against beta factor BAB, and the value factor HMLD.

Table 8 contains the best 4 models according to the log-marginal likelihood criterion. The best model is a Student-t

Table 8. Top 4 marginal likelihood models among all 114,674 models considered. Column 2 lists the *x*-factors in the model.

log(m)	*x*-factors	v_f
17,613.06	MKT, SMB, RMW, ROE, MOM, QMJ, BAB, HMLD	5
17,612.89	MKT, SMB, RMW, ROE, MOM, QMJ, BAB, HMLD	5.5
17,612.55	MKT, SMB, HML, RMW, ROE, MOM, QMJ, BAB, HMLD	5.5
17,612.38	MKT, SMB, HML, RMW, ROE, MOM, QMJ, BAB, HMLD	5

NOTES: Column 3 reports in each model specification what the number of degrees of freedom is in the factor process. The best model is the one with the highest marginal likelihood as reported in the second row. The data sample consists of monthly observations from January 1968 to December 2015. The first 60 observations form the training sample and the remaining 516 observations are the estimation sample.

Table 9. The highest log-marginal likelihood model vs. the best Gaussian model.

log(m)	*x*-factors	v_f
17,613.06	MKT, SMB, RMW, ROE, MOM, QMJ, BAB, HMLD	5
16,836.76	MKT, SMB, HML, RMW, ROE, MOM, QMJ, HMLD	∞

NOTES: The data sample consists of monthly observations from January 1968 to December 2015. The first 60 observations form the training sample and the remaining 516 observations are the estimation sample.

model with 5 degrees of freedom and 8 *x*-factors. The second best model has 5.5 degrees of freedom and the same pricing factors. The next two best models have 9 pricing factors but because the log-marginal likelihood values are within .6, according to Jeffreys' scale (Jeffreys 1961), the models are roughly equivalent.

Several findings can be observed in Table 8. First, that the factors MKT, ROE, MOM, QMJ, BAB, and HMLD are in each of the top 4 models. This suggests that these factors are jointly important drivers of the cross-section of expected returns. Our results show that size matters but it is in general better reflected in the SMB factor. Second, the factors CMA, IA, and LIQ do not appear in the collection of *x*-factors. Third, somewhat surprisingly, the two versions of profitability, that is, the RMW factor from the Fama-French five-factor model, and the ROE factor from the q-factor model, often appear jointly, despite the fact that both are measures of similar risks.

Another important point to note is that the Student-*t* distribution is much better supported by these data than the Gaussian distribution. The difference between the best overall model and the best Gaussian model is more than 700 on the log scale, as shown in Table 9. Moreover, the *x*-factors in the best Gaussian are also different. For example, the BAB factor, is not in the best Gaussian model. Given the large difference in log-marginal likelihoods, one should be cautious in interpreting results obtained under the normality assumption.

6. CONCLUDING REMARKS

In this article, we develop a Bayesian framework to isolate risk factors, from a given collection of potential risk factors, to explain the cross-section of expected returns. The framework relies on two important assumptions, that factors are traded portfolio excess returns or return spreads and the SDF is linear in the factors. The framework uses an enumerative model search strategy to jointly compare all possible factor pricing models and finds the one that is best supported by the data in terms of marginal likelihoods and posterior model probabilities. One aspect of this enumerative strategy, that we do not emphasize here, but is a potent justification for our approach, is that it avoids postmodel selection biases (in the frequentist sense) because all the models are enumerated before the data are encountered. Our construction of the proper prior, based on a training sample, with independent Student-*t* distributions for some location parameters, and a common inverse Wishart distribution for the covariance parameters, can act as template for prior construction in other similar problems with large-dimensional parameter spaces, and large-dimensional model spaces. We also supply a simulation-efficient Bayesian MCMC method for model estimation and marginal likelihood computation. The framework overall is self-contained and can be used with minimum user intervention.

Our article also highlights the importance of using multivariate Student-t distributions to model the fat tails in the factor data. In our empirical study, we show that the multivariate Student-*t* distribution is better supported than the multivariate normal distribution. Moreover, the best model under the normality assumption is quite different from the Student-*t* model that is most supported overall. This suggests caution in interpreting results that are predicated on the normality assumption. As the editor of this paper has pointed out, it would also be useful to allow for the possibility of multivariate stochastic volatility. Such a model extension is likely to lead to new findings about the risk factors. However, incorporating multivariate stochastic volatility in our model is a rather significant extension that requires a considerable elaboration and modification of our model and estimation framework. We intend to describe this extension in future work.

We conclude by noting that with the enormous continuing increase in computing power, our enumerative model search strategy can be extended to even larger collections of potential risk-factors than the 13 leading factors we have considered in this article. For instance, if we assume sparsity and suppose that at most 10 factors can be in the SDF, our Bayesian enumerative strategy, with a starting collection of 50 potential risk factors, would require the comparison of 13 billion models, which is within the reach of our method with current parallel and cloud computing resources. In general, as it turns out, the limiting factor to the enumerative model comparison strategy outlined here is not computing, but rather the lack of serious potential risk factors. Nonetheless, in ongoing work we are in the process of assembling a dataset with several additional factors, with the aim of applying the method of this article to that setting. We are hopeful that the proposed method will have a large impact on the practical discovery of risk-factors, with a consequent large impact on the investment and wealth management industries.

ACKNOWLEDGMENTS

REFERENCES

Affleck-Graves, J., and Mcdonald, B. (1989), "Nonnormalities and Tests of Asset Pricing Theories," *The Journal of Finance*, 44, 889–908. [773]

Ang, A. (2014), *Asset Management: A Systematic Approach to Factor Investing*, Financial Management Association Survey and Synthesis, New York, NY: Oxford University Press. [772]

Asness, C. and Frazzini, A. (2013), "The Devil in HML's Details," *The Journal of Portfolio Management*, 39, 49–68. [772,781]

Asness, C. S., Frazzini, A., and Pedersen, L. H. (2014), "Quality Minus Junk," Working paper. [772,781]

Avramov, D., and Chao, J. (2006), "An Exact Bayes Test of Asset Pricing Models with Application to International Markets," *The Journal of Business*, 79, 293–324. [773]

Barillas, F., and Shanken, J. (2017), "Which Alpha?" *The Review of Financial Studies*, 30, 1316–1338. [773]

—— (2018), "Comparing Asset Pricing Models," *The Journal of Finance*, 73, 715–754. [773]

Carhart, M. M. (1997), "On Persistence in Mutual Fund Performance," *The Journal of Finance*, 52, 57–82. [772,781]

Chib, S. (1995), "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321. [772,778]

—— (2001), "Markov Chain Monte Carlo methods: Computation and Inference," in *Handbook of Econometrics*, eds. Heckman, J. J. and Leamer, E., Amsterdam: North-Holland, pp. 3569–3649. [777]

Chib, S., and Ergashev, B. (2009), "Analysis of Multifactor Affine Yield Curve Models," *Journal of the American Statistical Association*, 104, 1324–1337. [772,776]

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis–Hastings Algorithm," *The American Statistician*, 49, 327–335. [775,777]

Chib, S., and Jeliazkov, I. (2001), "Marginal Likelihood From the Metropolis–Hastings Output," *Journal of the American Statistical Association*, 96, 270–281. [772,779]

Chib, S., Shin, M., and Simoni, A. (2018), "Bayesian Estimation and Comparison of Moment Condition Models," *Journal of the American Statistical Association*, 113, 1656–1668. [772,778]

Chib, S., Zeng, X., and Zhao, L. (2018), "On Comparing Asset Pricing Models," *The Journal of Finance*, in press. [773]

Cochrane, J. H. (2009), *Asset Pricing* (revised ed.). Princeton, NJ: Princeton University Press. [771]

Dufour, J.-M., Khalaf, L., and Beaulieu, M.-C. (2003), "Exact Skewness-Kurtosis Tests for Multivariate Normality and Goodness-of-Fit in Multivariate Regressions with Application to Asset Pricing Models," *Oxford Bulletin of Economics and Statistics*, 65, 891–906. [773]

Fama, E. F. (1965), "The Behavior of Stock-Market Prices," *The Journal of Business*, 38, 34–105. [773]

Fama, E. F., and French, K. R. (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56. [772,781]

—— (2015), "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116, 1–22. [772,781]

—— (2017), "Choosing Factors," *Chicago Booth Paper*, 16-17. [771]

Frazzini, A., and Pedersen, L. H. (2014), "Betting Against Beta," *Journal of Financial Economics*, 111, 1–25. [772,781]

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409. [776]

Hansen, L. P., and Jagannathan, R. (1991), "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262. [774]

Harvey, C. R., and Liu, Y. (2017), "Lucky Factors," *Working paper*. [771]

Harvey, C. R., and Zhou, G. (1990), "Bayesian Inference in Asset Pricing Tests," *Journal of Financial Economics*, 26, 221–254. [773]

Hou, K., Xue, C., and Zhang, L. (2015), "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, 28, 650–705. [772,781]

Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), London: Oxford University Press. [782]

Pástor, L., and Stambaugh, R. F. (2003), "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 111, 642–685. [772,781]

Pukthuanthong, K., Roll, R., and Subrahmanyam, A. (2017), "A Protocol for Factor Identification," Working paper. [771]

Richardson, M., and Smith, T. (1993), "A Test for Multivariate Normality in Stock Returns," *The Journal of Business*, 66, 295–321. [773]

Robert, C. P., and Casella, G. (2005), *Monte Carlo Statistical Methods (Springer Texts in Statistics)*, Secaucus, NJ: Springer-Verlag. [777]

Ruppert, D. (2010), *Statistics and Data Analysis for Financial Engineering*, Springer Texts in Statistics, New York: Springer. [773]

Shanken, J. (1987), "A Bayesian Approach to Testing Portfolio Efficiency," *Journal of Financial Economics*, 19, 195–215. [773]

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *Ann. Statist.*, 22, 1701–1728. [777]

Zhou, G. (1993), "Asset-pricing Tests under Alternative Distributions," *The Journal of Finance*, 48, 1927–1942. [773]