10-2010

Bayesian Hypothesis Testing in Latent Variable Models

Yong Li

Jun Yu
Singapore Management University, yujun@smu.edu.sg

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Bayesian Hypothesis Testing in Latent Variable Models

Yong Li & Jun Yu

October 2010

Paper No. 14-2010

Bayesian Hypothesis Testing in Latent Variable Models*

Yong Li
Sun Yat-Sen University

Jun Yu
Singapore Management University

Abstract: Hypothesis testing using Bayes factors (BFs) is known to suffer from several problems in the context of latent variable models. The first problem is computational. Another problem is that BFs are not well defined under the improper prior. In this paper, a new Bayesian method, based on decision theory and the EM algorithm, is introduced to test a point hypothesis in latent variable models. The new statistic is a by-product of the Bayesian MCMC output and, hence, easy to compute. It is shown that the new statistic is appropriately defined under improper priors because the method employs a continuous loss function. The finite sample properties are examined using simulated data. The method is also illustrated in the context of a one-factor asset pricing model and a stochastic volatility model with jumps using real data.

JEL classification: C11, C12, G12.
Keywords: Bayes factors; Kullback-Leibler divergence; Decision theory; EM Algorithm; Markov Chain Monte Carlo.

1 Introduction

Latent variable models have been widely used in economics, finance, and many other disciplines. They are appealing from both the practical and the theoretical perspectives. One advantage of using latent variables is that it reduces the dimensionality of data. A

*Li gratefully acknowledges the support of the Chinese Natural Science fund under Grant No.70901077, and the hospitality during his research visits to Sim Kee Boon Institute for Financial Economics at Singapore Management University. We would like to thank Robert Kohn, Jim Griffin, Havard Rue, Mike Titterington, Matt Wand, and participants to the Workshop on Recent Advances in Bayesian Computation for useful comments. Yong Li, Business School, Sun Yat-Sen University, Guangzhou, 510275, China. Jun Yu, School of Economics and Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yujun@smu.edu.sg. URL: http://www.mysmu.edu/faculty/yujun/.

well known example is the factor models. For example, in the arbitrage pricing theory (APT) of Ross (1976), and Roll and Ross (1980), returns of an infinite sequence of risky assets are assumed to depend linearly on a set of common factors. Another example is the modeling of the volatility of financial assets, where a separate stochastic process is often specified to describe the dynamics of the movement of volatility. This so-called stochastic volatility (SV) model has been proven to be an effective alternative to ARCH-type models; see Shephard (2005). The SV model is a special case of a more general class of models known as the state-space (SS) models. While statistical analysis of the linear Gaussian SS model is straightforward with the help of the Kalman filter technique, statistical analysis of a nonlinear or non-Gaussian SS model is much more challenging than its linear Gaussian counterpart.

For many latent variable models, it is difficult to use traditional frequentist estimation and inferential methods. The main reasons are as follows. First, for some latent variable models, such as the nonlinear or non-Gaussian SS models, the log-likelihood function of the observed variables (termed the observed data log-likelihood) often involves integrals which are not analytically tractable. When the dimension of the integrals is high, the classical numerical techniques may fail to work, and hence, the likelihood function becomes difficult to evaluate accurately. Consequently, the maximum likelihood (ML) method and all the tests based on ML, are difficult to use.

Second, for dynamic latent variable models, the frequentist inferential methods are almost always based on the asymptotic theory. The validity of the classical asymptotic theory requires a set of regularity conditions that may be too strong for economic data, to hold. For example, a regularity condition often used is stationarity. This condition may not be realistic for the macroeconomic and financial time series. In the context of a particular class of latent variable models, Chang, Miller, and Park (2009) discussed the impact of nonstationarity on the asymptotic distribution of the ML estimator.

Third, for the asymptotic theory to work well in finite samples, a large sample size is typically required. However, in many practical situations involved time series data, unfortunately, the sample size is not very large. In some cases, even if the sample size of available data is large, fully sampled data are not always utilized because of the concern over possible structural changes in the data. As a result, the classical asymptotic distribution may not be a good approximation to the finite sample distribution, and the inference based on the classical asymptotic theory may be misleading.

Due to the above mentioned difficulties in using the frequentist methods, there has been increasing interest in the Bayesian methods to deal with latent variable models.

With the advancement of MCMC algorithms and the rapidly expanding computing facility, the estimation of latent variable models has become increasingly easier. Since Bayesian inference is based on the posterior distribution, no asymptotic theory is needed for making statistical inferences.[1]

One of the most important statistical inferences is hypothesis testing, for which the formulation of the null hypothesis typically contains a unique value of a parameter which corresponds to the prediction of an important theory. Bayes factors (BFs) are the dominant method of Bayesian hypothesis testing (Kass and Raftery 1995; Geweke, 2007). One serious drawback is that they are not well defined when using an improper prior. The use of improper priors is typical in practice when noninformative priors are employed. Since the improper priors are specified only up to an undefined multiplicative constant, BFs contain undefined constants (Kass and Rafety, 1995), and hence, take arbitrary values.[2] Another drawback is computational. Calculation of BFs for comparing any two competing models requires the marginal likelihoods, and thus, a marginalization over the parameter vectors in each model. When the dimension of the parameter space is large, as is typical in latent variable models, the high-dimensional integration poses a formidable computational challenge, although there have been several interesting methods proposed in the literature for computing BFs from the MCMC output; see, for example, Chib (1995), and Chib and Jeliazkov (2001).

To define BFs with improper priors, a simple approach is to view part of the data as a training sample. The improper prior is then updated with the training sample to produce a new proper prior distribution. This leads to some variants of BFs; see, for example, the fractional BFs (O'Hagan 1995), and the intrinsic BFs (Berger and Perrichi, 1996). Instead of using BFs, Bernardo and Rueda (2002) suggested treating Bayesian hypothesis testing as a decision problem, and introduced a Bayesian test statistic that is well defined under improper priors. A crucial element in their approach is the specification of the loss function. They showed that the BFs approach to hypothesis testing is a special case of their decision structure with the loss function being a simple zero-one function.

In this paper, we generalize the Bayesian hypothesis testing approach of Bernardo and Rueda (2002) to deal with latent variable models. Like the approach of Bernardo and Rueda, our test statistic is also based on decision theory. However, our approach differs from that of Bernardo and Rueda in two ways. First, Bernardo and Rueda's approach is

[1]The posterior distribution is dependent on the choice of prior distributions, however. In some cases, the posterior distribution is sensitive to the specification of prior distributions; see, for example, Phillips (1991).

[2]If an informative and thus *proper* prior distributions are specified, BFs may be well defined.

based on the Kullback-Leibler (KL) loss function. Unfortunately, for the latent variable models, the KL function may involve calculation of intractable high-dimensional integrals. Instead we develop a new loss function based on the theory of the powerful EM algorithm that was originally proposed to estimate parameters of latent variable models (Dempster, et al., 1977). Second, we prove that the new test statistic is well defined under non-informative priors, show that it is a by-product of Bayesian estimation, and hence, make the computation relatively easy.

The paper is organized as follows. Section 2 introduces the setup of the latent variable models and reviews the Bayesian estimation method via MCMC. Section 3 motivates the use of continuous loss functions in Bayesian decision problems. In Section 4, the new Bayesian test statistic is introduced based on the decision theory and the EM algorithm in the context of latent variable models. Section 5 illustrates the finite sample behavior of the method in two Monte Carlo studies while Section 6 illustrates the new methods using real data. Section 7 concludes the paper, and Appendix collects the proof of the theoretical results in the paper.

2 Latent variable models and Bayesian estimation via MCMC

Without loss of generality, let $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n)^T$ denote observed variables and $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_n)^T$, the latent variables. The latent variable model is indexed by the parameter of interest, $\boldsymbol{\theta}$, and the nuisance parameter, $\boldsymbol{\psi}$. Let $p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$ be the likelihood function of the observed data, and $p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})$, the complete likelihood function. The relationship between these two functions is:

$$p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \int p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi}) d\boldsymbol{\omega}. \tag{1}$$

In many cases, the integral does not have an analytical expression. Consequently, the statistical inferences, such as estimation and hypothesis testing, are difficult to implement if they are based on the ML approach. In recent years, it has been documented that the latent variables models can be simply and efficiently estimated using MCMC techniques under the Bayesian framework. See Geweke, Koop, and van Dijk (2010) for algorithms, examples and references.

Let $p(\boldsymbol{\theta}, \boldsymbol{\psi})$ be the prior distribution of unknown parameter $\boldsymbol{\theta}, \boldsymbol{\psi}$. Due to the presence of the latent variables, the likelihood, $p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$, is intractable; hence it is difficult to compute the expectation of the posterior density, $p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})$. To alleviate this difficulty, the data-augmentation strategy of Tanner and Wong (1987) is applied to augment the

parameter space with the latent variable $\boldsymbol{\omega}$. Then, the Gibbs sampler can be used to generate random samples from the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\omega}|\boldsymbol{y})$. In particular, we start with an initial value, $[\boldsymbol{\theta}^{(0)}, \boldsymbol{\psi}^{(0)}, \boldsymbol{\omega}^{(0)}]$, and at the jth iteration, conditional on the current values $[\boldsymbol{\theta}^{(j)}, \psi^{(j)}, \boldsymbol{\omega}^{(j)}]$, we iterate:

(a) Generate $\boldsymbol{\theta}^{(j+1)}$ from $p(\boldsymbol{\theta}|\boldsymbol{\psi}^{(j)}, \boldsymbol{\omega}^{(j)}, \boldsymbol{y})$;

(b) Generate $\boldsymbol{\psi}^{(j+1)}$ from $p(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j+1)}, \boldsymbol{\omega}^{(j)}, \boldsymbol{y})$;

(c) Generate $\boldsymbol{\omega}^{(j+1)}$ from $p(\boldsymbol{\omega}|\boldsymbol{\psi}^{(j+1)}, \boldsymbol{\theta}^{(j+1)}, \boldsymbol{y})$.

After the effect of initialization dies off (namely with a sufficiently long period for the burning-in phase), the simulated random samples can be regarded as random observations from the joint distribution. Random observations drawn from the posterior simulation can be used to estimate the parameters. For example, Bayesian estimates of $\boldsymbol{\theta}$ and the latent variables $\boldsymbol{\omega}$ may be obtained via the corresponding sample mean of the generated random observations.

3 Bayesian hypothesis testing under decision theory

3.1 Hypothesis testing as a decision problem

After the model is estimated, often researchers are interested in testing a null hypothesis, of which the simplest contains a point. Typically, the point null hypothesis corresponds to the prediction of a theory. Assuming that the probabilistic behavior of observable data, $\mathbf{y} \in \mathbf{Y}$, is described appropriately by the probability model $M \equiv \{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})\}$ in term of the parameters of interest, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and the nuisance parameters, $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. Consider the following point null hypothesis:

$$\begin{array}{ll} H_0: & \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ H_1: & \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{array} . \tag{2}$$

Formally, this hypothesis testing problem can be taken as a decision problem where the action space has only two elements, namely, to accept (d_0) or to reject (d_1) the use of the null model, $M_0 \equiv \{p(\mathbf{y}|\theta_0, \boldsymbol{\psi}), \boldsymbol{\psi} \in \boldsymbol{\Psi}\}$, as a good proxy for the assumed model, $M_1 \equiv \{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}), \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\psi} \in \boldsymbol{\Theta}\}$.

For the decision problem, a loss function, $\{\mathcal{L}[d_i, (\boldsymbol{\theta}, \boldsymbol{\psi})], i = 0, 1\}$, which measures the loss of accepting H_0 or rejecting H_0 as a function of the actual value of the parameters $(\boldsymbol{\theta}, \boldsymbol{\psi})$, must be specified. Given the loss function and data $\mathbf{y}$, the optimal action is to reject H_0, if and only if (iff) the expected posterior loss of accepting H_0 is larger than the

expected posterior loss of rejecting H_0, that is,

$$\begin{aligned} & \int_\Theta \int_\Psi \mathcal{L}[d_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} - \int_\Theta \int_\Psi \mathcal{L}[d_1, (\boldsymbol{\theta}, \boldsymbol{\psi})] p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} \\ = & \int_\Theta \int_\Psi \left\{ \mathcal{L}[d_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] - \mathcal{L}[d_1, (\boldsymbol{\theta}, \boldsymbol{\psi})] \right\} p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \ \mathrm{d}\boldsymbol{\psi} > 0. \end{aligned}$$

Therefore, in practice, only the following net loss difference function is required to be specified:

$$\triangle \mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] = \mathcal{L}[d_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] - \mathcal{L}[d_1, (\boldsymbol{\theta}, \boldsymbol{\psi})].$$

It measures the evidence against H_0 as a function of $(\boldsymbol{\theta}, \boldsymbol{\psi})$. Following Berger (1985), any Bayesian admissible solution to the decision problem must satisfy,

$$\text{Reject} \quad H_0 \text{ iff } \quad T(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \int_\Theta \int_\Psi \triangle \mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} > 0, \tag{3}$$

for a pre-specified net loss difference function $\triangle \mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})]$.

3.2 Discrete loss function and Bayes factors

If the zero-one loss function is used, that is,

$$\mathcal{L}[d_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] = \begin{cases} 0 & if \ \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ 1 & if \ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \qquad \mathcal{L}[d_1, (\boldsymbol{\theta}, \boldsymbol{\psi})] = \begin{cases} 1 & if \ \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ 0 & if \ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases},$$

the net loss difference function $\triangle \mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})]$ is:

$$\triangle \mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi}))] = \begin{cases} -1 & if \ \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ 1 & if \ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}.$$

According to Equation (3), the corresponding decision rule is:

$$\text{Reject} \quad H_0 \text{ iff } \quad \int_\Psi (-1) p(\boldsymbol{\theta}_0, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\psi} + \int_\Theta \int_\Psi 1 p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} > 0.$$

In general, a positive probability ω, is assigned to the event $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, such that a reasonable prior for $\boldsymbol{\theta}$ with a discrete support at $\boldsymbol{\theta}_0$ is formulated as $p(\boldsymbol{\theta}) = \omega$, when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and $p(\boldsymbol{\theta}) = (1 - \omega)\pi(\boldsymbol{\theta})$. when $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, where $\pi(\boldsymbol{\theta})$ is a prior distribution. Hence, the decision criterion is:

$$\text{Reject } H_0 \text{ iff } \int_\Psi p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) \omega \pi(\boldsymbol{\psi}|\boldsymbol{\theta} = \boldsymbol{\theta}_0) \mathrm{d}\boldsymbol{\theta}_0 + \int_\Theta \int_\Psi p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \pi(\boldsymbol{\psi}|\boldsymbol{\theta}) (1-\omega) \pi(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\psi} > 0.$$

To represent the prior ignorance, the probability ω, is set to 0.5 and the criterion becomes:

$$\text{Reject} \quad H_0 \text{ iff } \quad B_{01} = \frac{\int_\Psi p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) \pi(\boldsymbol{\psi}|\boldsymbol{\theta} = \boldsymbol{\theta}_0) \mathrm{d}\boldsymbol{\psi}}{\int_\Theta \int_\Psi p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \pi(\boldsymbol{\theta}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi}} < 1,$$

where B_{01} is the well-known BF (Kass and Rafety, 1995).

When a subjective prior is not available, an objective prior or default prior may be used. Often, $\pi(\theta_k, \boldsymbol{\psi}|M_k)$ is taken as uninformative priors, such as Jeffreys or the reference prior (Jeffreys, 1961; Berger and Bernardo, 1992). These priors are generally improper, and it follows that $\pi(\boldsymbol{\theta}_k, \boldsymbol{\psi}|M_k) = C_k f(\boldsymbol{\theta}_k, \boldsymbol{\psi}|M_k)$, where $f(\boldsymbol{\theta}_k, \boldsymbol{\psi}|M_k)$ is a nonintegrable function, and C_k is an arbitrary positive constant, with $k = 0, 1$. In this case, the BF is

$$B_{01} = \frac{C_0}{C_1} \frac{\int_\Psi p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}_0) f(\boldsymbol{\theta}_0, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi}}{\int_\Theta \int_\Psi p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) f(\boldsymbol{\theta}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi}}. \tag{4}$$

Clearly the BF is ill-defined since it depends on the arbitrary constants, C_0/C_1. To overcome this problem, one may let part of the data be a training sample to generate a proper prior; see O'Hagan (1995), and Berger and Perrichi (1996). The choice of a training sample may be arbitrary.

3.3 KL continuous loss function

Bernardo and Rueda (2002) noted that it is more natural to assume the net loss function to be a continuous function of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. A particular function that was suggested in Bernardo and Rueda is the Kullback-Leibler (KL) divergence function. For any regular probability functions, $p(x)$ and $q(x)$, the KL divergence function can be expressed as:

$$K[p(x), q(\mathbf{x})] = \int p(x) \log \frac{p(x)}{q(x)} dx, \tag{5}$$

which is a non-negative measure and equal to 0 iff $p(x) = q(x)$.

Suppose that the log likelihood ratio for the two competing models is defined as:

$$R(\mathbf{Y}) := \log \frac{p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{Y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})}.$$

The larger the likelihood ratio, the stronger the evidence against the null hypothesis. The expected value of the log likelihood ratio is

$$E(R(\mathbf{Y})|H_1) = \int \log \frac{p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{Y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})} p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \mathrm{d}\mathbf{Y} = K[p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\psi}), p(\mathbf{Y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})],$$
$$E(R(\mathbf{Y})|H_0) = \int \log \frac{p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{Y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})} p(\mathbf{Y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) \mathrm{d}\mathbf{Y} = -K[p(\mathbf{Y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\psi})].$$

So $K[p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\psi}), p(\mathbf{Y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})]$ represents the expected log likelihood ratio when H_1 is true, and hence, may be interpreted as "the reminiscent of the power function in hypothesis testing, measuring the degree to which the data will reveal that the null hypothesis is false, when the alternative is in fact true", as noted in Eguchi and Copas (2006).

If the net loss function is chosen to be the KL function, then the decision criterion becomes:

$$\begin{aligned} T(\boldsymbol{\theta}, \boldsymbol{\theta}_0) &= \int_\Theta \int_\Psi \triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} \\ &= \int_\Theta \int_\Psi \left\{ \int \log \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})} p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \mathrm{d}\mathbf{y} \right\} p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}. \end{aligned}$$

This is the Bayesian hypothesis test statistic developed by Bernardo and Rueda (2002). To obtain some good properties such as symmetry, Bernardo and Rueda suggested using the following net loss function:

$$\triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] = \min\{K[p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}), p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})], K[p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})]\}. \tag{6}$$

Based on this loss function, the reference priors may be assigned to parameters to retain objectiveness. An obvious advantage is that this statistic is well defined under improper priors. Unfortunately, for latent variable models, KL may involve calculation of intractable high-dimensional integrals, and hence, may be difficult to evaluate.

4 A new loss function for latent variable models

The test statistic based on the KL divergence function requires that the observed data log-likelihood function be available analytically or be easy to calculate numerically. As argued above, however, for many latent variable models, evaluating the observed data log-likelihood function, and hence, the KL loss function is formidable. On the other hand, the EM algorithm has been widely used in the literature of latent variable models. The new difference loss function we propose in the present paper is based on the EM algorithm.

4.1 EM algorithm for latent variable models

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}, \boldsymbol{\psi})$ and $\mathbf{x} = (\mathbf{y}, \boldsymbol{\omega})$ be the complete-data set with a density $p(\mathbf{x}|\boldsymbol{\phi})$. The complete-data log-likelihood, $L_c(\mathbf{x}|\boldsymbol{\phi}) = \log p(\mathbf{x}|\phi)$, is often simple, whereas the observed data log-likelihood, $L_o(\mathbf{y}|\boldsymbol{\phi}) = \log p(\mathbf{y}|\boldsymbol{\phi})$, is very complicated in most situations because it often involves intractable integrals.

The basic idea of the EM algorithm is to replace maximization of the observed data log-likelihood function, $L_o(\mathbf{y}|\boldsymbol{\phi})$, with successful maximization of $Q(\phi|\phi^{(r)})$, the conditional expectation of the complete-data log-likelihood function, $L_c(\mathbf{x}|\boldsymbol{\phi})$, given the observation data $\mathbf{y}$ and a current fit $\boldsymbol{\phi}^{(r)}$ of the parameter. Thus, a standard EM algorithm consists of two steps: the *expectation* (E) step and the *maximization* (M) step. The E-step evaluates

the so-called Q function which is defined by

$$Q(\boldsymbol{\phi}|\boldsymbol{\phi}^{(r)}) = E_{\phi^{(r)}}\{L_c(\mathbf{x}|\boldsymbol{\phi})|\mathbf{y}, \phi^{(r)}\}, \tag{7}$$

where the expectation is taken with respect to the conditional distribution, $p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\phi}^{(r)})$. The M-step determines a $\boldsymbol{\phi}^{(r)}$ that maximizes $Q(\phi|\boldsymbol{\phi}^{(r)})$. Under some mild regularity conditions, the sequence, $\{\boldsymbol{\phi}^{(r)}\}$, obtained from the EM algorithm iterations converges to the ML estimate, $\widehat{\boldsymbol{\phi}}$. For details about the convergence properties of the sequence, $\boldsymbol{\theta}^{(r)}$, see Dempster, et al. (1977).

4.2 A new loss function

In a recent study, Ibrahim et al. (2008) proposed an information criterion for model selection based on $Q(\cdot|\cdot)$. Inspired by this study and the theoretical properties of the EM algorithm, we propose a new difference loss function for Bayesian point hypothesis testing in the context of latent variables models.

Consider the same nuisance parameter, $\boldsymbol{\psi}$. For any $\boldsymbol{\theta}, \boldsymbol{\theta}^* \in \Theta$, let $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = Q\left((\boldsymbol{\theta}, \boldsymbol{\psi})|(\boldsymbol{\theta}^*, \boldsymbol{\psi})\right)$. The new loss function is:

$$D(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \{Q(\boldsymbol{\theta}, \boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0, \boldsymbol{\theta})\} + \{Q(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\}.$$

The following lemma establishes some desirable properties of the new loss function, D. The proof of Lemma 1 can be found in Appendix 1.

Lemma 4.1 *The loss function D has the following properties:*

1. $D(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = D(\boldsymbol{\theta}_0, \boldsymbol{\theta})$;
2. $D(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \geq 0$;
3. $D(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = 0 \Longleftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Remark 4.1 *The new loss function is invariant under any one-to-one transformation of the parameters. This property is not shared by some simple loss functions, such as, the quadratic loss function.*

Remark 4.2 *If the observable variable* $\mathbf{y}$ *is independent on* $\boldsymbol{\omega}$*, the new loss function is reduced as a symmetric KL divergence function, that is,*

$$K(p(\boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi}), p(\boldsymbol{\omega}|\boldsymbol{\theta}_0, \boldsymbol{\psi})) + K(p(\boldsymbol{\omega}|\boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})). \tag{8}$$

Based on the new loss function, we define our Bayesian test statistic as the posterior mean of the loss function, namely,

$$T(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = E_{(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})} \{Q(\boldsymbol{\theta}, \boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0, \boldsymbol{\theta}) + Q(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\} \tag{9}$$

The following theorem gives the main result of this paper which shows how to compute the Bayesian test statistic from the MCMC output. The proof can be found in Appendix 2.

Theorem 4.1 *The Bayesian test statistic, $T(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$, can be expressed as*

$$T(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = E_{(\boldsymbol{\omega}, \boldsymbol{\psi}, \boldsymbol{\theta}|\mathbf{y})} \left\{ \log \frac{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_0, \boldsymbol{\psi})} \right\} + E_{(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})} \left\{ E_{(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi})} \left[\log \frac{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_0, \boldsymbol{\psi})}{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})} \right] \right\}. \tag{10}$$

Remark 4.3 *If Q function has a tractable form, it is obvious that the test statistic is only the by-product of the MCMC output under the alternative hypothesis. This is in shape contrast to BFs.*

Remark 4.4 *To implement the BF approach for hypothesis testing, numerical algorithms have to be applied to estimate BFs. However, it is difficult to assess the estimation accuracy. From Equation (10), it can be seen that the standard error of the newly proposed statistic is easily obtained.*

Remark 4.5 *If a prior distribution (such as Jeffrey's prior) is invariant under reparametrization, the Bayesian test statistic is robust to reparametrization.*

Remark 4.6 *While BFs are dependent on arbitrary constants under the non-informative prior, it can be shown that the proposed test statistic is well defined. The reason is that the arbitrary constants are canceled out in our statistic. The proof of this property can be found in Appendix 3.*

Remark 4.7 *In some interesting cases, unfortunately, the Q function does not have a tractable form. While the first term in (10) is only the by-product of the MCMC output under the alternative hypothesis, the second term in (10) is more difficult to calculate. In Appendix 4, we show how to approximate the second term by treating the nuisance parameter $\boldsymbol{\psi}$ as an additional latent variable, so that $T(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ can still be approximated using the MCMC output.*

Remark 4.8 *In practice, we need a threshold value for the rejection and the acceptance of H_0. Following Bernardo and Rueda (2002), we use the following decision rule:*

$$\textit{Accept } H_0 \textit{ if } T(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq C; \textit{ Reject } H_0 \textit{ if } T(\boldsymbol{\theta}, \boldsymbol{\theta}_0) > C,$$

where C is the threshold value. How to determine the threshold value is obviously important. Following McCulloch (1989), the comparison between two Bernoulli distributions may regarded as a reference case. McCulloch's idea is as follows. Consider two distributions, P_1 and P_2, whose corresponding densities are respectively denoted as p_1 and p_2. Set the KL divergence between P_1 and P_2 to be C, i.e., $K(P_1, P_2) = \int \log(p_1/p_2) dP_1 = C$, which measures the cost of predicting outcomes using P_2 when P_1 is the correct description of uncertainty. Let $B(p)$ be the Bernoulli distribution that assigns probability p to an event. We may find $q(C)$, such that

$$K(B(0.5), B(q(C))) = K(P_1, P_2) = C. \tag{11}$$

This means that the KL distance between P_1 and P_2 is required to be the same as that between $B(0.5)$ and $B(q(C))$. The latter distance is easier to be appreciated. In particular, it can be shown that $K(B(0.5), B(q(C))) = -\log(4q(C)(1-q(C)))/2$. Solving (11) for q (> 0.5), we get

$$q(C) = 0.5 + 0.5(1 - \exp(-2C))^{0.5}. \tag{12}$$

If $q(C) = 0.99$, the two Bernoulli distributions, $B(0.5)$ and $B(0.99)$, are very different. As a result, P_1 and P_2 must be very different too. This may be explained by the following analogy. The predicting outcomes with P_2, when the random variable is in fact P_1, is comparable with describing an unobserved Bernoulli event with probability 0.99, when in fact the probability is only 0.5. If $q(C) = 0.99$, using equation (12), we find that $C = 1.61$. For the new loss function developed in the present paper, C is the sum of the two KL divergence functions, as in (8) and (16) in Appendix 1. Consequently, we choose $3.22 = 2 \times 1.61$ to be the threshold value in the present paper. The use of threshold values is not new in the Bayesian literature. For example, Jeffreys' Bayes factor scale tells the strength of evidence in favor of one model versus another (Jeffreys, 1961). Perhaps a more natural approach is to obtain the empirical threshold value from the repeated simulated data. However, this model based calibration approach would be computationally more demanding.

5 Monte Carlo studies

In this section, we investigate the finite sample properties of the newly proposed statistic, using simulated data in the context of two financial models. The first is an asset pricing

model with a heavy-tailed distribution for which the Q function is available analytically. In the second study, we test for unit root in a stochastic volatility model with jumps for which the Q function does not have a closed form expression, and hence, has to be approximated using the MCMC output.

5.1 Asset pricing under multivariate t

Asset pricing theory is a pillar in modern finance. Various econometric approaches have been proposed to check the validity of various asset pricing models. For example, Gibbons, Ross, and Shanken (1989) developed a multivariate finite sample test. Bayesian tests have been proposed by Shanken (1987), Harvey and Zhou (1990), McCulloch and Rossi (1991), and Geweke and Zhou (1996). These tests were developed based on the normality assumption. Unfortunately, there has been overwhelming empirical evidence against normality for asset returns, which have led researchers to investigate asset pricing models with a heavy-tailed distribution, including the family of elliptical distributions discussed in Zhou (1993). In this section, we apply the new method to check the validity of a factor asset pricing model with a multivariate t distribution.

Let R_{it} be the excess return of portfolio i at period t with the following factor structure,

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i \boldsymbol{F}_t + \epsilon_{it}, \quad i = 1, 2, \cdots, N; t = 1, 2, \cdots, T, \tag{13}$$

where $\boldsymbol{F}_t$ is a $k \times 1$ vector of factor portfolio excess returns, $\boldsymbol{\beta}_i$ a $1 \times k$ vector of scaled covariances, ϵ_{it} the random error following the t distribution, N the number of portfolios, and T the length of the time series. This asset pricing model can be rewritten in the vector form,

$$\mathbf{R}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} \boldsymbol{F}_t + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \cdots, T, \tag{14}$$

where $\boldsymbol{\alpha}$ is a $1 \times k$ vector, β an $N \times K$ matrix, and $\epsilon_t \sim t(\mathbf{0}, \Psi, \nu)$. The density function of the multivariate t is given by

$$f(\boldsymbol{\epsilon}_t) = \frac{\Gamma(\frac{\nu+N}{2})}{(\pi\nu)^{\frac{2}{n}}\Gamma(\frac{\nu}{2})|\boldsymbol{\Psi}|^{\frac{1}{2}}} \left\{ 1 + \frac{\boldsymbol{\epsilon}_t' \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon}_t}{\nu} \right\}^{-\frac{\nu+N}{2}}.$$

The mean-variance efficiency implies that the excess premium $\boldsymbol{\alpha}$ should not be statistically different from zero. The hypothesis can be formulated as:

$$H_0 : \boldsymbol{\alpha} = 0 \times \mathbf{1}_N, \quad H_1 : \boldsymbol{\alpha} \neq 0 \times \mathbf{1}_N,$$

where $\mathbf{1}_N$ is $N \times 1$ vector with component 1.

It has been noted in Kan and Zhou (2006) that under the multivariate t specification, a direct numerical optimization of the observed data likelihood function is difficult. The scale mixture of multivariate normals may be used to represent the multivariate t distribution. As a consequence, Model (13) can be alternatively specified as:

$$\mathbf{R}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{F}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(0 \times \mathbf{1}_N, \boldsymbol{\Psi}/\omega_t), \quad \omega_t \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2}).$$

By treating ω_t as a latent variable, the powerful EM algorithm can be used to obtain the Q function in EM algorithm. Hence, one can obtain the Bayesian test statistic proposed earlier. The details are shown in Appendix 5.

In the simulation study, we consider a portfolio with $N = 8$, $k = 1$, and $\boldsymbol{\Psi}$ is diagonal. The true parameter values are set as $\nu = 8, \beta_i = 0.8, \psi_{ii} = 1.0, i = 1, 2, \cdots, 8$. In this asset pricing model, we consider five different cases, corresponding to $\boldsymbol{\alpha} = (0.00, 0.06, 0.07, 0.08, 0.10) \times \mathbf{1}_N$. Some vague conjugate prior distributions are specified for the parameters to represent the prior ignorance. In particular,

$$\mu_i \sim N[0, 100]; \beta_i \sim N[0, \psi_{ii}]; \psi_{ii}^{-1} \sim \Gamma[0.001, 0.001].$$

Three different sample sizes are considered, 250, 500 and 1000. The number of replications is always fixed at 100.

Since the Q function is known analytically, it is easy to obtain the Bayesian test statistic. The results are reported in Table 1 to 4. The estimates of $\boldsymbol{\alpha}$ are always close to the true value and the standard errors (SE) are always small, suggesting MCMC provides reliable estimates on $\boldsymbol{\alpha}$. Furthermore, the behavior of the estimates improves (smaller bias and smaller SE) when the sample size T increases. The finite sample behavior of our test is reasonable, judging by the proportion of the correct decisions over the 100 replications reported in Table 4. The power of the test increases with sample size.

To check the reliability of the threshold value of 3.22 obtained from the Bernoulli distribution, we repeat the experiment for 1,000 times under the null hypothesis for the three sample sizes and calculate the 99% quantile of the finite sample distribution of the test statistics. Since the Q function is analytically available, this Monte Carlo study is computationally feasible. The 99% quantile is 3.46, 3.50, and 3.49 when the sample size is 250, 500, and 1,000. All these values are close to 3.22, suggesting that the threshold value obtained from the Bernoulli distribution is reliable. Moreover, the three calibrated quantiles are very similar to each other, suggesting that the true threshold value is insensitive to the sample size.

Par	$\boldsymbol{\alpha} = 0 \times \mathbf{1}_N$	$\boldsymbol{\alpha} = 0.06 \times 1_N$	$\boldsymbol{\alpha} = 0.07 \times 1_N$	$\boldsymbol{\alpha} = 0.08 \times 1_N$	$\boldsymbol{\alpha} = 0.10 \times 1_N$
α_1	-0.0085	0.0753	0.0645	0.0713	0.0962
	(0.0672)	(0.0673)	(0.0674)	(0.0672)	(0.0669)
α_2	-0.0133	0.0676	0.0871	0.0666	0.0936
	(0.0675)	(0.0677)	(0.0674)	(0.0675)	(0.0675)
α_3	0.0027	0.0570	0.0740	0.0827	0.0934
	(0.0668)	(0.0676)	(0.0676)	(0.0668)	(0.0674)
α_4	-0.0104	0.0554	0.0667	0.0696	0.0950
	(0.0671)	(0.0674)	(0.0670)	(0.0672)	(0.0680)
α_5	0.0112	0.04986	0.0732	0.0912	0.0952
	(0.0679)	(0.0679)	(0.0674)	(0.0679)	(0.0680)
α_6	-0.0009	0.0613	0.0645	0.07914	0.1065
	(0.0677)	(0.0670)	(0.0675)	(0.0677)	(0.0672)
α_7	0.0011	0.0613	0.0587	0.0810	0.0965
	(0.0671)	(0.0670)	(0.0675)	(0.0671)	(0.0675)
α_8	-0.0023	0.0637	0.0550	0.07762	0.0953
	(0.0681)	(0.0676)	0.0677	(0.0680)	(0.0678)
Test	1.9820	2.7782	3.0626	3.2874	4.0274
	(0.8902)	(1.1108)	(1.1897)	(1.2248)	(1.3968)

Table 1: The average of the posterior mean and the posterior standard error of $\boldsymbol{\alpha}$ and the value of test statistic under 100 replications and $T = 250$ for the one factor asset pricing model. The numbers in parentheses are the standard errors.

5.2 Unit root test in a stochastic volatility model with jumps

Whether or not there is a unit root in volatility of financial assets has been a long-standing topic of interest to econometricians and empirical economists. In a log-normal stochastic volatility (SV) model, the volatility is often assumed to follow an AR(1) model with the autoregressive coefficient ϕ. The test of unit root amounts to testing $\phi = 1$. Based on the BF, So and Li (1999) proposed a Bayesian approach to test a unit root in the basic SV model. In this section, we consider the unit root test in the SV model with jumps. The presence of jumps in returns is an important stylized fact. Without including jumps, the jumps in the price will be mistakenly attributed to volatility, and hence, potentially change the dynamic properties of volatility. The model is specified as:

$$\begin{aligned} y_t &= s_t q_t + \exp(h_t/2)u_t,\ u_t \sim N(0,1), \\ h_t &= \tau + \phi(h_{t-1} - \tau) + \sigma v_t,\ v_t \sim N(0,1), \end{aligned}$$

where $t = 1, 2, \cdots, T$, q_t is an ordinary Bernoulli trial with $P(q_t = 1) = \pi$, and $\log(1+s_t) \sim N(-\eta^2/2, \eta^2)$. $s_t q_t$ can be viewed as a discretization of a finite activity Lévy process. This

Par	$\boldsymbol{\alpha} = 0 \times 1_N$	$\boldsymbol{\alpha} = 0.06 \times 1_N$	$\boldsymbol{\alpha} = 0.07 \times 1_N$	$\boldsymbol{\alpha} = 0.08 \times 1_N$	$\boldsymbol{\alpha} = 0.10 \times 1_N$
α_1	-0.0009	0.0527	0.0742	0.0866	0.0987
	(0.0476)	(0.0480)	(0.0477)	(0.0476)	(0.0477)
α_2	-0.0059	0.0589	0.0715	0.0746	0.0961
	(0.0473)	(0.0473)	(0.0474)	(0.0476)	(0.0474)
α_3	-0.0012	0.0505	0.0677	0.0781	0.1037
	(0.0474)	(0.0475)	(0.0475)	(0.0477)	(0.0474)
α_4	-0.0048	0.0595	0.0682	0.0736	0.0923
	(0.0478)	(0.0475)	(0.0473)	(0.0475)	(0.0479)
α_5	0.0051	0.0708	0.0746	0.0817	0.1040
	0.0476	(0.0475)	(0.0474)	(0.0475)	(0.0477)
α_6	0.0063	0.0601	0.0658	0.0847	0.1060
	(0.0477)	(0.0477)	(0.0473)	(0.0476)	(0.0477)
α_7	-0.0006	0.0587	0.0715	0.0882	0.1003
	(0.0472)	(0.0474)	(0.0472)	(0.0476)	(0.0473)
α_8	0.0020	0.0511	0.0642	0.0834	0.1005
	(0.0480)	(0.0481)	(0.0475)	(0.0477)	(0.0480)
Test	2.0082	3.4745	4.2915	4.9429	6.4886
	(0.8771)	(1.2395)	(1.4152)	(1.5231)	(1.7851)

Table 2: The average of the posterior mean and the posterior standard error of $\boldsymbol{\alpha}$ and the value of test statistic under 100 replications and $T = 500$ for the one factor asset pricing model. The numbers in parentheses are the standard errors.

model was introduced in Chib et al. (2002). The estimation of ϕ is complicated by the fact that volatility and jump components are both latent. For the same reason, the frequentist tests, including the Dickey-Fuller method, are difficult to use, and so are the BFs.

Following So and Li (1999), three values are considered for ϕ, 1.00, 0.98, 0.95, corresponding to the nonstationary, the nearly nonstationary, and the stationary case. The other parameters are set at $\tau = -9, \sigma^2 = 0.1, \pi = 0.08$, and $\eta = 0.03$. These values are empirically reasonable for daily equity returns. As in So and Li (1999) and Chib et al. (2002), we specify some proper prior distributions for the nuisance parameters:

$$\tau \sim N[0.0, 100], \frac{1}{\sigma_\eta} \sim Gamma(2+10^{-10}, 0.1), \pi \sim Beta(2, 100), \log(\eta) \sim N(-3.07, 0.149).$$

For ϕ, we consider a prior density that assigns a positive mass at unity, namely,

$$f(\phi) = \pi I(\phi = 1) + (1 - \pi) Uniform(0, 1), \pi \sim Uniform(0, 1), \tag{15}$$

where $I(x)$ is the indicator function, such that $I(x) = 1$ if x is true and 0 otherwise, π the weight that represents the prior probability for model M_0 formulated under the null

Par	$\boldsymbol{\alpha} = 0 \times 1_N$	$\boldsymbol{\alpha} = 0.06 \times 1_N$	$\boldsymbol{\alpha} = 0.07 \times 1_N$	$\boldsymbol{\alpha} = 0.08 \times 1_N$	$\boldsymbol{\alpha} = 0.10 \times 1_N$
α_1	0.0008	0.0535	0.0738	0.0820	0.1032
	(0.0338)	(0.0334)	(0.0337)	(0.0336)	(0.0335)
α_2	-0.0040	0.0623	0.0643	0.0796	0.1018
	(0.0337)	(0.0336)	(0.0336)	(0.0334)	(0.0336)
α_3	-0.0039	0.0561	0.0680	0.0752	0.1020
	(0.0335)	(0.0337)	(0.0336)	(0.0336)	(0.0337)
α_4	-0.0020	0.0626	0.0677	0.0850	0.0997
	(0.0336)	(0.0335)	(0.0336)	(0.0336)	(0.0335)
α_5	0.0026	0.0637	0.0735	0.0818	0.1017
	(0.0336)	(0.0335)	(0.0337)	(0.0337)	(0.0335)
α_6	-0.0022	0.0536	0.0721	0.0741	0.0934
	(0.0336)	(0.0336)	(0.0335)	(0.0336)	(0.0336)
α_7	0.0002	0.0513	0.0702	0.0778	0.1005
	(0.0335)	(0.0335)	(0.0336)	(0.0335)	(0.0335)
α_8	0.0015	0.0556	0.0679	0.0774	0.0953
	(0.0336)	(0.0336)	(0.0338)	(0.0336)	(0.0335)
Test	2.0386	4.9698	6.3204	7.5784	10.9372
	(0.8759)	(1.5064)	(1.7138)	(1.9147)	(2.3423)

Table 3: The average of the posterior mean and the posterior standard error of $\boldsymbol{\alpha}$ and the value of test statistic under 100 replications and $T = 1000$ for the one factor asset pricing model. The numbers in parentheses are the standard errors.

hypothesis. The Uniform distribution is assigned for π to represent the prior ignorance for model uncertainty. Since the Q function is not analytically available, Appendix 6 shows how to compute $T(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$.

Three different sample sizes are considered, 500, 1000 and 1500 in this simulation study. The number of replications is fixed at 100. The results are reported in Table 5 and Table 6, and the following conclusions may be drawn. First, the estimates of ϕ are always close to the true value and the SEs are always small, suggesting MCMC provides reliable estimates on ϕ with both sets of priors. Second, the behavior of the estimates improves (smaller bias and smaller SE) when the sample size increases. Third, the finite sample behavior of our test is reasonable, judging by the proportion of the correct decisions over the 100 replications, as reported in Table 6. For example, the test accepts the correct unit root hypothesis 75%, 91% and 93% of the time when the sample size is increased from 500 to 1500 observations. In addition, the correct stationary model is chosen 95% and 100% of the time when 500 observations are used, 100% and 100% of the time when 1000 observations are used, and 100% and 100% of the time when 1500 observations are used.

Parameter	$T = 250$	$T = 500$	$T = 1000$
$\boldsymbol{\alpha} = 0.00 \times \mathbf{1}_N$	95	98	97
$\boldsymbol{\alpha} = 0.06 \times 1_N$	27	63	93
$\boldsymbol{\alpha} = 0.07 \times 1_N$	42	75	99
$\boldsymbol{\alpha} = 0.08 \times 1_N$	48	87	100
$\boldsymbol{\alpha} = 0.10 \times 1_N$	74	100	100

Table 4: The number of correct decisions out of 100 replications for the one-factor asset pricing model.

n	Prior	Statistic	$\phi = 1$	$\phi = 0.98$	$\phi = 0.95$
500	Uniform	$\hat{\phi}$	0.9941	0.9747	0.9416
		$SE(\hat{\phi})$	0.0043	0.0137	0.0267
		$Test$	2.4750	13.1041	58.6337
1000	Uniform	$\hat{\phi}$	0.9971	0.9786	0.9509
		$SE(\hat{\phi})$	0.0019	0.0086	0.0159
		$Test$	1.7258	16.5421	75.8355
1500	Uniform	$\hat{\phi}$	0.9981	0.9782	0.9497
		$SE(\hat{\phi})$	0.0012	0.0071	0.0129
		$Test$	1.3566	24.7524	102.3799

Table 5: The average of the posterior mean and the posterior standard error of ϕ and the value of test statistic under 100 replications for the SV model with jumps.

6 Empirical Study

6.1 Testing the market price model

In the first empirical study, we illustrate our method by testing a simple asset pricing model – the market price model. This single-factor model is given as:

$$R_{it} = \alpha_i + \beta_i R_{Mt} + \epsilon_{it},$$

where R_{Mt} is the excess return of the market, and ϵ_{it} is independent over i. We consider the monthly returns of 25 portfolios and the market excess return. The portfolios, constructed at the end of each June, are the intersections of 5 portfolios formed on size (market equity, ME) and 5 portfolios formed on the ratio of book equity to market equity (BE/ME). This sample period is from July 1927 to December 2009, so that $N = 25$, $T = 1002$. The data are freely available from the data library of Kenneth French.[3]

[3]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

ϕ	$n = 500$	$n = 1000$	$n = 1500$
1.00	75	91	93
0.98	95	100	100
0.95	100	100	100

Table 6: The number of correct decisions under different sample sizes for the SV model with jumps.

Before estimating the model, we first use the Shapiro-Wilk statistic (Shapiro and Wilk, 1965) to test normality and found overwhelming evidence against it. Consequently, we replace normality with the t distribution. In the Bayesian analysis, we specify the vague conjugate prior distributions to represent the prior ignorance, namely,

$$\mu_i \sim N[0, 100], \beta_i \sim N[0, 100], \psi_{ii}^{-1} \sim \Gamma[0.001, 0.001].$$

Under these prior specifications, we run 30,000 Gibbs iterations with a burning-in sample of 20,000. The remaining 10,000 iterations are regarded as effective random samples for the posterior Bayesian inference. The convergence of Gibbs sampling is checked using the Raftery-Lewis diagnostic test statistic (Raftery and Lewis, 1992). The posterior mean of the degrees of freedom is 2.444, with standard error 0.1175. The other estimation results are reported in Table 7. The Bayesian test statistic for $\boldsymbol{\alpha} = 0 \times \mathbf{1}_{25}$ is 19.08, with standard error 4.057. Hence, we conclude that the asset pricing model is strongly rejected.

6.2 Unit root testing in the volatility of S&P 500 index

In the second empirical study, we test the unit root hypothesis in volatility of S&P 500 index sampled over the period that covers the 2007-2008 subprime crisis. The data are the demeaned daily returns of S&P 500 from January 3, 2005 to January 31, 2009. There are 1512 observations in the data. Following Chib et al. (2002), we specify the proper prior distributions as follows:

$$\tau \sim N[0.0, 100], \frac{1}{\sigma_\eta} \sim Gamma(2.5, 0.025), \pi \sim Beta(2, 100), \log(\eta) \sim N(-3.07, 0.149),$$
$$\phi \sim \pi I(\phi = 1) + (1 - \pi) Uniform(0, 1), \pi \sim Uniform(0, 1).$$

The empirical results are obtained based on 30,000 iterations after a burn-in of 20,000. The convergence of Gibbs sampling is checked using the Raftery-Lewis diagnostic test statistic. The results are reported in Table 8 and show that the unit root hypothesis is rejected.

Portfolio	$\boldsymbol{\alpha}$		$\boldsymbol{\beta}$		$\boldsymbol{\Psi}$	
	EST	SE	EST	SE	EST	SE
S1B1	-0.0087	0.0014	1.4040	0.0370	0.0025	0.0001
S1B2	-0.0028	0.0010	1.2700	0.0274	0.0014	0.0001
S1B3	-0.0014	0.0009	1.1690	0.0244	0.0011	0.0001
S1B4	-0.0006	0.0008	1.0870	0.0215	0.0008	0.0001
S1B5	0.0013	0.0009	1.1600	0.0245	0.0011	0.0001
S2B1	-0.0038	0.0009	1.2970	0.0232	0.0010	0.0001
S2B2	-0.0003	0.0007	1.1830	0.0183	0.0006	0.0000
S2B3	0.0014	0.0006	1.0890	0.0168	0.0005	0.0000
S2B4	0.0016	0.0007	1.1020	0.0176	0.0006	0.0003
S2B5	0.0012	0.0009	1.2130	0.0226	0.0010	0.0000
S3B1	-0.0016	0.0006	1.2350	0.0186	0.0006	0.0000
S3B2	0.0010	0.0006	1.1180	0.0150	0.0004	0.0000
S3B3	0.0016	0.0005	1.0720	0.0143	0.0004	0.0000
S3B4	0.0018	0.0006	1.0630	0.0154	0.0004	0.0000
S3B5	0.0014	0.0008	1.1590	0.0213	0.0008	0.0000
S4B1	-0.0006	0.0005	1.1380	0.0138	0.0004	0.0000
S4B2	-0.0001	0.0004	1.0670	0.0119	0.0003	0.0000
S4B3	-0.0006	0.0005	1.0580	0.0128	0.0003	0.0000
S4B4	-0.0007	0.0006	1.0560	0.0157	0.0005	0.0000
S4B5	-0.0001	0.0008	1.1970	0.0228	0.0010	0.0000
S5B1	0.0001	0.0004	0.9937	0.0109	0.0002	0.0000
S5B2	-0.0009	0.0004	0.9573	0.0107	0.0002	0.0000
S5B3	-0.0001	0.0005	0.8935	0.0128	0.0003	0.0000
S5B4	0.0001	0.0006	0.9629	0.0156	0.0004	0.0000
S5B5	0.0003	0.0010	1.0630	0.0283	0.0014	0.0001

Table 7: Bayesian estimation and standard error of the parameters for the market model with the multivariate t distribution.

Model	π	η	τ	ϕ	σ^2	Test
EST	0.0096	0.0504	-0.8130	0.9822	0.0281	47.84
SE	0.0065	0.0202	0.2597	0.0092	0.0113	NA

Table 8: Empirical results for S&P 500

7 Conclusion and discussion

In this paper, we have proposed a new loss function for Bayesian point hypothesis testing in the context of latent variable models. The loss function is based on the Q function of the EM algorithm. Based on the new loss function, a new Bayesian test statistic is developed. The main advantages of the new statistic is that it is a by-product of the MCMC output under the alternative hypothesis, and hence, easy to compute. The second advantage is that it is well-defined even under a non-informative prior specification.

While it is necessary to specify a threshold value to implement our test, various strategies are available for calibrating the threshold value. McCulloch (1989) provided a simple and effective approach. Soofi, Ebrahimi, and Habibullah (1995) extended McCulloch's method to cases that involve distributions other than Bernoulli, and proposed a calibration method based on a normalized transformation of the KL information. Both approaches are independent of the data. Perhaps a more natural approach is to borrow the idea from the bootstrap method by generating the empirical threshold value from the data. However, this necessitates higher computational cost.

The new approach has been applied to test a simple one-factor asset pricing model and the unit root hypothesis in a SV model with jumps. However, the technique itself is quite general and can be applied in many other contexts. Examples includes the Fama-French three factor models with dependent covariance structure and the testing of the number of factors in latent factor models, just to name a few.

8 Appendix

8.1 Appendix 1: Proof of Lemma 4.1

For any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, by the definition of $Q(\cdot|\cdot)$,

$$\begin{aligned} Q(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) &= E\{L_c(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_1, \boldsymbol{\psi})|\mathbf{y}, \boldsymbol{\theta}_2, \boldsymbol{\psi}\} = \int_\Omega \log p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_1, \boldsymbol{\psi}) p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_2, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\omega} \\ = & \int_\Omega \log p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_1, \boldsymbol{\psi}) p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_2, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\omega} = \int_\Omega \log[p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_1, \boldsymbol{\psi}) p(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\psi})] p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_2, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\omega} \\ = & \int_\Omega \log p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_1, \boldsymbol{\psi}) p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_2, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\omega} + \log p(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\psi}) = H(\boldsymbol{\theta_1}|\boldsymbol{\theta}_2) + \log p(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\psi}) \end{aligned}$$

It follows that,

$$\begin{aligned} & Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}) = H(\boldsymbol{\theta}|\boldsymbol{\theta}) + \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - H(\boldsymbol{\theta}_0|\boldsymbol{\theta}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) \\ = & H(\boldsymbol{\theta}|\boldsymbol{\theta}) - H(\boldsymbol{\theta}_0|\boldsymbol{\theta}) + \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) \\ = & \int_\Omega \log \frac{p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi})}{p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi})} p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\omega} + \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) \\ = & K[p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi}), p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi})] + \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) \\ & Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0) = K[p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi})] + \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) - \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}). \end{aligned}$$

where $K[\cdot, \cdot]$ is the KL divergence function. Therefore,

$$\begin{aligned} & D(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \{Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0|\boldsymbol{\theta})\} + \{Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0)\} \\ = & K[p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi}), p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi})] + K[p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi})], \end{aligned} \tag{16}$$

and the three properties stated in Lemma 4.1 naturally follow.

8.2 Appendix 2: Proof of Theorem 4.1

From Lemma 4.1, we have.

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}) &= \int_\Omega \log \frac{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_0, \boldsymbol{\psi})} p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\omega} \\ &= E_{(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi})} \left\{ \log \frac{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_0, \boldsymbol{\psi})} \right\} \\ Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0) &= \int_\Omega \log \frac{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_0, \boldsymbol{\psi})}{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})} p(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\omega} \\ &= E_{(\boldsymbol{\omega}|\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi})} \left\{ \log \frac{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}_0, \boldsymbol{\psi})}{p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})} \right\} \end{aligned}$$

Hence, the Bayesian test statistic can be expressed as:

$$\begin{aligned} T &= \int_\Theta \int_\Psi \{Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}) + Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0)\} \, p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} \\ &= E_{(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})}\{Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}|\boldsymbol{\theta})\} + E_{(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})}\{Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0)\} \end{aligned}$$

It can be shown that,

$$
\begin{aligned}
& E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\{Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0|\boldsymbol{\theta})\} = E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{E_{(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta},\boldsymbol{\psi})}\left[\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{\psi})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta}_0,\boldsymbol{\psi})}\right]\right\} \\
= & \int_\Theta\int_\Psi E_{(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta},\boldsymbol{\psi})}\left\{\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{\psi})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta}_0,\boldsymbol{\psi})}\right\}p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})d\boldsymbol{\theta}d\boldsymbol{\psi} \\
= & \int_\Theta\int_\Psi\left\{\int_\Omega\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{\psi})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta}_0,\boldsymbol{\psi})}p(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta},\boldsymbol{\psi})d\boldsymbol{\omega}\right\}p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} \\
= & \int_\Theta\int_\Psi\int_\Omega\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{\psi})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta}_0,\boldsymbol{\psi})}p(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})\mathrm{d}\boldsymbol{\omega}\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} \\
= & E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{\psi})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta}_0,\boldsymbol{\psi})}\right\},
\end{aligned}
$$

which proves Theorem 4.1.

8.3 Appendix 3

In this Appendix, we will show that the proposed statistic, $T(\boldsymbol{\theta},\boldsymbol{\theta}_0)$, is free of arbitrary constants. First, assume that some general improper priors satisfy $p(\boldsymbol{\psi}|\boldsymbol{\theta},H_k) = A_k f(\boldsymbol{\psi}|\boldsymbol{\theta},H_k)$, $p(\boldsymbol{\theta}|H_k) = B_k f(\boldsymbol{\theta}|H_k)$ where $f(\boldsymbol{\psi}|\boldsymbol{\theta},H_k)$, $f(\boldsymbol{\theta}|H_k)$ are the nonintegrable function, and A_k, B_k are arbitrary positive constants with $k = 0, 1$. Then, it can be shown that,

$$
\begin{aligned}
& p(\boldsymbol{\omega},\boldsymbol{\psi},\boldsymbol{\theta}|\mathbf{y},H_k) = \frac{p(\boldsymbol{\omega},\boldsymbol{\psi},\boldsymbol{\theta}|\mathbf{y},H_k)}{p(\mathbf{y}|H_k)} = \frac{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi},\boldsymbol{\theta}|H_k)}{\int_\Omega\int_\Theta\int_\Psi p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi},\boldsymbol{\theta}|H_k)\mathrm{d}\boldsymbol{\omega}\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{\theta}} \\
= & \frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta},H_k)p(\boldsymbol{\psi},\boldsymbol{\theta}|H_k)}{\int_\Omega\int_\Theta\int_\Psi p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta},H_k)p(\boldsymbol{\psi},\boldsymbol{\theta}|H_k)\mathrm{d}\boldsymbol{\omega}\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{\theta}} \\
= & \frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta},H_k)A_k f(\boldsymbol{\psi}|\boldsymbol{\theta},H_k)B_k f(\boldsymbol{\theta}|H_k)}{\int_\Omega\int_\Theta\int_\Psi p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta},H_k)A_k f(\boldsymbol{\psi}|\boldsymbol{\theta},H_k)B_k f(\boldsymbol{\theta}|H_k)\mathrm{d}\boldsymbol{\omega}\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{\theta}} \\
= & \frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta},H_k)f(\boldsymbol{\psi},\boldsymbol{\theta}|H_k)}{\int_\Omega\int_\Theta\int_\Psi p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta},H_k)f(\boldsymbol{\psi},\boldsymbol{\theta}|H_k)\mathrm{d}\boldsymbol{\omega}\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{\theta}}.
\end{aligned}
$$

Hence, $p(\boldsymbol{\omega},\boldsymbol{\psi},\boldsymbol{\theta}|\mathbf{y},H_k)$ is independent on A_k, B_k. Similarly, we can show that $p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y},H_k)$ and $p(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta},\boldsymbol{\psi},H_k)$ are also independent on A_k, B_k. Furthermore, from Appendix 1 and

Appendix 2, we have,

$$
\begin{aligned}
& E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{\log\frac{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} = E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})p(\boldsymbol{\psi}|\boldsymbol{\theta})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)p(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} \\
= \;& E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})A_1 f(\boldsymbol{\psi}|\boldsymbol{\theta})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)A_0 f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} \\
= \;& E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})f(\boldsymbol{\psi}|\boldsymbol{\theta})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} + \log\frac{A_1}{A_0} \\
& E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{E_{(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta}_0,\boldsymbol{\psi})}\left[\log\frac{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}\right]\right\} \\
= \;& E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{E_{(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta}_0,\boldsymbol{\psi})}\left[\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)p(\boldsymbol{\psi}|\boldsymbol{\theta})}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})p(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right]\right\} \\
= \;& E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{E_{(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta}_0,\boldsymbol{\psi})}\left[\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)A_0 f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})A_1 f(\boldsymbol{\psi}|\boldsymbol{\theta}))}\right]\right\} \\
= \;& E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{E_{(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta}_0,\boldsymbol{\psi})}\left[\log\frac{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})f(\boldsymbol{\psi}|\boldsymbol{\theta}))}\right]\right\} + \log\frac{A_0}{A_1}.
\end{aligned}
$$

From Theorem 4.1, it can be seen that the arbitrary constants are cancelled. As a result, the Bayesian test statistic is free of the arbitrary constants.

8.4 Appendix 4

In this Appendix, we will propose a method to calculate $T(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ when Q is not analytically tractable. To do so, we treat the nuisance parameter $\boldsymbol{\psi}$ as the latent variable. The Bayesian test statistic is shown to take the form of:

$$
T(\boldsymbol{\theta},\boldsymbol{\theta}_0) = E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})}\left\{\log\frac{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} + E_{(\boldsymbol{\theta}|\mathbf{y})}\left\{E_{(\boldsymbol{\omega},\boldsymbol{\psi}|\mathbf{y},\boldsymbol{\theta}_0)}\left[\log\frac{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}\right]\right\}.
$$

The first expectation is only a by-product of Bayesian estimation under the alternative hypothesis and can be easily approximated with the MCMC output. To approximate the second expectation, let

$$
f(\boldsymbol{\theta}) = \int_{\Omega}\int_{\Psi}\log p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})p(\boldsymbol{\omega},\boldsymbol{\psi}|\mathbf{y},\boldsymbol{\theta}_0)\mathrm{d}\boldsymbol{\omega}\mathrm{d}\boldsymbol{\psi},
$$

and

$$
\dot{f}(\boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}, \ddot{f}(\boldsymbol{\theta}) = \frac{\partial^2 f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}.
$$

Taking the second Taylor expansion of f at $\boldsymbol{\theta}_0$, we get,

$$
f(\boldsymbol{\theta}) \approx f(\boldsymbol{\theta}_0) + \dot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\ddot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0).
$$

It follows that,

$$
\begin{aligned}
& E_{(\boldsymbol{\theta}|\mathbf{y})}\left\{E_{(\boldsymbol{\omega},\boldsymbol{\psi}|\mathbf{y},\boldsymbol{\theta}_0)}\left\{\log\frac{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}\right\}\right\} = E_{(\boldsymbol{\theta}|\mathbf{y})}\{f(\boldsymbol{\theta}_0)-f(\boldsymbol{\theta})\} \\
\approx & \int_{\Theta}\left\{-\dot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)-(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\ddot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right\}p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \\
= & E_{(\boldsymbol{\theta}|\mathbf{y})}\left\{-\dot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)-(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\ddot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right\}.
\end{aligned}
$$

Assuming the exchange between the integration and the differentiation in the $\boldsymbol{\theta}$, we then get,

$$
\begin{aligned}
\dot{f}(\boldsymbol{\theta}) &= \frac{\partial f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \int_{\Omega}\frac{\partial\log p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}p(\boldsymbol{\omega},\boldsymbol{\psi}|\mathbf{y},\boldsymbol{\theta}_0)\mathrm{d}\boldsymbol{\omega} \\
\ddot{f}(\boldsymbol{\theta}) &= \frac{\partial^2 f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = \int_{\Omega}\frac{\partial^2\log p(\mathbf{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\boldsymbol{\theta}^T}p(\boldsymbol{\omega},\boldsymbol{\psi}|\mathbf{y},\boldsymbol{\theta}_0)\mathrm{d}\boldsymbol{\omega}.
\end{aligned}
$$

At $\boldsymbol{\theta}_0$, the first-order and the second-order differentiations can be easily approximated using MCMC samples of the posterior distribution, $p(\boldsymbol{\omega},\boldsymbol{\psi}|\mathbf{y},\boldsymbol{\theta}_0)$.

8.5 Appendix 5: Calculation of $T(\ \boldsymbol{\theta},\ \boldsymbol{\theta}_0)$ for the factor asset pricing model

Let $y=\{y_1,y_2,\cdots,\boldsymbol{y}_n\}$, $\boldsymbol{\omega}=\{\omega_1,\omega_2,\cdots,\omega_n\}$. The observed data log-likelihood function, $L_o(y|\boldsymbol{\theta},\boldsymbol{\psi})$, is expressed as:

$$
C-\frac{T}{2}\sum_{i=1}^{k}\log\phi_{ii}-\frac{\nu+k}{2}\sum_{t=1}^{T}\sum_{i=1}^{k}\log\left(1+\frac{(y_{it}-\alpha_i-\boldsymbol{\beta}_i\boldsymbol{F}_t)^2}{\nu\phi_{ii}}\right),
$$

where C is a constant. Based on the multivariate normal-gamma mixture representation for the multivariate t distribution, the complete log-likelihood, $L_c(y,\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{\psi})$, can be expressed as

$$
C+\frac{N}{2}\sum_{t=1}^{T}\log\omega_t-\frac{T}{2}\sum_{i=1}^{k}\log\phi_{ii}-\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{k}\omega_t\phi_{ii}^{-1}(y_{it}-\alpha_i-\boldsymbol{\beta}_i\boldsymbol{F}_t)^2.
$$

Thus, the posterior expectation of ω_t given the data and parameters is

$$
E(\boldsymbol{\omega}_t|\boldsymbol{\theta},\boldsymbol{\psi},\mathbf{y}_t)=\frac{\nu+N}{\nu+\sum_{i=1}^{k}\phi_{ii}^{-1}(y_{it}-\alpha_i-\boldsymbol{\beta}_i\boldsymbol{F}_t)^2}, t=1,2,\cdots,T.
$$

For the asset pricing model considered in the simulation study, we can show that,

$$
\begin{aligned}
& Q(\boldsymbol{\theta}|\boldsymbol{\theta})-Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}) = \int[L_c(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{\psi})-L_c(\mathbf{y},\boldsymbol{\omega}|\boldsymbol{\theta}_0,\boldsymbol{\psi})]p(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\omega} \\
= & \int\sum_{t=1}^{T}\sum_{i=1}^{N}\left\{\omega_t\phi_{ii}^{-1}[(y_{it}-\boldsymbol{\beta}_i\boldsymbol{F}_t)\alpha_i-\frac{1}{2}\alpha_i^2]\right\}p(\boldsymbol{\omega}|\mathbf{y},\boldsymbol{\theta},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\omega} \\
= & \sum_{t=1}^{T}\sum_{i=1}^{N}\left\{E(\omega_t|\boldsymbol{\theta},\boldsymbol{\psi},\mathbf{y}_t)\phi_{ii}^{-1}[(y_{it}-\boldsymbol{\beta}_i\boldsymbol{F}_t)\alpha_i-\frac{1}{2}\alpha_i^2]\right\},
\end{aligned}
$$

and that

$$Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0) = \sum_{t=1}^{T}\sum_{i=1}^{N}\left\{-E(\omega_t|\boldsymbol{\theta}_0, \boldsymbol{\psi}, \mathbf{y}_t)\phi_{ii}^{-1}[(y_{it} - \boldsymbol{\beta}_i\boldsymbol{F}_t)\alpha_i - \frac{1}{2}\alpha_i^2]\right\}.$$

Therefore, the Bayesian test statistic is given by,

$$\begin{aligned} & T(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left[Q(\boldsymbol{\theta}|\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}) + Q(\boldsymbol{\theta}_0|\boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0)\right] \\ = \quad & E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{\sum_{t=1}^{T}\sum_{i=1}^{N}[E(\omega_t|\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y}_t) - E(\omega_t|\boldsymbol{\theta}_0, \boldsymbol{\psi}, \mathbf{y}_t)]\phi_{ii}^{-1}[(y_{it} - \boldsymbol{\beta}_i\boldsymbol{F}_t)\alpha_i - \frac{1}{2}\alpha_i^2]\right\}. \end{aligned}$$

8.6 Appendix 6: Calculation of $T(\boldsymbol{\theta},\ \boldsymbol{\theta}_0)$ for the SV model

Let $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \boldsymbol{y}_n\}$, $\mathbf{h} = \{h_1, h_2, \cdots, h_n\}$, $\mathbf{s} = \{s_1, s_2, \cdots, s_n\}$, $\mathbf{q} = \{q_1, q_2, \cdots, q_n\}$. The joint density function is:

$$\begin{aligned} & p(\mathbf{y}, \boldsymbol{h}, \boldsymbol{s}, \boldsymbol{q}|\pi, \eta, \tau, \phi, \sigma^2) = \prod_{t=1}^{T} p(y_t, h_t, s_t, q_t|h_{t-1}, \pi, \eta, \tau, \phi, \sigma^2) \\ = \quad & \prod_{t=1}^{T}\left\{p(y_t|h_t, s_t, q_t)p(h_t|h_{t-1}, \tau, \phi, \sigma^2)p(q_t|\pi)p(s_t|\eta)\right\} \\ = \quad & \prod_{t=1}^{T}\{C\sigma^{-1}\exp\left[\frac{-(y_t - s_t q_t)^2\exp(-h_t) + h_t}{2} - \frac{(h_t - \tau - \phi(h_{t-1} - \tau))^2}{2\sigma^2}\right] \\ & \times\pi^{q_t}(1-\pi)^{(1-q_t)}\frac{1}{\eta(1+s_t)}\exp\left[-\frac{(\log(1+s_t) - 0.5\eta^2)^2}{2\eta^2}\right]\}, \end{aligned} \tag{17}$$

where C is a known constant. The observed data log-likelihood function is given by,

$$L_o(\mathbf{y}|\pi, \eta, \tau, \phi, \sigma^2) = \log\left\{\int p(\mathbf{y}, \boldsymbol{h}, \boldsymbol{s}, \boldsymbol{q}|\pi, \eta, \tau, \phi, \sigma^2)\mathrm{d}\mathbf{h}\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{q}\right\}.$$

We can see that this function involves a $3T$-dimensional integral. When T is large, the optimization is extremely difficult.

For the SV model with jumps, the method shown in Appendix 4 can be used to approximate the Bayesian test statistic, $T(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$. To do, several components are required. For example,

$$\begin{aligned} & \log p(\mathbf{y}, \mathbf{h}, \mathbf{s}, \mathbf{q}, \pi, \eta, \tau, \sigma^2|\phi) - \log p(\mathbf{y}, \mathbf{h}, \mathbf{s}, \mathbf{q}, \pi, \eta, \tau, \sigma^2|\phi_0) \\ = \quad & \log p(\mathbf{y}, \mathbf{h}, \mathbf{s}, \mathbf{q}|\pi, \eta, \tau, \sigma^2, \phi) + \log p(\pi, \eta, \tau, \sigma^2|\phi) \\ & - \log p(\mathbf{y}, \mathbf{h}, \mathbf{s}, \mathbf{q}|\pi, \eta, \tau, \sigma^2, \phi_0) - \log p(\pi, \eta, \tau, \sigma^2|\phi_0) \\ = \quad & \log p(\mathbf{y}, \mathbf{h}, \mathbf{s}, \mathbf{q}|\pi, \eta, \tau, \sigma^2, \phi) - \log p(\mathbf{y}, \mathbf{h}, \mathbf{s}, \mathbf{q}|\pi, \eta, \tau, \sigma^2, \phi_0). \end{aligned}$$

It follows that,

$$\begin{aligned} & \log p(\mathbf{y},\mathbf{h},\mathbf{s},\mathbf{q}|\pi,\eta,\tau,\sigma^2,\phi) - \log p(\mathbf{y},\mathbf{h},\mathbf{s},\mathbf{q}|\pi,\eta,\tau,\sigma^2,\phi_0) \\ = & \frac{1}{2\sigma^2}\sum_{t=1}^{T}\{(\phi_0^2-\phi^2)(\omega_{t-1}-\tau)^2 - 2(\phi_0-\phi)(\omega_t-\tau)(\omega_{t-1}-\mu)\}. \end{aligned}$$

Moreover,

$$\begin{aligned} \dot{f}(\phi) &= \int \frac{\partial \log p(\mathbf{y},\mathbf{h},\mathbf{s},\mathbf{q},\pi,\eta,\tau,\sigma^2|\phi)}{\partial\phi} p(\mathbf{h},\mathbf{s},\mathbf{q},\pi,\eta,\tau,\sigma^2|\mathbf{y},\phi_0)\mathrm{d}\mathbf{h}\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{q}\mathrm{d}\pi\mathrm{d}\eta\mathrm{d}\tau\mathrm{d}\sigma^2 \\ &= E_{(\mathbf{h},\mathbf{s},\mathbf{q},\pi,\eta,\tau,\sigma^2|\mathbf{y},\phi_0)}\left\{\frac{1}{\sigma^2}\sum_{t=1}^{T}[(h_t-\tau-\phi(h_{t-1}-\tau))(h_{t-1}-\tau)]\right\}, \\ \ddot{f}(\phi) &= \int \frac{\partial^2 \log p(\mathbf{y},\mathbf{h},\mathbf{s},\mathbf{q},\pi,\eta,\tau,\sigma^2|\phi)}{\partial^2\phi} p(\mathbf{h},\mathbf{s},\mathbf{q},\pi,\eta,\tau,\sigma^2|\mathbf{y},\phi_0)\mathrm{d}\mathbf{h}\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{q}\mathrm{d}\pi\mathrm{d}\eta\mathrm{d}\tau\mathrm{d}\sigma^2 \\ &= E_{(\mathbf{h},\mathbf{s},\mathbf{q},\pi,\eta,\tau,\sigma^2|\mathbf{y},\phi_0)}\left\{-\frac{1}{\sigma^2}\sum_{t=1}^{n}[(\omega_{t-1}-\tau)^2]\right\}, \\ \dddot{f}(\phi) &= 0. \end{aligned}$$

References

Bernardo, J.M. and Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review*, **70**, 351-372.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Spring-Verlag, 2th edition.

Berger, J.O. and Bernardo, J.M. (1992). On the development of reference priors. *Bayesian Statistics 4* edited by J. M. Bernardo, J.O. Berger, D.V. Lindley and A. F. M. Smith. Oxford: Oxford University Press, 35-60.

Berger, J.O., and Perrichi, L.R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association*, **91**, 109-122.

Chang, Y., Isaac Miller, J., Park, J.Y. (2009). Extracting a common stochastic trend: theory with some applications. *Journal of Econometrics*, **150**, 231-247.

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

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, **96**, 270-281.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, **108**, 281–316.

Dempster, A.P., Laird, N.M. and Rublin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, **1**, 1-38.

Eguchi, S. and Copas, J. (2006). Interpreting Kullback–Leibler divergence with the Neyman-Pearson lemma. *Journal of Multivariate Analysis*, **97**, 2034–2040.

Gibbons, M.R., Ross, S.A and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, **57**, 1121- 1152.

Geweke, J. (2007). Bayesian model comparison and validation. *American Economic Review*, **97**, 60-64.

Geweke, J.F., Koop, G. and van Dijk, H.K. (2010). *Handbook of Bayesian Econometrics*. Oxford: Oxford University Press.

Geweke, J., and G. Zhou. (1996). Measuring the pricing errors of the arbitrage pricing theory. *Review of Financial Studies*, **9**, 557-587.

Harvey, C., and Zhou, G. (1990). Bayesian inference in asset pricing tests. *Journal of Financial Economics*, **26**, 221-254.

Ibrahim, J.G., Zhu, H.T. and Tang, N.S. (2008). Model Selection Criteria for Missing-Data Problems Using the EM Algorithm. *Journal of the American Statistical Association*, **103**, 1648-1658.

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

Kass, R. E. and Raftery, A. E. (1995). Bayes Factor. *Journal of the Americana Statistical Association*, **90**, 773-795.

Kan, R. and Zhou, G.F. (2006). Modeling Non-normality using multivariate t: Implications for asset pricing. Working paper, University of Toronto.

McCulloch, R.E. (1989). Local Model Influence. *Journal of the American Statistical Association*, **84**, 673-478.

McCulloch, R.E., and Rossi, P.E. (1991). A Bayesian approach to testing the arbitrage pricing theory. *Journal of Econometrics*, **49**,141-168.

O'Hagan. (1995). Fractional Bayes Factors for Model Comparison (with discussion). *Journal of the Royal Statistical Society, Series B*, **57**, 99-138.

Phillips, P.C.B. (1991). To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends. *Journal of Applied Econometrics*, **6**, 333-364.

Raftery, A. and Lewis, S. (1992). How many iterations in the Gibbs sampler? in Bernardo, J., Berger, J., Dawid, A. and Smith, A. eds, *Bayesian Statistics*, Vol,4, Claredon Press, Oxford, 763-774.

Roll, R., Ross, S.A. (1980). An Empirical Investigation of the Arbitrage Pricing Theory. *Journal of Finance*, **35**, 1073-1103.

Ross, S.A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, **13**, 341-360.

Shanken, J. (1987). A Bayesian approach to testing portfolio efficiency. *Journal of Financial Economics*, **19**, 217-244.

Shephard, N. (2005). *Stochastic Volatility Models, Selected Reading.* Oxford: Oxford University Press.

Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika*, **52**, 591-611.

Soofi, E. S., Ebrahimi, N., and Habibullah, M. (1995). Information distinguishability with application to the analysis of failure Data. *Journal of the American Statistical Association*, **90**, 657-668.

So, M.K.P. and Li, W.K. (1999). Bayesian unit-root testing in stochastic volatility models. *Journal of Business and Economic statistics*, **17**, 491-496.

Tanner, T.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528-549.

Zhou, G.F. (1993). Asset pricing testing under alternative distributions. *Journal of Finance*, **5**, 1927-1942.