Journal of Econometrics 166 (2012) 237–246

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Bayesian hypothesis testing in latent variable models☆

Yong Li [a], Jun Yu [b,*]

[a] Sun Yat-Sen Business School, Sun Yat-Sen University, Guangzhou, 510275, China

[b] Sim Kee Boon Institute for Financial Economics, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Received 17 October 2010
Received in revised form
8 August 2011
Accepted 21 September 2011
Available online 7 October 2011

JEL classification:
C11
C12
G12

Keywords:
Bayes factors
Kullback–Leibler divergence
Decision theory
EM algorithm
Markov chain Monte Carlo

ABSTRACT

Hypothesis testing using Bayes factors (BFs) is known not to be well defined under the improper prior. In the context of latent variable models, an additional problem with BFs is that they are difficult to compute. In this paper, a new Bayesian method, based on the 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. In addition, it is easy to interpret. The method is illustrated using a one-factor asset pricing model and a stochastic volatility model with jumps.

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 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 stochastic volatility (SV) model that 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 et al. (2009) discussed the impact of

☆ Li gratefully acknowledges the financial support of the Chinese Natural Science fund (No. 70901077), the Chinese Education Ministry Social Science fund (No. 09YJC790266), the Fundamental Research Funds for the Central Universities 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, Peter Phillips, Mike Titterington, Matt Wand, and participants to the Workshop on Recent Advances in Bayesian Computation and the AMES2011 meeting, especially the Co-editor, an Associate Editor, and two referees for useful comments.

* Corresponding author. Tel.: +65 68280858; fax: +65 68280833.
E-mail address: yujun@smu.edu.sg (J. Yu).
URL: http://www.mysmu.edu/faculty/yujun/ (J. Yu).

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.09.040

nonstationarity on the asymptotic distribution of the ML estimator. In the case of general hidden Markov models, the asymptotic properties of the ML estimate remain largely unknown, with the exception of consistency which was recently developed in Douc et al. (2011).

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. This property is true for all models, including models with latent variables. 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 Raftery, 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).[3] Instead of using BFs, Bernardo and Rueda (2002), BR hereafter 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.[4]

In this paper, we generalize the Bayesian hypothesis testing approach of BR to deal with latent variable models. Like the approach of Bernardo and Rueda, our test statistic is also based on the decision theory. However, our approach differs from theirs in two ways. First, BR's approach is based on the Kullback–Leibler (KL) loss function. Unfortunately, for the latent variable models, the KL function used in BR 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 do the maximum likelihood estimation of parameters in latent variable models (Dempster et al., 1977). Second, we prove that the new test statistic is well defined under improper 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 MCMC method. 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 new method using two models, a one-factor asset pricing model and a stochastic volatility model with jumps. Section 6 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 $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^T$ denote observed variables and $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_n)^T$, the latent variables. The latent variable model is indexed by the parameter of interest, $\boldsymbol{\theta}$, and the nuisance parameter, $\boldsymbol{\psi}$. Let $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$ be the likelihood function of the observed data, and $p(\boldsymbol{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, \boldsymbol{\psi})$, the complete likelihood function. The relationship between these two functions is:

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

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. 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, $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$, is intractable; hence it is difficult to compute the expectation of the posterior density, $p(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{y})$. 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 $p(\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\omega}|\mathbf{y})$. After the effect of initialization dies off (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. For further details about Bayesian estimation of latent variable models via MCMC such as algorithms, examples and references, see Geweke et al. (2011).

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.

[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 distribution is specified, BFs may be well defined.

[3] Alternatively, one may use model selection criteria, such as the deviance information criterion proposed by Spiegelhalter et al. (2002) and applied to the stochastic volatility models by Berg et al. (2004).

[4] Poirier (1997) developed a loss function approach for hypothesis testing for models without latent variables.

Typically, the point null hypothesis corresponds to the prediction of a theory. Assuming that the probabilistic behavior of observable data, $\boldsymbol{y} \in \boldsymbol{Y}$, is described appropriately by the probability model $M \equiv \{p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\psi})\}$ 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:

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

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, $M_0 \equiv \{p(\boldsymbol{y}|\theta_0, \boldsymbol{\psi}), \boldsymbol{\psi} \in \boldsymbol{\Psi}\}$, as a good proxy for the assumed model, $M_1 \equiv \{p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\psi}), \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\psi} \in \boldsymbol{\Theta}\}$.

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 $\boldsymbol{y}$, 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,

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

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

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

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,

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

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

3.2. Discrete loss function and Bayes factors

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

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

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

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

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

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

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

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:

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

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

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

where B_{01} is the well-known BF (Kass and Raftery, 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 the Jeffreys or the reference prior (Jeffreys, 1961; BR, 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

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

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

BR (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 BR is the KL divergence function. For any regular probability functions, $p(x)$ and $q(x)$, the KL divergence can be expressed as:

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

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

If the net loss function is chosen to be the KL divergence between the unrestricted and the restricted likelihood functions, the decision criterion becomes:

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

This is the Bayesian hypothesis test statistic developed in BR. To obtain some good properties such as symmetry, BR suggested using the following net loss function:

$$\begin{aligned} \Delta\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] = \min\{&K[p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\psi}), p(\boldsymbol{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})], \\ &K[p(\boldsymbol{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\psi})]\}. \end{aligned} \tag{6}$$

As shown in BR, this net loss function measures the minimum amount of information (in natural information units) that one observation may be expected to provide in order to discriminate between $p(y|\boldsymbol{\theta}, \boldsymbol{\psi})$ and $p(y|\boldsymbol{\theta}_0, \boldsymbol{\psi})$. 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 so-called Q function used in the EM algorithm.

4.1. EM algorithm for latent variable models

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

The basic idea of the EM algorithm is to replace maximization of the observed data log-likelihood function, $L_o(\boldsymbol{y}|\boldsymbol{\rho})$, with successful maximization of $Q\left(\boldsymbol{\rho}|\boldsymbol{\rho}^{(r)}\right)$, the conditional expectation of the complete-data log-likelihood function, $L_c(\boldsymbol{x}|\boldsymbol{\rho})$, given the observation data $\boldsymbol{y}$ and a current fit $\boldsymbol{\rho}^{(r)}$ 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 Q function which is defined by

$$Q\left(\boldsymbol{\rho}|\boldsymbol{\rho}^{(r)}\right) = E_{\boldsymbol{\rho}^{(r)}}\{L_c(\boldsymbol{x}|\boldsymbol{\rho})|\boldsymbol{y}, \boldsymbol{\rho}^{(r)}\}, \tag{7}$$

where the expectation is taken with respect to the conditional distribution, $p(\boldsymbol{\omega}|\boldsymbol{y}, \boldsymbol{\rho}^{(r)})$. The M-step determines a $\boldsymbol{\rho}^{(r)}$ that maximizes $Q(\boldsymbol{\rho}|\boldsymbol{\rho}^{(r)})$. Under some mild regularity conditions, the sequence, $\{\boldsymbol{\rho}^{(r)}\}$, obtained from the EM algorithm iterations converges to the ML estimate, $\widehat{\boldsymbol{\rho}}$. For details about the convergence properties of the sequence, $\boldsymbol{\rho}^{(r)}$, 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 A.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. In Appendix A.1, it is shown that

$$\begin{aligned} D(\boldsymbol{\theta}, \boldsymbol{\theta}_0) &= K[p(\boldsymbol{\omega}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\psi}), p(\boldsymbol{\omega}|\boldsymbol{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi})] \\ &\quad + K[p(\boldsymbol{\omega}|\boldsymbol{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\boldsymbol{\omega}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\psi})]. \end{aligned} \tag{8}$$

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

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

Although it is difficult to interpret the new loss function from the original definition, Eq. (8) suggests that it can be interpreted using the KL divergence between the two posterior densities of the latent variables. Similar to BR, the new loss function may be explained as a measure for the amount of information that the sample observations may be expected to provide in order to discriminate between $p(\boldsymbol{\omega}|y, \boldsymbol{\theta}, \boldsymbol{\psi})$ and $p(\boldsymbol{\omega}|y, \boldsymbol{\theta}_0, \boldsymbol{\psi})$. It seems reasonable to use this KL divergence as the loss function for the latent variable models.

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

$$\begin{aligned} T(\boldsymbol{y}, \boldsymbol{\theta}_0) &= E_{(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{y})}\{Q(\boldsymbol{\theta}, \boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \\ &\quad + Q(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) - Q(\boldsymbol{\theta}, \boldsymbol{\theta}_0)\}. \end{aligned} \tag{9}$$

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 A.2.

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

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

Remark 4.3. If the 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 and the approach of BR.

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 Eq. (9) to (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 the Jeffreys 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 improper 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 A.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 A.4, we show how to approximate the second term by treating the nuisance parameter $\boldsymbol{\psi}$ as an additional latent variable, so that $T(\boldsymbol{y}, \boldsymbol{\theta}_0)$ 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 BR, we use the following decision rule:

$$\text{Accept } H_0 \text{ if } T(\boldsymbol{y}, \boldsymbol{\theta}_0) \leq \delta; \qquad \text{Reject } H_0 \text{ if } T(\boldsymbol{y}, \boldsymbol{\theta}_0) > \delta,$$

where δ 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

Table 1
The threshold values based on the probability.

$q(C)$	50%	55%	65%	75%	85%	95%	98%	99%	99.5%	99.9%
$2C$	0	0.010	0.094	0.288	0.732	1.660	2.544	3.220	3.918	5.522

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. For the new loss function developed in the present paper, the threshold value is the sum of the two KL divergence functions, as in (8). Hence, we use $\delta = 2C$ as the threshold value in this paper. Using Eq. (12), we tabulate some threshold values in Table 1.

While the choice of a threshold value is up to the user, in this paper we choose $q(C) = 0.99$ so that the two probability distributions are distinctly different. Hence, we use 3.22 as the threshold value in the present paper. Obviously, other threshold values are possible. 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.

Remark 4.9. In the Bayesian framework, Bayesian credible intervals may be used for hypothesis testing. In practice, however, this approach has several problems. First, for any given coverage probability, there is no unique Bayesian credible interval. This is an important concern as the posterior distributions are often asymmetric. Although one may use the highest probability density (HPD) regions, the derivation of the HPD intervals require intensive numerical calculations. Moreover, there is no way to derive a marginal HPD region from a joint HPD region. In addition, as argued in Lindley (1965) and Kim (1991), the usage of a HPD region for the purpose of hypothesis testing "is appropriate only for circumstances in which prior knowledge of the hypothesized parameter is vague or diffuse". Second, Bayesian credible intervals are not robust to a one-to-one transformation of parameters.

5. Empirical illustrations

In this section, we first illustrate the proposed method using an asset pricing model with a heavy-tailed distribution for which the Q function is available analytically. We then 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]

5.1. Testing asset pricing models under multivariate t

Asset pricing theory is a pillar in modern finance. Under the Bayesian framework, Bayesian analysis of the asset pricing theory has attracted a considerable amount of attentions in the empirical asset pricing literature. For example, Avramov and Zhou (2010) is an excellent review of the literature on Bayesian portfolio analysis. The asset pricing test is an important topic in asset pricing theory. Various econometric approaches have been proposed to check the validity of various asset pricing models. For example, Bayesian tests have been proposed by Shanken (1987), Harvey and Zhou (1990), McCulloch and Rossi (1991), and Geweke and Zhou (1996), etc. In addition, under the frequentist framework, Gibbons et al. (1989) developed a multivariate finite sample test. All 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,

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

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,

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

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

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

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

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

where $\boldsymbol{1}_N$ 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 (14) can be alternatively specified as:

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

By treating ω_t as a latent variable, the powerful EM algorithm can be used to obtain the Q function. Hence, one can obtain the

[5] In an earlier version, Li and Yu (2010) examined the finite sample performances of the proposed method in the same context and found strong evidence to support the method in both cases.

proposed Bayesian test statistic. The detail of the derivation is as follows.

Let $\boldsymbol{R} = \{R_1, R_2, \ldots, R_T\}, \boldsymbol{\omega} = \{\omega_1, \omega_2, \ldots, \omega_T\}, \boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\Phi}, \nu)$. In the present paper, we consider $\boldsymbol{\Phi}$ as a diagonal matrix. The observed data log-likelihood function, $L_o(\boldsymbol{R}|\boldsymbol{\alpha}, \boldsymbol{\psi})$, is:

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

where C is a remainder, independent of the parameters of interest. Based on the normal-gamma mixture representation for the multivariate t distribution, the complete log-likelihood, $L_c(\boldsymbol{R}, \boldsymbol{\omega}|\boldsymbol{\alpha}, \boldsymbol{\psi})$, can be expressed as

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

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

$$E(\boldsymbol{\omega}_t|\boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{R}_t) = \frac{\nu+N}{\nu+\sum_{i=1}^{N}\phi_{ii}^{-1}(R_{it}-\alpha_i-\boldsymbol{\beta}_i\boldsymbol{F}_t)^2},$$

$t = 1, 2, \ldots, T$.

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

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

and that

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

Therefore, the Bayesian test statistic is given by,

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

When $K = 1$ in Eq. (14), the model becomes the single-factor market price model given by:

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

where R_{Mt} is the excess return of the market, and ϵ_{it} is independent over i. In the first empirical study, we test the market price model.

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 1926 to December 2009, so that $N = 25$, $T = 1002$. The data are freely available from the data library of Kenneth French.[6]

Before estimating the model, we need to check the normality assumption. Young (1993) considered some Bayesian diagnostic measures and showed that the ratio of the expectations of the posteriors of the slope and variance and the expectation of the ratio give similar results to the Shapiro–Wilk statistic (Shapiro and Wilk, 1965) when non-informative priors are used. When calculating the Shapiro–Wilk statistic, we found overwhelming evidence against normality. Consequently, we replace normality with the t distribution. Since the Q function is known analytically, it is easy to obtain the proposed Bayesian test statistic.[7] In the Bayesian analysis, we specify the vague conjugate prior distributions to represent the prior ignorance, namely,

$$\alpha_i \sim N[0, 100], \qquad \beta_i \sim N[0, 100],$$
$$\phi_{ii}^{-1} \sim \Gamma[0.001, 0.001].$$

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 2. The Bayesian test statistic for $\boldsymbol{\alpha} = 0 \times \mathbf{1}_{25}$ is 19.08, which is more extreme than the difference between $B(0.5)$ and $B(0.999)$. Hence, we conclude that the asset pricing model is strongly rejected.

It is important to emphasize that, although our method is motivated from the case of objective priors, informative priors can be also used in our method. In a recent study, Tu and Zhou (2010) explored a general approach to forming informative priors based on economic objectives and found that the proposed informative priors outperform significantly the objective priors in terms of investment performance. Our method can be used in conjunction with the informative prior specifications.

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:

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

where $t = 1, 2, \ldots, T$, 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 model was

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

[7] Even in this simple case, the approach of BR cannot be implemented because a closed-form expression for $KL[p(\boldsymbol{y}|\boldsymbol{\theta}), p(\boldsymbol{y}|\boldsymbol{\theta}_0)]$ cannot be obtained.

Table 2
Bayesian estimation and standard error of the parameters for the market model with the multivariate t distribution.

Portfolio	α		β		Ψ	
	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

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.

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. As in So and Li (1999), Chib et al. (2002) and Yu (2005), we specify some proper prior distributions for the nuisance parameters:

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

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 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 A.5 shows how to compute $T(\boldsymbol{y}, \phi_0)$ where $\phi_0 = 1$.

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 3 and show that the unit root hypothesis is rejected.

6. 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 and can be interpreted meaningfully using the KL functions. 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.

Table 3
Empirical results for S&P 500.

Model	π	η	τ	ϕ	σ^2	Test
EST	0.0096	0.0504	−0.8130	0.9822	0.0281	6.2675
SE	0.0065	0.0202	0.2597	0.0092	0.0113	NA

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 et al. (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.

Appendix

A.1. Proof of Lemma 4.1

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

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

It follows that,

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

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

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

and the three properties stated in Lemma 4.1 naturally follow.

A.2. Proof of Theorem 4.1

From Lemma 4.1, we have.

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

Hence, the Bayesian test statistic can be expressed as:

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

It can be shown that,

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

which proves Theorem 4.1.

A.3

In this Appendix, we will show that the proposed statistic, $T(\boldsymbol{y},\boldsymbol{\theta}_0)$, 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,

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

Hence, $p(\boldsymbol{\omega},\boldsymbol{\psi},\boldsymbol{\theta}|\boldsymbol{y},H_k)$ is independent on A_k, B_k. Similarly, we can show that $p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y},H_k)$ and $p(\boldsymbol{\omega}|\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{\psi},H_k)$ are also independent on A_k, B_k. Furthermore, from Appendices A.1 and A.2, we have,

$$\begin{aligned} &E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{\log\frac{p(\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}{p(\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} \\ &= E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{\log\frac{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})p(\boldsymbol{\psi},\boldsymbol{\theta})}{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)p(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} \\ &= E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{\log\frac{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})A_1B_1 f(\boldsymbol{\psi},\boldsymbol{\theta})}{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)A_0 f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} \\ &= E_{(\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{\log\frac{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})f(\boldsymbol{\psi},\boldsymbol{\theta})}{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}\right\} + \log\frac{A_1B_1}{A_0} \\ &E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{E_{(\boldsymbol{\omega}|\boldsymbol{y},\boldsymbol{\theta}_0,\boldsymbol{\psi})}\left[\log\frac{p(\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}\right]\right\} \\ &= E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{E_{(\boldsymbol{\omega}|\boldsymbol{y},\boldsymbol{\theta}_0,\boldsymbol{\psi})}\left[\log\frac{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)p(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})p(\boldsymbol{\psi},\boldsymbol{\theta})}\right]\right\} \\ &= E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{E_{(\boldsymbol{\omega}|\boldsymbol{y},\boldsymbol{\theta}_0,\boldsymbol{\psi})}\left[\log\frac{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)A_0 f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})A_1B_1 f(\boldsymbol{\psi},\boldsymbol{\theta})}\right]\right\} \\ &= E_{(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}\left\{E_{(\boldsymbol{\omega}|\boldsymbol{y},\boldsymbol{\theta}_0,\boldsymbol{\psi})}\left[\log\frac{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta}_0)f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\boldsymbol{y},\boldsymbol{\omega}|\boldsymbol{\psi},\boldsymbol{\theta})f(\boldsymbol{\psi},\boldsymbol{\theta})}\right]\right\} + \log\frac{A_0}{A_1B_1}. \end{aligned}$$

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

A.4

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

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

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

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

and

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

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

$$f(\boldsymbol{\theta}) \approx f(\boldsymbol{\theta}_0) + \dot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\frac{\ddot{f}(\boldsymbol{\theta}_0)}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0).$$

It follows that,

$$\begin{aligned} &E_{(\boldsymbol{\theta}|\boldsymbol{y})}\left\{E_{(\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{y},\boldsymbol{\theta}_0)}\left\{\log\frac{p(\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta}_0)}{p(\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{\psi}|\boldsymbol{\theta})}\right\}\right\} \\ &= E_{(\boldsymbol{\theta}|\boldsymbol{y})}\{f(\boldsymbol{\theta}_0) - f(\boldsymbol{\theta})\} \\ &\approx \int_\Theta\left\{-\dot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\frac{\ddot{f}(\boldsymbol{\theta}_0)}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right\} p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} \\ &= E_{(\boldsymbol{\theta}|\boldsymbol{y})}\left\{-\dot{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\frac{\ddot{f}(\boldsymbol{\theta}_0)}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right\}. \end{aligned}$$

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

$$\dot{f}(\boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \int_{\Omega} \frac{\partial \log p(\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{\psi}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\boldsymbol{\omega}, \boldsymbol{\psi}|\boldsymbol{y}, \boldsymbol{\theta}_0) d\boldsymbol{\omega}$$

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

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

A.5. Calculation of $T(y, \phi_0)$ for the SV model

Let $\boldsymbol{y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T\}$, $\boldsymbol{h} = \{h_1, h_2, \ldots, h_T\}$, $\boldsymbol{s} = \{s_1, s_2, \ldots, s_T\}$, $\boldsymbol{q} = \{q_1, q_2, \ldots, q_T\}$. The joint density function is:

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

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

$$\begin{aligned}
&\mathcal{L}_o(\boldsymbol{y}|\pi, \eta, \tau, \phi, \sigma^2) \\
&= \log \left\{ \int p(\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{s}, \boldsymbol{q}|\pi, \eta, \tau, \phi, \sigma^2) d\boldsymbol{h} d\boldsymbol{s} d\boldsymbol{q} \right\}.
\end{aligned}$$

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 A.4 can be used to approximate the Bayesian test statistic, $T(\boldsymbol{y}, \phi_0)$. In this case, ϕ is the parameter of interest and $\pi, \eta, \tau, \sigma^2$ are the nuisance parameters. Following Appendix A.4, we treat the nuisance parameters as latent variables. Based on the prior distributions specified in Section 5.2, we get

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

To compute the Bayes test statistic, several components are required. For example,

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

It follows that,

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

Moreover,

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

References

Avramov, D., Zhou, G.F., 2010. Bayesian portfolio analysis. Annual Review of Financial Economics 3, 25–47.

Bernardo, J.M., Rueda, R., 2002. Bayesian hypothesis testing: a reference approach. International Statistical Review 70, 351–372.

Berg, A., Meyer, R., Yu, J., 2004. Deviance information criterion for comparing stochastic volatility models. Journal of Business and Economic Statistics 22, 107–120.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, second ed. Spring-Verlag.

Berger, J.O., Bernardo, J.M., 1992. On the development of reference priors. In: Bernardo, J.M., Berger, J.O., Lindley, D.V., Smith, A.F.M. (Eds.), Bayesian Statistics 4. Oxford University Press, Oxford, pp. 35–60.

Berger, J.O., 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., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis-Hastings output. Journal of the American Statistical Association 96, 270–281.

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

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

Douc, R., Moulines, E., Olsson, J., van Handel, R., 2011. Consistency of the maximum likelihood estimator for general hidden Markov models. Annals of Statistics 39, 474–513.

Gibbons, M.R., Ross, S.A., 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., van Dijk, H.K., 2011. Handbook of Bayesian Econometrics. Oxford University Press, Oxford.

Geweke, J., Zhou, G., 1996. Measuring the pricing errors of the arbitrage pricing theory. Review of Financial Studies 9, 557–587.

Harvey, C., Zhou, G., 1990. Bayesian inference in asset pricing tests. Journal of Financial Economics 26, 221–254.

Ibrahim, J.G., Zhu, H.T., 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, third ed. Oxford University Press, Oxford.

Kan, R., Zhou, G.F., 2006. Modeling non-normality using multivariate t: implications for asset pricing. Working paper, University of Toronto.

Kass, R.E., Raftery, A.E., 1995. Bayes factor. Journal of the Americana Statistical Association 90, 773–795.

Kim, D., 1991. A Bayesian significance test of the stationarity of regression parameters. Biometrika 78, 667–675.

Li, Y., Yu, J., 2010. Bayesian hypothesis testing in latent variable models, working paper, Sim Kee Boon Institute for Financial Economics, Singapore Management University, CoFiE-WP-11-2010.

Lindley, D.V., 1965. Introduction to Probability and Statistics from a Bayesian Viewpoint, vol. 2. Cambridge University Press.

McCulloch, R.E., 1989. Local model influence. Journal of the American Statistical Association 84, 473–478.
McCulloch, R.E., 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.
Poirier, D.J., 1997. A predictive motivation for loss function specification in parametric hypothesis testing. Economic letters 56, 1–3.
Raftery, A., Lewis, S., 1992. In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), How Many Iterations in the Gibbs Sampler? In: Bayesian Statistics, vol. 4. Claredon Press, Oxford, pp. 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 University Press, Oxford.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality. Biometrika 52, 591–611.
Soofi, E.S., Ebrahimi, N., 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., Li, W.K., 1999. Bayesian unit-root testing in stochastic volatility models. Journal of Business and Economic Statistics 17, 491–496.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit with discussion. Journal of the Royal Statistical Society, Series B 64, 583–639.
Tanner, T.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–549.
Tu, J., Zhou, G.F., 2010. Incorporating economic objectives into Bayesian priors: portfolio choice under parameter uncertainty. Journal of Financial and Quantitative Analysis 45, 959–986.
Young, K.D.S., 1993. Bayesian diagnostics for checking assumptions of normality. Journal of Statistical Computation and Simulation 47, 167–180.
Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.
Zhou, G.F., 1993. Asset pricing testing under alternative distributions. Journal of Finance 5, 1927–1942.