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:

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

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.