Author's personal copy

Journal of Econometrics 178 (2014) 602–612

Contents lists available at ScienceDirect

Journal of Econometrics

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

A new approach to Bayesian hypothesis testing☆

Yong Li [a], Tao Zeng [b], Jun Yu [c,*]

[a] Hanqing Advanced Institute of Economics and Finance, Renmin University of China, 100872, Beijing, PR China

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

[c] 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 8 April 2013
Received in revised form
6 August 2013
Accepted 29 August 2013
Available online 7 September 2013

JEL classification:
C11
C12
G12

Keywords:
Bayes factor
Decision theory
EM algorithm
Deviance
Markov chain Monte Carlo
Latent variable models

ABSTRACT

In this paper a new Bayesian approach is proposed to test a point null hypothesis based on the deviance in a decision-theoretical framework. The proposed test statistic may be regarded as the Bayesian version of the likelihood ratio test and appeals in practical applications with three desirable properties. First, it is immune to Jeffreys' concern about the use of improper priors. Second, it avoids Jeffreys–Lindley's paradox, Third, it is easy to compute and its threshold value is easily derived, facilitating the implementation in practice. The method is illustrated using some real examples in economics and finance. It is found that the leverage effect is insignificant in an exchange time series and that the Fama–French three-factor model is rejected.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Hypothesis testing plays a fundamental role in making statistical inference about the model specification. After models are estimated, empirical researchers would often like to test a relevant hypothesis to look for evidence to support or to be against a particular theory. An important class of hypotheses involve a single parameter value in the null.

In this paper we are concerned about testing a single point hypothesis under Bayesian paradigm. So far Bayes factor (BF) is the dominant statistic for Bayesian hypothesis testing (Kass and Raftery, 1995; Geweke, 2007). The wide range of applicability of BF comes with no surprise. BF computes the posterior odds of the null hypothesis and hence provides a general and intuitive way to evaluate the evidence in favor of the null hypothesis.

In the meantime, unfortunately, BF also suffers from several theoretical and practical difficulties. First, when improper prior distributions are used, BF contains undefined constants and takes arbitrary values. This is known as Jeffreys' concern (Kass and Raftery, 1995). Second, when a proper but vague prior distribution with a large spread is used to represent prior ignorance, BF tends to favor the null hypothesis. The problem may persist even when the sample size is large. This is known as Jeffreys–Lindley's paradox (Kass and Raftery, 1995; Poirier, 1995). Third, the calculation of BF generally requires the evaluation of marginal likelihoods. In many models, the marginal likelihoods may be difficult to compute.

Several approaches have been proposed in the literature to deal with Jeffreys' concern and Jeffreys–Lindley's paradox. One simple approach is to split the data into two parts, one as a training set, the other for statistical analysis. The non-informative prior is then updated by the training data, which produces a new proper informative prior distribution for computing BF. This idea is shared by the fractional BF (O'Hagan, 1995), and the intrinsic BF (Berger, 1985). In many practical situations, unfortunately, it is not clear

☆ We would like to thank the co-editor, the associate editor, two referees, Peter Phillips and Yinghui Zhou for helpful comments. Li gratefully acknowledges the financial support of the Chinese Natural Science fund (No. 71271221) and the hospitality during his research visits to Sim Kee Boon Institute for Financial Economics at Singapore Management University. Yu would like to acknowledge the financial support from Singapore Ministry of Education Academic Research Fund Tier 2 under the grant number MOE2011-T2-2-096.

* 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 © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.08.035

Author's personal copy

how to split the sample. Moreover, the sample split may have a major impact on statistical inference. Without a need to split the sample, several Bayesian hypothesis testing approaches have been proposed based on the decision theory. Noting that the BF approach to Bayesian hypothesis testing is a decision problem with a simple zero–one loss function, Bernardo and Rueda (2002) (BR hereafter) and Li and Yu (2012) (LY hereafter) suggested extending the zero–one loss function into continuous loss functions, resulting in Bayesian test statistics that is well defined under improper priors.

The test statistics of BR and LY relies on threshold values. While in theory these threshold values may be calibrated from simulated data generated from the null hypothesis, in practice they are computationally expensive to obtain. Following McCulloch (1989), LY proposed to choose the threshold values based on the Bernoulli distribution. Although this choice makes the determination of threshold values convenient, there are obvious drawbacks. Not only is the choice of the Bernoulli distribution arbitrary, but also are the threshold values independent of the data and the candidate models. Moreover, it is not clear if the test statistic of LY can resolve Jeffreys–Lindley's paradox.

The main purpose of this paper is to develop a new Bayesian hypothesis testing approach for the point null hypothesis testing. The test statistic is based on the Bayesian deviance and constructed in a decision theoretical framework. It can be regarded as the Bayesian version of the likelihood ratio test. We show that the statistic appeals in four aspects. First, it does not suffer from Jeffreys' concern and, hence, can be used under improper priors. Second, it does not suffer from Jeffreys–Lindley's paradox and, hence, can be used under vague priors. Third, it is easy to compute. Finally, the threshold values can be easily determined and are dependent on the data as well as the candidate models.

The paper is organized as follows. Section 2 reviews the Bayesian literature on testing the point null hypothesis from the viewpoint of decision theory. Section 3 develops the new Bayesian test statistic and establishes its properties. Section 4 illustrates the new method by using three real examples in economics and finance. Section 5 concludes the paper. Appendix collects the proof of theoretical results.

2. Point null hypothesis testing: a literature review

2.1. The setup

Denote $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ the vector of observables. Denote $p(\mathbf{y}|\boldsymbol{\vartheta})$ the likelihood function of the observed data. Denote $\pi(\boldsymbol{\vartheta})$ the prior distribution and $p(\boldsymbol{\vartheta}|\mathbf{y})$ the posterior. Suppose that researchers may wish to test a hypothesis, the simplest of which contains only a point which may correspond to the prediction of a theory (Robert, 2001). Denote $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, whose dimension is p, the parameters of interest, and $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, whose dimension is q, the nuisance parameters. So $\boldsymbol{\vartheta} = (\boldsymbol{\theta}, \boldsymbol{\psi})' \in \boldsymbol{\Theta} \times \boldsymbol{\Psi}$. Assume that the observed data, $\mathbf{y} \in \mathbf{Y}$, is described a probabilistic model $M \equiv \{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})\}$. The point null hypothesis is:

$$\begin{cases} H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \end{cases} \tag{1}$$

From the viewpoint of decision theory, the hypothesis testing may be viewed as a decision problem where the action space has two elements, i.e., to accept H_0 (name it d_0) or to reject H_0 (name it d_1). Denote the null model $M_0 \equiv \{p(\mathbf{y}|\theta_0, \boldsymbol{\psi}), \boldsymbol{\psi} \in \boldsymbol{\Psi}\}$, and $M_1 \equiv M$. Suppose a loss is incurred as a function of the actual value of the parameters $(\boldsymbol{\theta}, \boldsymbol{\psi})$ when one accepts H_0 or rejects H_0. Assume the loss function is given by $\{\mathcal{L}[d_i, (\boldsymbol{\theta}, \boldsymbol{\psi})], i = 0, 1\}$. Naturally, one would like to reject H_0 when the expected posterior loss of accepting H_0 is sufficiently larger than the expected posterior loss of rejecting H_0, i.e.,

$$\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = \int_{\Theta}\int_{\Psi} \triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})]p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} > C,$$

where C is a threshold value, $\triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] = \mathcal{L}[d_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] - \mathcal{L}[d_1, (\boldsymbol{\theta}, \boldsymbol{\psi})]$ is the net loss function which can be used to measure the evidence against H_0 as a function of $(\boldsymbol{\theta}, \boldsymbol{\psi})$.

2.2. Bayes factors and the discrete loss function

BF employs the zero–one loss function. In particular, if

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

we can get

$$\begin{aligned} \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = & \int_{\Psi} (-1)\frac{p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})p(\boldsymbol{\psi}|\boldsymbol{\theta}_0)p(\boldsymbol{\theta}_0)}{p(\mathbf{y})}\mathrm{d}\boldsymbol{\psi} \\ & + \int_{\Theta}\int_{\Psi} 1\frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})p(\boldsymbol{\psi}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}, \end{aligned}$$

where $p(\mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\vartheta})\mathrm{d}\boldsymbol{\vartheta}$ is the marginal likelihood. In general, to represent a prior ignorance, an equal probability 0.5 is assigned to H_0 and to H_1. A reasonable prior for $\boldsymbol{\theta}$ with a discrete support at $\boldsymbol{\theta}_0$ is formulated as $p(\boldsymbol{\theta}) = 0.5$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $p(\boldsymbol{\theta}) = 0.5\pi(\boldsymbol{\theta})$ when $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, where $\pi(\boldsymbol{\theta})$ is a prior distribution. Hence, when $C = 0$, the decision criterion is given by:

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

which is equivalent to

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

where BF_{01} is the well-known BF (Kass and Raftery, 1995) and is the ratio of two marginal likelihood values.

When a subjective prior is not available, an objective prior or default prior may be used. Often, $\pi(\boldsymbol{\theta})$ is taken as non-informative priors, such as the Jeffreys or the reference prior (Jeffreys, 1961; Bernardo and Rueda, 2002). These non-informative priors are generally improper, and it follows that $\pi(\boldsymbol{\theta}) = C_0 f(\boldsymbol{\theta})$, where $f(\boldsymbol{\theta})$ is a nonintegrable function, and C_0 is an arbitrary positive constant. In this case, the BF is

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

Clearly, the BF is not well defined since it depends on the arbitrary constant C_0, giving rise to Jeffreys' concern. In addition, if a proper prior is used but has a large variance, the likelihood function may take low values under the alternative hypothesis. This often leads to a smaller marginal likelihood value for the alternative model. Consequently, BF has a tendency to favor H_0, giving rise to Jeffreys–Lindley's paradox; see Poirier (1995) and Robert (2001).

The formulation of BF generally requires a positive probability for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ to be assigned. When $\boldsymbol{\theta}$ is continuous, the prior concentrates a positive probability mass on the single point $\boldsymbol{\theta}_0$. As pointed out by BR, Jeffreys–Lindley's paradox is the consequence of using this non-regular prior structure.

Author's personal copy

2.3. BR and the KL loss function

Instead of using a zero–one loss function, BR (2002) advocated using a continuous function of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ to formulate the loss function. In particular, they suggested using the KL divergence. For any regular probability functions, $p(x)$ and $q(x)$, the KL divergence is defined as:

$$KL[p(x), q(x)] = \int p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x. \tag{2}$$

It can be shown that $KL \geq 0$ for any p and q, and equal to 0 iff $p(x) = q(x)$. In this case, the decision criterion is:

$$\begin{aligned} \mathbf{T}_{BR}(\mathbf{y}, \boldsymbol{\theta}_0) &= \int_{\Theta} \int_{\Psi} \triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} \\ &= \int_{\Theta} \int_{\Psi} \left\{ \int \log \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})} p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \mathrm{d}\mathbf{y} \right\} \\ &\quad \times p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} > C. \end{aligned} \tag{3}$$

To ensure the symmetry, BR suggested using the following net loss function:

$$\begin{aligned} \triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] = \min\{&KL[p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}), p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})], \\ &KL[p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})]\}. \end{aligned} \tag{4}$$

Obviously, $\mathbf{T}_{BR}(\mathbf{y}, \boldsymbol{\theta}_0) = 0$ under the null hypothesis but is positive under the alternative hypothesis. According to BR, this loss function can be used under the reference priors to maintain objectiveness, overcoming Jeffreys' concern. Although the statistic of BR is well defined under improper priors and has other desirable properties, it has certain practical difficulties. First, when the KL loss function is not available analytically, the test statistic of BR is difficult to use, especially when the dimension of the integral in the KL loss is high. Second, threshold values for C, are needed but difficult to find in general.

2.4. LY and the $\mathcal{Q}$ loss function

In the context of latent variable models, the likelihood function and the KL loss are not available analytically and the test statistic of BR is difficult to use. To solve this problem, LY developed a Bayesian test statistic based on the $\mathcal{Q}$ function used in the EM algorithm. Denote $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ the vector of observables and $\mathbf{z} = (z_1, z_2, \ldots, z_n)'$ the vector of latent variables. Let $\mathbf{x} = (\mathbf{y}, \mathbf{z})'$. The latent variable model is dependent on a set of parameters $\boldsymbol{\vartheta}$. Denote $p(\mathbf{y}|\boldsymbol{\vartheta})$ and $p(\mathbf{x}|\boldsymbol{\vartheta})$ the likelihood function of the observed data and the likelihood function of complete data, respectively. The two functions are related to each other by

$$p(\mathbf{y}|\boldsymbol{\vartheta}) = \int p(\mathbf{x}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{z} = \int p(\mathbf{y}, \mathbf{z}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{z}. \tag{5}$$

When the above integral at the right hand side does not have a closed-form solution, instead of using maximum likelihood (ML) method, it is numerically more tractable to carry out Bayesian analysis based on the MCMC algorithm for estimating the latent variable models; see, for example Geweke et al. (2011).

For latent variable models, the complete-data log-likelihood, $L_c(\mathbf{x}|\boldsymbol{\vartheta}) = \log p(\mathbf{x}|\boldsymbol{\vartheta})$, is related to the observed data log-likelihood, $L_o(\mathbf{y}|\boldsymbol{\vartheta}) = \log p(\mathbf{y}|\boldsymbol{\vartheta})$. While $L_c(\mathbf{x}|\boldsymbol{\vartheta})$ is often simple, but $L_o(\mathbf{y}|\boldsymbol{\vartheta}) = \log p(\mathbf{y}|\boldsymbol{\vartheta})$ is often complicated because the integral equation (5) does not have an analytical solution. The EM algorithm is a way to obtain the ML estimator (Dempster et al., 1977). A standard EM algorithm consists of two steps: the *expectation* (E) step and the *maximization* (M) step. The E-step evaluates the $\mathcal{Q}$ function which is defined by:

$$\mathcal{Q}\left(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(r)}\right) = E_{\mathbf{z}}\left\{L_c(\mathbf{x}|\boldsymbol{\vartheta})|\mathbf{y}, \boldsymbol{\vartheta}^{(r)}\right\}, \tag{6}$$

where the expectation is taken with respect to the conditional distribution of latent variables given $\mathbf{y}$ and $\boldsymbol{\vartheta}^{(r)}$, $p(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta}^{(r)})$. The M-step determines a $\boldsymbol{\vartheta}^{(r+1)}$ that maximizes $\mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^{(r)})$.

Let $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\psi})$. LY (2012) introduced a continuous net loss function as:

$$\begin{aligned} \triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] = &\{\mathcal{Q}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}) - \mathcal{Q}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta})\} \\ &+ \{\mathcal{Q}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_0) - \mathcal{Q}(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0)\}, \end{aligned}$$

and proposed a Bayesian test statistic as:

$$\mathbf{T}_{LY}(\mathbf{y}, \boldsymbol{\theta}_0) = E_{\boldsymbol{\vartheta}|\mathbf{y}}[\triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})]]. \tag{7}$$

Like the statistic of BR, the test statistic, $T_{LY}(\mathbf{y}, \boldsymbol{\theta}_0)$, is well defined under improper priors and hence is immune to Jeffreys' concern. Also, it is easy to compute if the MCMC output is available. However, like the statistic of BR, the threshold values are needed in practice. Following McCulloch (1989), LY proposed to base the threshold values on two Bernoulli distributions. Although the use of the threshold values is not new in the Bayesian literature (see, for example, Jeffreys' BF scales), it is awkward that these threshold values are independent of the data and the candidate models. It was remarked in LY that a more natural approach is to obtain threshold values from simulated data in repeated sampling, which is computationally time consuming in general.

3. A new method for Bayesian hypothesis testing

3.1. The test statistic

BR's approach requires the KL loss function must have a closed-form expression and the threshold values for Bayesian hypothesis testing are difficult to obtain. LY's approach is easy to compute, but the threshold values are independent of the data and the candidate models. To avoid these theoretical and computational difficulties, in this section, we introduce a new Bayesian approach for hypothesis testing. Denote the net loss function as:

$$\begin{aligned} \triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] &= -2 \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) - (-2 \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})) \\ &= 2 \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - 2 \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}), \end{aligned} \tag{8}$$

where $-2 \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$ represents the residual information in data $\mathbf{y}$ given $\boldsymbol{\theta}$, $\boldsymbol{\psi}$ in the alternative model. According to Good (1956), $-2 \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$ measures the surprise or uncertainty. Similarly, one can interpret $-2 \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})$. The net loss function is the difference of the two Bayesian deviances, if the Bayesian deviance is defined in the same way as in Spiegelhalter et al. (2002) (Section 2.5). The new Bayesian test statistic is then defined by:

$$\begin{aligned} \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = 2 \int &[\log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})] \\ &\times p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}. \end{aligned} \tag{9}$$

Under the null, $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = 0$, whereas under the alternative, $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) \neq 0$. When the deviance of the null hypothesis is sufficiently smaller than that of the alternative, it is reasonable to believe that we should reject the null hypothesis.

BF essentially compares the relative magnitude of

$$\int_{\Psi} p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}_0) \mathrm{d}\boldsymbol{\psi}$$

and

$$\int_{\Theta} \int_{\Psi} p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi},$$

whereas our test statistic compares the relative magnitude of

$$\int \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} = \int \log p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\psi}$$

Author's personal copy

and

$$\int_{\Theta}\int_{\Psi} \log p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}.$$

Clearly there are two major differences between the two approaches. First, the likelihood functions in BF are replaced with the log-likelihood functions in our test. Second and more importantly, the (log-)likelihood functions are averaged over the prior distributions in BF but over the posterior distributions in our method. The second difference suggests that our statistic is less sensitive to the prior distributions.

The first result in this present paper shows that the Bayes risk of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta_0})$ is just two times the test statistic proposed by BR.

Theorem 3.1. *It can be shown that*

$$E_{\mathbf{y}}\left[\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)\right] = \int \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) p(\mathbf{y}) \mathrm{d}\mathbf{y} = 2E_{\mathbf{y}}\left[\mathbf{T}_{BR}(\mathbf{y}, \boldsymbol{\theta}_0)\right].$$

Remark 3.1. $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ may be explained as the Bayesian version of the likelihood ratio test since it is the likelihood ratio averaged over the posterior distribution under the alternative hypothesis.

Remark 3.2. To show how the new statistic is immune to Jeffreys' concern, consider general improper priors, $p(\boldsymbol{\psi}|\boldsymbol{\theta}) = Af(\boldsymbol{\psi}|\boldsymbol{\theta})$, $p(\boldsymbol{\theta}) = Bf(\boldsymbol{\theta})$, $p(\boldsymbol{\psi}|\boldsymbol{\theta}_0) = C_0 f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)$ where $f(\boldsymbol{\psi}|\boldsymbol{\theta})$, $f(\boldsymbol{\theta})$ and $f(\boldsymbol{\psi}|\boldsymbol{\theta}_0)$ are nonintegrable functions, and A, B, C_0 are arbitrary positive constants. It can be shown that,

$$\begin{aligned} p(\boldsymbol{\psi}, \boldsymbol{\theta}|\mathbf{y}) &= \frac{p(\mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\theta})}{p(\mathbf{y})} = \frac{p(\mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\theta})}{\int\int p(\mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\theta})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{\theta}} \\ &= \frac{p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) p(\boldsymbol{\psi}, \boldsymbol{\theta})}{\int\int p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) p(\boldsymbol{\psi}, \boldsymbol{\theta})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{\theta}} \\ &= \frac{p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) ABf(\boldsymbol{\psi}, \boldsymbol{\theta})}{\int\int p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) ABf(\boldsymbol{\psi}, \boldsymbol{\theta})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{\theta}} \\ &= \frac{p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) f(\boldsymbol{\psi}, \boldsymbol{\theta})}{\int\int p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) f(\boldsymbol{\psi}, \boldsymbol{\theta})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{\theta}}. \end{aligned}$$

Hence, $p(\boldsymbol{\psi}, \boldsymbol{\theta}|\mathbf{y})$ is independent of the arbitrary constants. Similarly, we can show that $p(\boldsymbol{\psi}|\mathbf{y})$ is also independent of C_0. Consequently, $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta_0})$ is well defined under improper priors.

Remark 3.3. To see how the new statistic can avoid Jeffreys–Lindley's paradox, we consider a well known example in the literature; see, for example, Robert (1993). Let $y \sim N(\theta, 1)$. Suppose we want to test the simple point null hypothesis $H_0 : \theta = 0$. The prior distribution of θ can be set as $N(\mu, \tau^2)$ with $\mu = 0$. Then the posterior distribution of θ is $N(\mu(y), \omega^2)$ with

$$\mu(y) = \frac{\mu + \tau^2 y}{1 + \tau^2}, \qquad \omega^2 = \frac{\tau^2}{1 + \tau^2}.$$

BF is given by:

$$BF_{10} = \frac{1}{BF_{01}} = \sqrt{\frac{1}{1 + \tau^2}} \exp\left[\frac{\tau^2 y^2}{2(1 + \tau^2)}\right].$$

As $\tau^2 \to +\infty$, $BF_{10} \to 0$ which means that the test always supports the null hypothesis regardless whether or not it holds true, giving rise to Jeffreys–Lindley's paradox. The reason for the paradox is that BF compares $\int p(y|\theta)p(\theta)\mathrm{d}\theta$ with $p(y|\theta = 0)$. When $p(\theta)$ has a large variance, even if y is far away from 0, there is a fair chance that $p(y|\theta = 0)$ is larger. On the other hand, it is easy to show:

$$\begin{aligned} \mathbf{T}(y, 0) &= 2\left[\int \log p(y|\theta) p(\theta|y)\mathrm{d}\theta - \log p(y|\theta = 0)\right] \\ &= 2y\mu(y) - \mu^2(y) - \omega^2. \end{aligned}$$

Table 1
Using BF and the new test to test $\theta = 0$ when $y = 3$.

τ	1	100	1000
$\log BF_{01}$	−1.90	0.11	2.41
$\mathbf{T}(\mathbf{y}, \theta_0)$	6.25	8.00	8.00

As $\tau^2 \to +\infty$, $\mu(y) \to y$, $\omega^2 \to 1$. In this case, the posterior distribution converges to $N(y, 1)$ and $\mathbf{T}(y, 0) \to y^2 - 1$ which is distributed exactly as $\chi^2(1) - 1$ when H_0 is true. Consequently, our proposed test statistic avoids Jeffreys–Lindley's paradox. Essentially, we compare $\int \log p(y|\theta)\mathrm{d}N(\theta; y, 1)$ with $\log p(y|\theta = 0)$. Since the posterior distribution $N(\theta; y, 1)$ puts much more weight in the area near y, when y is far away from zero, the former quantity should take a much larger value than the latter. To illustrate the point, if $y = 3$ which is 3 standard deviation away under the null hypothesis, we expect a reasonable test should reject the null hypothesis. Table 1 reports $\log BF_{01}$ and $\mathbf{T}(y, 0)$ when $\tau = 1, 100, 1000$. It can be seen that while our method always rejects the null the BF fails to reject the null when $\tau = 100, 1000$.

Remark 3.4. When $p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$ is available in closed-form and the model under alternative hypothesis is estimated by MCMC, it is straightforward to calculate $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta_0})$ by

$$\frac{1}{M}\sum_{m=1}^{M}\left(\log p\left(\mathbf{y}|\boldsymbol{\theta}^{(m)}, \boldsymbol{\psi}^{(m)}\right) - \log p\left(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}^{(m)}\right)\right),$$

where $\{\boldsymbol{\theta}^{(m)}, \boldsymbol{\psi}^{(m)}\}$, $m = 1, 2, \ldots, M$, are the draws, generated by the MCMC technique, from the posterior distribution under the alternative hypothesis.

3.2. Latent variable models

In many cases, $p(\mathbf{y}|\boldsymbol{\vartheta})$ does not have a closed-form expression. For example, in latent variable models, $p(\mathbf{y}|\boldsymbol{\vartheta})$ often involves integrals that cannot not be solved analytically. In this section, we show how to approximate $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta_0})$ with the EM algorithm and the MCMC output. To do so, we first impose the following set of regularity conditions.

Assumption 1. The likelihood of the model considered is regular.

Assumption 2. The data generating process is stationary.

Assumption 3. There exists a finite sample size n^*, so that, for $n > n^*$, there is a local maximum at $\hat{\boldsymbol{\vartheta}}$ such that $L_n^{(1)}(\hat{\boldsymbol{\vartheta}}) = 0$ and $L_n^{(2)}(\hat{\boldsymbol{\vartheta}})$ is negative definite, where $L_n(\boldsymbol{\vartheta}) = \log p(\boldsymbol{\vartheta}|\mathbf{y})$, $L_n^{(1)}(\boldsymbol{\vartheta}) = \partial \log p(\boldsymbol{\vartheta}|\mathbf{y})/\partial\boldsymbol{\vartheta}$, $L_n^{(2)}(\boldsymbol{\vartheta}) = \partial^2 \log p(\boldsymbol{\vartheta}|\mathbf{y})/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'$.

Assumption 4. The largest eigenvalue λ_n of $\left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}})\right]^{-1}$ goes to zero when $n \to \infty$.

Assumption 5. For any $\epsilon > 0$, there exists an integer N and some $\delta > 0$ such that for any $n > \max\{N, n^*\}$ and $\boldsymbol{\vartheta} \in H(\hat{\boldsymbol{\vartheta}}, \delta) = \{\boldsymbol{\vartheta} : \|\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}\| \le \delta\}$, $L_n^{(2)}(\boldsymbol{\vartheta})$ exists and satisfies

$$-A(\epsilon) \le L_n^{(2)}(\boldsymbol{\vartheta}) L_n^{-(2)}(\hat{\boldsymbol{\vartheta}}) - \mathbf{I}_{p+q} \le A(\epsilon),$$

where $\mathbf{I}_{p+q}$ is an identity matrix and $A(\epsilon)$ is a positive semidefinite symmetric matrix whose largest eigenvalue goes to zero as $\epsilon \to 0$. When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, this assumption also holds.

Assumption 6. For any $\delta > 0$, as $n \to \infty$,

$$\int_{\boldsymbol{\Omega} - H(\hat{\boldsymbol{\vartheta}}, \delta)} p(\boldsymbol{\vartheta}|\mathbf{y})\mathrm{d}\boldsymbol{\vartheta} \to 0,$$

where $\boldsymbol{\Omega}$ is the support space of $\boldsymbol{\vartheta}$.

Author's personal copy

Assumption 7. For any $\delta > 0$, when $\boldsymbol{\vartheta} \in H(\hat{\boldsymbol{\vartheta}}, \delta)$, $L_n^{(2)}(\boldsymbol{\vartheta})/n = O_p(1)$.

Remark 3.5. These assumptions are mild regularity conditions and have been used in the literature to develop Bayesian large sample theory; see, for example, Chen (1985), Kim (1994, 1998) and Geweke (2005). Based on these regularity conditions, Li et al. (2012) showed that, conditional on the observed data $\mathbf{y}$,

$$\bar{\boldsymbol{\vartheta}} = E[\boldsymbol{\vartheta}|\mathbf{y}, H_1] = \int \boldsymbol{\vartheta} p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}} + o(n^{-1/2}),$$

$$V(\hat{\boldsymbol{\vartheta}}) = -L_n^{-(2)}(\hat{\boldsymbol{\vartheta}}) + o(n^{-1}),$$

where

$$V(\tilde{\boldsymbol{\vartheta}}) = E\left[(\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}})'|\mathbf{y}, H_1\right] = \int (\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \tilde{\boldsymbol{\vartheta}})' p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}.$$

Theorem 3.2. *Let $\bar{\boldsymbol{\vartheta}} = (\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\psi}})'$ be the posterior mean of $\boldsymbol{\vartheta}$ under H_1, $\bar{\boldsymbol{\vartheta}}_* = (\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}})'$, $\bar{\boldsymbol{\vartheta}}_b = (1-b)\bar{\boldsymbol{\vartheta}}_* + b\bar{\boldsymbol{\vartheta}}$, $b \in [0,1]$, $S(\mathbf{x}|\boldsymbol{\vartheta}) = \partial \log p(\mathbf{x}|\boldsymbol{\vartheta})/\partial \boldsymbol{\vartheta}$,*

$$D = \int_0^1 \left\{(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \left[E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b} S_1(\mathbf{x}|\bar{\boldsymbol{\vartheta}}_b)\right]\right\} \mathrm{d}b, \tag{10}$$

where $S_1(\mathbf{x}|\boldsymbol{\vartheta})$ is the subvector of $S(\mathbf{x}|\boldsymbol{\vartheta})$ corresponding to $\boldsymbol{\theta}$. Let

$$\begin{aligned}\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta}_0) &= 2D + 2\left[\log p(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\psi}}) - \log p(\bar{\boldsymbol{\psi}}|\boldsymbol{\theta}_0)\right] \\ &\quad - 2\left[\int \log p(\boldsymbol{\theta}|\boldsymbol{\psi}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}\right] \\ &\quad - \left[p + q - \mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\psi}}) V_{22}(\bar{\boldsymbol{\vartheta}})]\right]\end{aligned} \tag{11}$$

where $V_{22}(\bar{\boldsymbol{\vartheta}}) = E[(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})'|\mathbf{y}, H_1]$, which is the submatrix of $V(\bar{\boldsymbol{\vartheta}})$ corresponding to $\boldsymbol{\psi}$ and

$$L_{0n}^{(2)}(\boldsymbol{\psi}) = \frac{\partial^2 \log p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}.$$

Under Assumptions 1–7, it can be shown that

$$\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta}_0) = \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) + o_p(1). \tag{12}$$

Remark 3.6. According to (12) we can approximate $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ by $\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta}_0)$.

Remark 3.7. In many cases, the analytical form of D is not available. Following Gelman and Meng (1998), if D does not have a closed form expression, we can numerically approximate it using the trapezoidal rule. In particular, we can choose a set of fixed grids $\{b_{(s)}\}_{s=0}^{S}$ such that $b_0 = 0 < b_{(1)} < b_{(2)} < \cdots < b_{(S)} < b_{(S+1)} = 1$, and then approximate D by

$$\begin{aligned}\widehat{D} &= \frac{1}{2}\left(\bar{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0\right)' \sum_{s=0}^{S} \left(b_{(s+1)} - b_{(s)}\right) \\ &\quad \times \left(E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_{b_{(s)}}} \left[S_1\left(\mathbf{x}|\bar{\boldsymbol{\vartheta}}_{b_{(s)}}\right)\right] + E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_{b_{(s+1)}}} \left[S_1\left(\mathbf{x}|\bar{\boldsymbol{\vartheta}}_{b_{(s+1)}}\right)\right]\right).\end{aligned} \tag{13}$$

To calculate $E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_{b_{(s)}}} \left[S_1\left(\mathbf{x}|\bar{\boldsymbol{\vartheta}}_{b_{(s)}}\right)\right]$, we use

$$\begin{aligned}E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_{b_{(s)}}} \left[S_1\left(\mathbf{x}|\bar{\boldsymbol{\vartheta}}_{b_{(s)}}\right)\right] &= E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_{b_{(s)}}} \left[S_1\left(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_{b_{(s)}}\right)\right] \\ &\approx M^{-1} \sum_{m=1}^{M} S_1\left(\mathbf{y}, \mathbf{z}^{(m)}|\bar{\boldsymbol{\vartheta}}_{b_{(s)}}\right),\end{aligned}$$

where $\{\mathbf{z}^{(m)}, m = 1, 2, \ldots, M\}$ are efficient random observations simulated from $p\left(\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_{b_{(s)}}\right)$ with $\bar{\boldsymbol{\vartheta}}_{b_{(s)}} = (1 - b_{(s)})\bar{\boldsymbol{\vartheta}} + b_{(s)}\bar{\boldsymbol{\vartheta}}_*$ after discarding some burn-in samples. With D being replaced by $\widehat{D}$ in (13), we can approximate $\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta}_0)$ by $\hat{\mathbf{T}}_1(\mathbf{y}, \boldsymbol{\theta}_0)$.

Remark 3.8. The test statistic clearly requires the evaluation of the observed information matrix, the second derivative of the observed-data likelihood function. For most latent variable models, the observed-data likelihood function does not have a closed-from expression so that the second derivatives are difficult to evaluate. It is noted that

$$L_n^{(2)}(\boldsymbol{\vartheta}) = \frac{\partial^2 L_o(\mathbf{y}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} + \frac{\partial^2 p(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'}.$$

In the EM algorithm, under the mild regularity conditions, if $\mathcal{Q}(\cdot|\cdot)$ has a closed form expression, Oakes (1999) showed that the observed information matrix can be expressed as:

$$\mathbf{I}(\boldsymbol{\vartheta}) = -\frac{\partial^2 L_o(\mathbf{y}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} = \left\{-\frac{\partial^2 \mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 \mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^{*\prime}}\right\}_{\boldsymbol{\vartheta}^* = \boldsymbol{\vartheta}}. \tag{14}$$

When $\mathcal{Q}(\cdot|\cdot)$ does not have a closed form expression, Louis (1982) derived the observed information matrix as:

$$\begin{aligned}\mathbf{I}(\boldsymbol{\vartheta}) &= E_{(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta})}\left\{-\frac{\partial^2 L_c(\mathbf{x}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'}\right\} - Var_{(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta})}\{S(\mathbf{x}|\boldsymbol{\vartheta})\} \\ &= E_{(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta})}\left\{-\frac{\partial^2 L_c(\mathbf{x}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - S(\mathbf{x}|\boldsymbol{\vartheta}) S(\mathbf{x}|\boldsymbol{\vartheta})'\right\} \\ &\quad + E_{(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta})}\{S(\mathbf{x}|\boldsymbol{\vartheta})\} E_{(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta})}\{S(\mathbf{x}|\boldsymbol{\vartheta})\}',\end{aligned} \tag{15}$$

where the expectations are taken with respect to the conditional distribution of $\mathbf{z}$ given $\mathbf{y}$ and $\boldsymbol{\vartheta}$. Hence, the information matrix can be approximated by:

$$\begin{aligned}&E_{(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta})}\left\{-\frac{\partial^2 L_c(\mathbf{x}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - S(\mathbf{x}|\boldsymbol{\vartheta}) S(\mathbf{x}|\boldsymbol{\vartheta})'\right\} \\ &\quad \approx -\frac{1}{M} \sum_{m=1}^{M} \left\{\frac{\partial^2 L_c\left(\mathbf{y}, \mathbf{z}^{(m)}|\boldsymbol{\vartheta}\right)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} + S\left(\mathbf{y}, \mathbf{z}^{(m)}|\boldsymbol{\theta}\right) S\left(\mathbf{y}, \mathbf{z}^{(m)}|\boldsymbol{\vartheta}\right)'\right\}, \\ &E_{(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta})}\{S(\mathbf{x}|\boldsymbol{\vartheta})\} \approx \frac{1}{M} \sum_{m=1}^{M} S\left(\mathbf{y}, \mathbf{z}^{(m)}|\boldsymbol{\vartheta}\right),\end{aligned}$$

where $\{\mathbf{z}^{(m)}, m = 1, 2, \ldots, M\}$ are the efficient random draws from the conditional distribution $p(\mathbf{z}|\mathbf{y}, \boldsymbol{\vartheta})$.

3.3. Choosing threshold values

To implement the proposed method, we need to specify a threshold value. We shall use the following decision rule to test the hypothesis:

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

where C is the threshold value to be specified. The following theorem gives the asymptotic distribution of the test statistic. The threshold value can be then set to be a certain percentile of the asymptotic distribution. This compares favorably with Jeffreys' subjective threshold values for BF (Jeffreys, 1961) and the threshold values used in LY.

Theorem 3.3. *When the likelihood information dominates the prior information, under Assumptions 1–7, we have, under the null hypothesis*

Author's personal copy

$$\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)+\left[p+q-\mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\vartheta}})V_{22}(\bar{\boldsymbol{\vartheta}})]\right]$$
$$\overset{a}{\sim}\boldsymbol{\epsilon}'\left[\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\right]\boldsymbol{\epsilon}, \tag{16}$$
$$\mathbf{T}_1(\mathbf{y},\boldsymbol{\theta}_0)+\left[p+q-\mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\vartheta}})V_{22}(\bar{\boldsymbol{\vartheta}})]\right]$$
$$\overset{a}{\sim}\boldsymbol{\epsilon}'\left[\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\right]\boldsymbol{\epsilon}, \tag{17}$$

where $\boldsymbol{\epsilon}$ *is a standard multivariate normal variate,* $\boldsymbol{\vartheta}_0=(\boldsymbol{\theta}_0,\boldsymbol{\psi}_0)$ *the true value of* $\boldsymbol{\vartheta}$, $\mathbf{J}(\boldsymbol{\vartheta}_0)$ *the Fisher information matrix given by*

$$\mathbf{J}(\boldsymbol{\vartheta}_0)=\frac{1}{n}\int -L_n^{(2)}(\boldsymbol{\vartheta}_0)p(\mathbf{y}|\boldsymbol{\vartheta}_0)\mathrm{d}\mathbf{y},$$

$\mathbf{IJ}(\boldsymbol{\vartheta}_0)$ *the inverse of* $\mathbf{J}(\boldsymbol{\vartheta}_0)$, $\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)$ *and* $\mathbf{IJ}_{11}(\boldsymbol{\vartheta}_0)$ *the submatrices of* $\mathbf{J}(\boldsymbol{\vartheta}_0)$ *and* $\mathbf{IJ}(\boldsymbol{\vartheta}_0)$, *respectively, corresponding to* $\boldsymbol{\theta}$.

Remark 3.9. In general, the asymptotic distributions of $\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)$ and $\mathbf{IJ}_{11}(\boldsymbol{\vartheta}_0)$ are not known. However, both $\mathbf{J}(\boldsymbol{\vartheta}_0)$ and $\mathbf{IJ}(\boldsymbol{\vartheta}_0)$ can be consistently estimated by

$$\mathbf{J}(\boldsymbol{\vartheta}_0)\approx-\frac{1}{n}L_n^{(2)}(\bar{\boldsymbol{\vartheta}}),\qquad \mathbf{IJ}(\boldsymbol{\vartheta}_0)\approx nV(\bar{\boldsymbol{\vartheta}}).$$

This greatly facilitates the calculation of the asymptotic distribution.

Remark 3.10. To obtain the asymptotic distribution and the threshold values, since the middle term in the asymptotic distribution, $\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)$, only depends on the model and the data, one only needs to simulate from the standard multivariate normal.

In some cases, there is no need to simulate the asymptotic distributions of $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)$ and $\mathbf{T}_1(\mathbf{y},\boldsymbol{\theta}_0)$. The following theorem gives such a situation.

Theorem 3.4. *If* $\boldsymbol{\theta}$ *and* $\boldsymbol{\psi}$ *are orthogonal,* $\mathbf{tr}\left[\mathbf{J}_{22}(\boldsymbol{\vartheta}_0)\mathbf{IJ}_{22}(\boldsymbol{\vartheta}_0)\right]=q$, $\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)=\mathbf{I}_p$, $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)\overset{a}{\sim}\chi^2(p)-p$, *and* $T_1(\mathbf{y},\boldsymbol{\theta}_0)\overset{a}{\sim}\chi^2(p)-p$, *where* $\mathbf{J}_{22}(\boldsymbol{\vartheta}_0)$ *and* $\mathbf{IJ}_{22}(\boldsymbol{\vartheta}_0)$ *are the submatrices of* $\mathbf{J}(\boldsymbol{\vartheta}_0)$ *and* $\mathbf{IJ}(\boldsymbol{\vartheta}_0)$ *corresponding to* $\boldsymbol{\psi}$.

Remark 3.11. Theorem 3.4 can be simply derived from Theorem 3.3. While the likelihood ratio statistic asymptotically follows $\chi^2(p)$ and is always positive, the Bayesian version of the likelihood ratio statistic proposed in the present paper asymptotically follows $\chi^2(p)-p$. The mean of the asymptotic distribution is zero and hence it is possible that our statistic takes a negative value, a property shared by the logarithmic BF. This is not surprising as both the new statistic and the BF are obtained by integrating over the parameter space rather than by maximizing over the parameter space.

Remark 3.12. It is well known in the literature that BFs are conservative compared to the likelihood ratio (LR) test; see, for example, Edwards et al. (1963) and Kass and Raftery (1995). It is important to point out that the LR test, like other frequentist's tests, is conducted based on the following Fisher's scale. If the critical level is between 95% and 97.5%, the evidence for the alternative is "moderate"; between 97.5% and 99%, "substantial"; between 99% and 99.5%, "strong"; between 99.5% and 99.9%, "very strong"; larger than 99.9%, "overwhelming". Inferences based on BFs use Jeffreys' scale instead. If $2\log BF_{10}$ is less than 0, there is "negative" evidence for the alternative; between 0 and 2, "not worth more than a bare mention"; between 2 and 6, "positive"; between 6 and 10, "strong"; larger than 10, "very strong". To show the difference between our statistic and the LR test as well as BFs, let $y_1,\ldots,y_n\overset{\text{i.i.d.}}{\sim}N(\theta,1)$. The prior distribution of θ can be set as $N(0,\tau^2)$. We want to test the simple point null hypothesis $H_0:\theta=0$. Suppose $\bar{y}=\frac{1}{n}\sum_{i=1}^n y_i=\sqrt{6.634897/n}$ so that the critical level of the LR test is always kept at 99%. According to Fisher's scale, we have "strong" evidence for the alternative when using LR. In this case, it can be shown that $2\log BF_{10}=\frac{\tau^2n^2\bar{y}^2}{n\tau^2+1}-\log(n\tau^2+1)$, $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)=nw^2(2-nw^2)n\bar{y}^2-nw^2$, where $w^2=\frac{\tau^2}{n\tau^2+1}$. Table 2 gives the values of $2\log BF_{10}$, $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)+1$, *LR*, and the decision from BFs according to Jeffreys' scale, when $\tau=1$. It can be seen that BFs find the evidence for the alternative hypothesis to be "positive" when $n=10$. The evidence turns to be "not worth more than a bare mention" when $n=100$, but to "negative" when $n=1000$, 10,000. This result is consistent with the conservative property of BFs relative to LR. In the meantime, our test statistic is slightly more conservative than LR although the difference is small and they converge to each other as the sample size grows.

Table 2
Comparison of $2\log BF_{10}$, $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)+1$, and *LR* when the prior distribution of θ is $N(0,1)$ and $\bar{y}=\sqrt{6.634897/n}$ so that the critical level of LR is always 99%.

n	10	100	1000	10,000
$2\log BF_{10}$	3.63383	1.95408	−0.28049	−2.57621
Decision	Positive	Not worth mention	Negative	Negative
$\mathbf{T}(\mathbf{y},\theta_0)+1$	6.67097	6.64415	6.63589	6.63500
LR	6.63490	6.63490	6.63490	6.63490

4. Examples

In this section, we illustrate the proposed theory using three examples in economics and finance. In the first example, we compare the performance of BF and that of $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)$ in the context of simple linear regression model, aiming to explore the presence of Jeffreys–Lindley's paradox in BF and the absence of Jeffreys–Lindley's paradox in the proposed method. In the second example, we check the quality of the approximation of $\mathbf{T}_1(\mathbf{y},\boldsymbol{\theta}_0)$ and $\hat{\mathbf{T}}_1(\mathbf{y},\boldsymbol{\theta}_0)$ to $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)$ in the context of linear asset pricing model. In this case both the observed-data log-likelihood and the complete-data log-likelihood have the analytical form. In the third example, we test the presence of leverage effect in a stochastic volatility (SV) model. Since the observed-data log-likelihood is not available in closed-form for the SV model, only $\hat{\mathbf{T}}_1(\mathbf{y},\boldsymbol{\theta}_0)$ is obtained. In all examples, the convergence of Gibbs sampler is checked using the Raftery–Lewis diagnostic test statistic (Raftery and Lewis, 1992).

4.1. Testing the significance in a simple linear regression model

Consider the following simple linear regression model:

$$y_i=\beta x_i+\varepsilon_i,\qquad \varepsilon_i\sim i.i.d.\ N\left(0,\sigma^2\right),\quad i=1,\ldots,n. \tag{18}$$

Denote $\mathbf{y}=(y_1,y_2,\ldots,y_n)'$ and $X=(x_1,x_2,\ldots,x_n)'$. We are interested in knowing whether or not the explanatory variable x_i has an explanatory power for y_i, i.e., we test

$$H_0:\beta=0,\qquad H_1:\beta\neq0.$$

The prior distributions for β and σ^2 are set at

$$\beta\sim N\left(\mu_\beta,\sigma^2V_\beta\right),\qquad \sigma^2\sim IG\left(a,b\right).$$

In this example, $\boldsymbol{\theta}=\beta$, $\boldsymbol{\psi}=\sigma^2$, and the likelihood function has a closed-form expression. Thus, $\mathbf{T}\left(\mathbf{y},\theta_0\right)$ can be computed. Also note that β is orthogonal to σ^2 and, hence, $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)\overset{a}{\sim}\chi^2(1)-1$. The marginal likelihood of data can be expressed, under H_0, as:

$$p_0\left(\mathbf{y}\right)=\frac{b^a\Gamma\left(a+\frac{n}{2}\right)}{(2\pi)^{n/2}\Gamma\left(a\right)}\left[b+\frac{1}{2}\mathbf{y}'\mathbf{y}\right]^{-(a+n/2)},$$

Author's personal copy

Table 3
Testing the significance in a simple linear regression model.

	$V_\beta = 0.1$	$V_\beta = 100$	$V_\beta = 10^5$	$V_\beta = 10^{22}$	$V_\beta = 10^{25}$	$V_\beta = 10^{35}$
BF_{01}	2.95×10^{-10}	2.63×10^{-9}	8.32×10^{-8}	26.3051	831.8407	8.31×10^{7}
$\mathbf{T}(\mathbf{y}, \theta_0)$	40.1209	40.1205	40.1205	40.1205	40.1205	40.1205
β	0.2447	0.2561	0.2562	0.2562	0.2562	0.2562
$SE(\beta)$	0.1322	0.1361	0.1361	0.1361	0.1361	0.1361
σ	0.5066	0.5036	0.5036	0.5036	0.5036	0.5036
$SE(\sigma)$	0.0250	0.0249	0.0249	0.0249	0.0249	0.0249

and under H_1, as:

$$p_1(\mathbf{y}) = \frac{b^a \Gamma\left(a + \frac{n}{2}\right)\sqrt{|V^*|}}{(2\pi)^{n/2}\,\Gamma(a)\sqrt{|V_\beta|}}\left[b + \frac{1}{2}\left(\mu_\beta' V_\beta^{-1}\mu_\beta + \mathbf{y}'\mathbf{y} - \mu^{*\prime} V^{*-1}\mu^*\right)\right]^{-(a+n/2)},$$

where

$$\mu^* = V^*\left(V_\beta^{-1}\mu_\beta + X'\mathbf{y}\right), \qquad V^* = \left(V_\beta^{-1} + X'X\right)^{-1}.$$

Hence, $BF_{01} = p_0(\mathbf{y})/p_1(\mathbf{y})$ has an analytical expression.

To explore the presence of Jeffreys–Lindley's paradox in BF and the absence of Jeffreys–Lindley's paradox in our proposed test, we consider an example used in Wooldridge (2009) (Page 45). In this example, a linear relationship between CEO salary and firm sales is established. To focus on the parameter of interest β, we subtract their respective sample mean from $\mathbf{y}$ and X and only estimate (18) without the intercept. To compute $\mathbf{T}(\mathbf{y}, \theta_0)$, we apply Gibbs sampler to the model corresponding to the alternative hypothesis to carry out the Bayesian analysis. We set the parameters in the priors at:

$$\mu_\beta = 0, \qquad a = 0.001, \qquad b = 0.001,$$

but leave the value of the prior variance V_β varied for the purpose of examining how V_β influences the decision based on BF_{01} and $\mathbf{T}(\mathbf{y}, \theta_0)$, respectively. For the Bayesian MCMC analysis, 10,000 random draws are sampled from the posterior distribution after 1000 burn-in periods.

The testing results and parameter estimates (both the posterior means and the posterior standard errors) are reported in Table 3. From this table, we see that as V_β increases, BF_{01} also increases. When the prior variance V_β is moderate, BF is less than 1 and tends to reject the null hypothesis. However, when V_β is large enough, the BF tends to support the null hypothesis. This clearly demonstrates Jeffreys–Lindley's paradox. On the other hand, the posterior distributions of β and σ remain nearly unchanged, and most importantly, $\mathbf{T}(\mathbf{y}, \theta_0)$ take nearly identical values with different V_β. Consequently, $\mathbf{T}(\mathbf{y}, \theta_0)$ is immune to Jeffreys–Lindley's paradox. To test the hypothesis using the proposed theory, since θ and σ^2 are orthogonal to each other, the asymptotic distribution of $\mathbf{T}(\mathbf{y}, \theta_0)$ is $\chi^2(1) - 1$. The 99%, 95%, 90% percentiles of $\chi^2(1) - 1$ are 5.63, 2.84, 1.71. The test statistic $\mathbf{T}(\mathbf{y}, \theta_0)$ is 40.12, suggesting that the null hypothesis is rejected under the 99%, 95%, 90% probability levels. When the frequentist's approach is used, the OLS estimate of β is 0.26 and the standard error is 0.03. This suggests that the null hypothesis has to be rejected, consistent with the finding from our method.

4.2. Hypothesis tests in asset pricing models with heavy tails

Asset pricing theory is a central focus of modern finance. Many econometric approaches have been developed to test asset pricing models. Most of the tests were developed based on the normality assumption, which is often violated in return data due to the presence of heavy tails. The heavy tails have motivated some researchers to develop asset pricing models with heavy-tailed distributions, see Zhou (1993), Kan and Zhou (2006), and Li and Yu (2012). In this subsection, we apply the proposed method to check the validity of Fama–French three factor asset pricing model (Fama and French, 1993) with a multivariate t distribution.

This asset pricing model with multivariate t distribution can be simply expressed as:

$$\mathbf{R}_t = \boldsymbol{\alpha} + \beta_1 M_t + \beta_2 SMB_t + \beta_3 HML_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim t(\mathbf{0}, \boldsymbol{\Sigma}, \nu),$$

where $\mathbf{R}_t$ is the excess return of portfolio at period t with $N \times 1$ dimension, M_t the excess return of the whole stock market, SMB_t and HML_t stands for "small (market capitalization) minus big" and for "high (book-to-market ratio) minus low" which measures the historical excess returns of small caps over big caps and of value stocks over growth stocks, $\boldsymbol{\Sigma}$ a diagonal matrix, and ν the freedom of degree of t distribution which is assumed to be known for the illustrative purpose and for convenience.

Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$, $\mathbf{F}_t = (M_t, SMB_t, HML_t)'$. As noted in Kan and Zhou (2006), using the scale mixture representation for t distribution, this model can be equivalently specified as:

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

The mean–variance efficiency suggests that the excess premium $\boldsymbol{\alpha}$ should be zero. Hence, the hypothesis to be tested is given by:

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

where $\mathbf{1}_N$ is an N-dimensional vector with unit elements.

As in the previous example, the likelihood function has a closed-form expression and, hence, both D and $\mathbf{T}(\mathbf{y}, \theta_0)$ can be computed. The purpose of this example is to check the quality of approximation of $\mathbf{T}_1(\mathbf{y}, \theta_0)$ and $\widehat{\mathbf{T}}_1(\mathbf{y}, \theta_0)$ when ω_t is regarded as latent variables.

In this empirical analysis, we consider the monthly returns of 25 portfolios constructed at the end of each June on the basis of 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 ranged from July 1926 to July 2011 so that $N = 25$, $T = 1021$. The data are freely available from the data library of Kenneth French.[1]

As noted in Kan and Zhou (2006), it is not easy to make the statistical inference using optimization-based ML methods. Hence, we consider Bayesian statistical inference coupled with MCMC techniques. Following Li and Yu (2012), we assign the vague conjugate prior distributions to represent the prior ignorance as follows:

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

and set $\nu = 3$.

In this Bayesian analysis, 110,000 random samples are draw from the posterior distribution using Gibbs sampler. The first 10,000 random samples are discarded as burning-in samples and

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

Author's personal copy

Table 4
Bayesian estimation and the standard error of the parameters for the Fama–French three factor model with the multivariate t distribution.

Portfolio	$\boldsymbol{\alpha}$		Portfolio	$\boldsymbol{\alpha}$	
	EST	SE		EST	SE
S1B1	−0.0083	0.0010	S1B2	−0.0031	0.0007
S1B3	−0.0017	0.0005	S1B4	−0.0001	0.0004
S1B5	0.0003	0.0005	S2B1	−0.0024	0.0005
S2B2	0.0000	0.0004	S2B3	0.0012	0.0003
S2B4	0.0006	0.0004	S2B5	−0.0003	0.0005
S3B1	−0.0003	0.0005	S3B2	0.0012	0.0004
S3B3	0.0012	0.0004	S3B4	0.0009	0.0004
S3B5	−0.0005	0.0005	S4B1	0.0005	0.0004
S4B2	−0.0002	0.0004	S4B3	0.0007	0.0004
S4B5	−0.0001	0.0004	S4B5	−0.0011	0.0006
S5B1	0.0006	0.0003	S5B2	0.0003	0.0004
S5B3	0.0004	0.0005	S5B4	−0.0012	0.0004
S5B5	−0.0026	0.0008			

the remaining samples are retained as effective observations. To check the quality of approximation of $\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta}_0)$ and $\hat{\mathbf{T}}_1(\mathbf{y}, \boldsymbol{\theta}_0)$ to $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$, we choose $S = 20$ and set the equal distance between $b_{(s)}$ and $b_{(s+1)}$ for $s = 0, 1, \ldots, 21$.

To save place, in Table 4, we only reported the Bayesian estimate and the standard error of the parameter of interest, namely $\boldsymbol{\alpha}$. The results for hypothesis testing are reported in Table 5. From these two tables, we find that $\widehat{D}$ well approximates D. Not surprisingly, $\hat{\mathbf{T}}_1(\mathbf{y}, \boldsymbol{\theta}_0)$, which is based on $\widehat{D}$, well approximates $\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta}_0)$ which in turn well approximates $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$. All three values are around 141. To obtain the threshold values, we estimate $\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)$ in (16), simulate 1000 random vectors from the standard multivariate normal variate, and then obtain 1000 random numbers for $\hat{\mathbf{T}}_1(\mathbf{y}, \boldsymbol{\theta}_0)$. From these random numbers, we obtain the following threshold values: $C = 20.2657$ under 99%, $C = 15.5040$ under 95% and $C = 11.3610$ under 90%. Consequently, we reject the null hypothesis under all the probability levels. Hence, it can be concluded that, despite its empirical popularity, the Fama–French three factor asset pricing model does not hold in this market.

4.3. Testing the leverage effect in a stochastic volatility model

Stochastic volatility (SV) models have been widely used for pricing options. An important and well documented empirical feature in many financial time series is the financial leverage effect (Black, 1976). Following Yu (2005), we define the leverage effects SV model as follows:

$$y_t|h_t = \exp(h_t/2)\, u_t, \quad t = 1, \ldots, n,$$
$$h_{t+1}|h_t, \mu, \phi, \tau^2, \rho = \mu + \phi(h_t - \mu) + \tau v_{t+1}, \quad t = 0, \ldots, n,$$

with

$$\begin{pmatrix} u_t \\ v_{t+1} \end{pmatrix} \overset{i.i.d.}{\sim} N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\},$$

and $h_0 = \mu$, where y_t is the return at time t, h_t the return volatility at period t. In this model, ρ the leverage effect parameter. When $\rho < 0$, there is a negative relationship between the expected future volatility and the current return (Yu, 2005). In particular, volatility tends to rise in response to bad news but fall in response to good news (Black, 1976). The hypothesis that we test is $H_0 : \rho = 0$.

To carry out Bayesian test of the hypothesis, we use the data that consist of daily returns on Pound/Dollar exchange rates $\{x_t\}$ from 01/10/81 to 28/06/85. The series $\{y_t\}$ is the daily mean-corrected returns. We first estimate the model using the Bayesian MCMC method. The following vague priors are specified:

$$\mu \sim N[0, 100], \qquad \phi \sim Beta[1, 1],$$
$$\tau^{-2} \sim \Gamma[0.001, 0.001], \qquad \rho \sim U[-1, 1].$$

Table 5
Asset pricing testing for the Fama–French three factor models.

Statistics	D	$\widehat{D}$	$\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta}_0)$	$\hat{\mathbf{T}}_1(\mathbf{y}, \boldsymbol{\theta}_0)$	$\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$
Value	82.6173	82.3551	141.1888	140.6644	140.5191

Table 6
Estimation results for the stochastic volatility model with the leverage effect.

Parameter	Mean	SE
μ	−0.6658	0.3507
ϕ	0.9788	0.0153
ρ	−0.0343	0.1481
τ	0.1685	0.0447

We draw 110,000 from the posterior distribution, discard the first 10,000 as build-in period and store every 20th value of the remaining samples as effective observations. The estimation results are reported in Table 6. To calculate $\widehat{\mathbf{T}}_1(\mathbf{y}, \theta_0)$, we take $s = 20$, set the equal distances between $b_{(s)}$ and $b_{(s+1)}$ for $s = 0, 1, \ldots, 20$ and find $\widehat{\mathbf{T}}_1(\mathbf{y}, \theta_0) = -1.7244$. From simulations, the threshold values are $C = 5.1041$ under 99%, $C = 2.2291$ under 95% and $C = 1.0600$ under 90%. Hence, the null hypothesis cannot be rejected under all three probability levels. While a strong leverage effect has been found in the equity markets (see, for example, Yu (2005, 2012) and Ait-Sahalia et al. (2013)), there is no significant leverage effect in the exchange rate.

5. Conclusion

In this paper, we have proposed a new Bayesian statistic to test a point null hypothesis with the hope that the new statistic is less sensitive to the choice of priors than the well known BF. The test statistic is based on the difference of the two deviances averaged over the posterior distribution. It can be motivated from a decision theoretical framework. The main advantages of the new statistic are fourfold. First, it is immune to Jeffreys' concern. Second, it avoids Jeffreys–Lindley's paradox. Third, it can be easily computed using the MCMC outputs from the posterior distribution. Fourth, the asymptotic distribution can be derived for calibrating the threshold values. The proposed method is illustrated using a simple linear regression model, an asset pricing model and a stochastic volatility model with real data. We have found very strong evidence against the popular Fama–French three factor asset pricing model with equity data. For an exchange rate series, on the other hand, we cannot find a strong support for the presence of leverage effect.

It is known that BFs are more conservative than the LR test (see for example, Edwards et al. (1963), Kass and Raftery (1995) and Efron et al. (2001)). This is because BFs has a built-in penalty term that depends on the sample size. On the other hand, the likelihood ratio test does not have such a penalty term and tends to reject the null in very large samples even when the null is meaningful (Raftery, 1986). Similarly, the proposed approach here does not have a penalty term in association with the dimension of a model. In this sense, we caution in general against basing hypothesis testing solely on the proposed statistic when the user is conservative and has a highly informative prior.

Appendix A. Proof of Theorem 3.1

It can be shown that

$$E_{\mathbf{y}}\left\{ \int \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \right\}$$
$$= \int \int \int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} p(\mathbf{y}) \mathrm{d}\mathbf{y}$$

Author's personal copy

$$
\begin{aligned}
&= \int\int\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} p(\mathbf{y}, \boldsymbol{\vartheta}) \mathrm{d}\mathbf{y}\mathrm{d}\boldsymbol{\vartheta} \\
&= \int \left\{ \int \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right] p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right\} p(\boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \int \left\{ \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right] \left[\int p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right] \right\} p(\boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \int \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right] p(\boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \int \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \right] p(\mathbf{y}) \mathrm{d}\mathbf{y}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&E_{\mathbf{y}} \left\{ \int \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}_0) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \right\} \\
&= \int \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}_0) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \right] p(\mathbf{y}) \mathrm{d}\mathbf{y}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E_{\mathbf{y}} [T_{BR}(\mathbf{y}, \boldsymbol{\theta}_0)] &= E_{\mathbf{y}} \left\{ \int KL [p(\mathbf{y}|\boldsymbol{\theta}), p(\mathbf{y}|\boldsymbol{\theta}_0)] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \right\} \\
&= E_{\mathbf{y}} \left\{ \int \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \right\} \\
&\quad - E_{\mathbf{y}} \left\{ \int \left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}_0) p(\mathbf{y}|\boldsymbol{\vartheta}) \mathrm{d}\mathbf{y} \right] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \right\} \\
&= \int\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} p(\mathbf{y}) \mathrm{d}\mathbf{y} \\
&\quad - \int\int \log p(\mathbf{y}|\boldsymbol{\vartheta}_0) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} p(\mathbf{y}) \mathrm{d}\mathbf{y} \\
&= E_{\mathbf{y}} \int [\log p(\mathbf{y}|\boldsymbol{\vartheta}) - \log p(\mathbf{y}|\boldsymbol{\vartheta}_0)] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}.
\end{aligned}
$$

Theorem 3.1 is proven.

Appendix B. Proof of Theorem 3.2

Applying the Taylor expansion on the logarithm of the posterior density, we get

$$
\begin{aligned}
\log p(\boldsymbol{\vartheta}|\mathbf{y}) &= \log p(\hat{\boldsymbol{\vartheta}}|\mathbf{y}) + L_n^{(1)}(\hat{\boldsymbol{\vartheta}})'(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) \\
&\quad + \frac{1}{2}(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' L_n^{(2)}(\tilde{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) \\
&= \log p(\hat{\boldsymbol{\vartheta}}|\mathbf{y}) + \frac{1}{2}(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' L_n^{(2)}(\tilde{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}),
\end{aligned}
$$

where $\tilde{\boldsymbol{\vartheta}}$ lies on the segment between $\boldsymbol{\vartheta}$ and $\hat{\boldsymbol{\vartheta}}$. Note that

$$
p(\boldsymbol{\vartheta}|\mathbf{y}) = \frac{p(\mathbf{y}, \boldsymbol{\vartheta})}{p(\mathbf{y})}.
$$

Hence,

$$
\begin{aligned}
&\log p(\boldsymbol{\vartheta}|\mathbf{y}) - \log p(\hat{\boldsymbol{\vartheta}}|\mathbf{y}) = \log p(\mathbf{y}, \boldsymbol{\vartheta}) \\
&\quad - \log p(\mathbf{y}) - \log p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) + \log p(\mathbf{y}) \\
&= \log p(\mathbf{y}, \boldsymbol{\vartheta}) - \log p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) = \frac{1}{2}(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' L_n^{(2)}(\tilde{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}).
\end{aligned}
$$

Then, for any $\epsilon > 0$, there exists an integer N_2 such that for any $n > N_2$, $L_n^{(2)}(\tilde{\boldsymbol{\vartheta}})$ satisfies

$$
\begin{aligned}
&\left[\boldsymbol{I}_{\boldsymbol{p}+\boldsymbol{q}} - A(\epsilon) \right] \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] \le -L_n^{(2)}(\tilde{\boldsymbol{\vartheta}}) = \left[L_n^{(2)}(\tilde{\boldsymbol{\vartheta}}) L_n^{-(2)}(\hat{\boldsymbol{\vartheta}}) \right] \\
&\quad \times [-L_n^{(2)}(\hat{\boldsymbol{\vartheta}})] \le \left[\boldsymbol{I}_{\boldsymbol{p}+\boldsymbol{q}} + A(\epsilon) \right] \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right].
\end{aligned}
$$

Following the proof of Theorem 3.2 in Li et al. (2012), under Assumptions 1–7, we note that there exists N, when $n > N$, we have

$$
\begin{aligned}
&\int (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' \left[-L_n^{(2)}(\tilde{\boldsymbol{\vartheta}}) \right] (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \int_{H(\hat{\boldsymbol{\vartheta}}, \delta)} [(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' \left[-L_n^{(2)}(\tilde{\boldsymbol{\vartheta}}) \right] (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \int_{H(\hat{\boldsymbol{\vartheta}}, \delta)} (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' \left[L_n^{(2)}(\tilde{\boldsymbol{\vartheta}}) L_n^{-(2)}(\hat{\boldsymbol{\vartheta}}) \right] \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] \\
&\quad \times (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta},
\end{aligned}
$$

which is bounded above by

$$
\begin{aligned}
&\int_{H(\hat{\boldsymbol{\vartheta}}, \delta)} [(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' \left[\boldsymbol{I}_{\boldsymbol{p}+\boldsymbol{q}} + A(\epsilon) \right] \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \mathbf{tr} \left\{ \left[\boldsymbol{I}_{\boldsymbol{p}+\boldsymbol{q}} + A(\epsilon) \right] \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] V(\hat{\boldsymbol{\vartheta}}) \right\},
\end{aligned}
$$

and below by

$$
\begin{aligned}
&\int_{H(\hat{\boldsymbol{\vartheta}}, \delta)} [(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' \left[\boldsymbol{I}_{\boldsymbol{p}+\boldsymbol{q}} - A(\epsilon) \right] \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \mathbf{tr} \left\{ \left[\boldsymbol{I}_{\boldsymbol{p}+\boldsymbol{q}} - A(\epsilon) \right] \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] V(\hat{\boldsymbol{\vartheta}}) \right\}.
\end{aligned}
$$

Hence, under the regularity conditions, for $\epsilon \to 0$, we have

$$
\begin{aligned}
&\lim_{n \to \infty} \int (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' \left[-L_n^{(2)}(\tilde{\boldsymbol{\vartheta}}) \right] (\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \mathbf{tr} \left\{ \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] V(\hat{\boldsymbol{\vartheta}}) \right\}.
\end{aligned}
$$

Furthermore, it can be shown that

$$
\begin{aligned}
\mathbf{tr} \left\{ \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] V(\hat{\boldsymbol{\theta}}) \right\} &= \mathbf{tr} \left\{ \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right]^{-1} \right\} + o(1) \\
&= p + q + o(1).
\end{aligned}
$$

Hence, conditional on the observed data $\mathbf{y}$, we get

$$
\begin{aligned}
&\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} = \int [\log p(\mathbf{y}, \boldsymbol{\vartheta}) - \log p(\boldsymbol{\vartheta})] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \int \log p(\mathbf{y}, \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} - \int \log p(\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \int \left[\frac{1}{2}(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' L_n^{(2)}(\tilde{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) \right] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} + \log p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) \\
&\quad - \int \log p(\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= -\frac{1}{2} \mathbf{tr} \left\{ \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] V(\hat{\boldsymbol{\theta}}) \right\} + o(1) + \log p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) \\
&\quad - \int \log p(\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \log p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) - \int \log p(\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&\quad - \frac{1}{2} \mathbf{tr} \left\{ \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) \right] V(\hat{\boldsymbol{\vartheta}}) \right\} + o(1) \\
&= \log p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) - \int \log p(\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} - \frac{1}{2}(p + q) + o(1).
\end{aligned}
$$

Furthermore, it is noted that

$$
\log p(\mathbf{y}, \bar{\boldsymbol{\vartheta}}) = \log p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) + \frac{1}{2}(\bar{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}})' L_n^{(2)}(\tilde{\boldsymbol{\theta}})(\bar{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}),
$$

where $\tilde{\boldsymbol{\theta}}$ lies on the segment between $\bar{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$. Using Assumption 7, we can show that $\log p(\mathbf{y}, \bar{\boldsymbol{\vartheta}}) = \log p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) + o_p(1)$.

Author's personal copy

Similarly, under the null hypothesis, it can be shown that

$$\log p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0) = \log p(\mathbf{y}, \bar{\boldsymbol{\psi}}|\boldsymbol{\theta}_0) + \left.\frac{\log p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}} (\boldsymbol{\psi} - \bar{\boldsymbol{\psi}}) + \frac{1}{2}(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})' \left[\left.\frac{\partial^2 \log p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}=\tilde{\boldsymbol{\psi}}^*}\right](\boldsymbol{\psi} - \bar{\boldsymbol{\psi}}),$$

where $\tilde{\boldsymbol{\psi}}^*$ lies on the segment between $\boldsymbol{\psi}$ and $\bar{\boldsymbol{\psi}}$. When $n \to \infty$, we have $H(\bar{\boldsymbol{\psi}}, \delta_1) \subset H(\hat{\boldsymbol{\psi}}, \delta)$ and $\tilde{\boldsymbol{\psi}}^* \in H(\bar{\boldsymbol{\psi}}, \delta_1) \subset H(\hat{\boldsymbol{\psi}}, \delta)$. Then,

$$\begin{aligned}
&\int (\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})' \left[-L_{0n}^{(2)}(\tilde{\boldsymbol{\psi}}^*)\right](\boldsymbol{\psi} - \bar{\boldsymbol{\psi}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \mathbf{tr}\left\{\left[-L_{0n}^{(2)}(\tilde{\boldsymbol{\psi}}^*)\right]\left[\int (\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})' p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}\right]\right\} \\
&= \mathbf{tr}\left\{\left[-L_{0n}^{(2)}(\hat{\boldsymbol{\psi}})\right] E[(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})'|\mathbf{y}, H_1]\right\} + o_p(1).
\end{aligned}$$

Moreover, we get

$$\begin{aligned}
&\int \left[\left.\frac{\log p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}}\right](\boldsymbol{\psi} - \bar{\boldsymbol{\psi}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&= \left[\left.\frac{\log p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}}\right](\bar{\boldsymbol{\psi}} - \bar{\boldsymbol{\psi}}) = 0
\end{aligned}$$

and

$$\begin{aligned}
&\int \log p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} = \log p(\mathbf{y}, \bar{\boldsymbol{\psi}}|\boldsymbol{\theta}_0) - \frac{1}{2}\mathbf{tr}\left\{\left[-L_{0n}^{(2)}(\hat{\boldsymbol{\psi}})\right]\right. \\
&\left.\times\ E[(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})'|\mathbf{y}, H_1]\right\} + o_p(1).
\end{aligned}$$

Hence,

$$\begin{aligned}
E[(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})'|\mathbf{y}, H_1] &= E[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})'|\mathbf{y}, H_1] \\
&\quad + 2E[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})|\mathbf{y}, H_1](\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})' + (\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})(\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})' \\
&= E[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})'|\mathbf{y}, H_1] + o_p(n^{-1/2}) o_p(n^{-1/2}) \\
&= E[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})'|\mathbf{y}, H_1] + o_p(n^{-1}) = -L_n^{-(2)}(\hat{\boldsymbol{\vartheta}}) + o_p(n^{-1}) \\
&= \frac{1}{n}\left[\frac{1}{n}L_n^{(2)}(\hat{\boldsymbol{\vartheta}})\right]^{-1} + o_p(n^{-1}) = \frac{1}{n}O_p(1) + o_p(n^{-1}) = O_p(n^{-1}),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbf{tr}\left\{\left[-L_{0n}^{(2)}(\hat{\boldsymbol{\psi}})\right] E[(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})'|\mathbf{y}, H_1]\right\} \\
&= \mathbf{tr}\left\{\left[-\frac{1}{n}L_{0n}^{(2)}(\hat{\boldsymbol{\psi}})\right] nE[(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})'|\mathbf{y}, H_1]\right\} \\
&= \mathbf{tr}\left\{\left[-\frac{1}{n}L_{0n}^{(2)}(\bar{\boldsymbol{\psi}}) + o_p(1)\right] nE[(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})'|\mathbf{y}, H_1]\right\} \\
&= \mathbf{tr}\left\{\left[-L_{0n}^{(2)}(\bar{\boldsymbol{\psi}})\right] E[(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \bar{\boldsymbol{\psi}})'|\mathbf{y}, H_1]\right\} + o_p(1).
\end{aligned}$$

We can further show that

$$\begin{aligned}
T(\mathbf{y}, \boldsymbol{\theta}_0) &= 2\left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} - \int \log p(\mathbf{y}|\boldsymbol{\vartheta}_0) p(\boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\psi}\right] \\
&= 2\log p(\mathbf{y}, \bar{\boldsymbol{\vartheta}}) - 2\int \log p(\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} - (p+q) \\
&\quad - 2\log p(\mathbf{y}, \bar{\boldsymbol{\psi}}|\boldsymbol{\theta}_0) + 2\int \log p(\boldsymbol{\psi}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
&\quad + \mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\psi}}) V_{22}(\bar{\boldsymbol{\vartheta}})] + o_p(1) \\
&= 2[\log p(\mathbf{y}, \bar{\boldsymbol{\vartheta}}) - \log p(\mathbf{y}, \bar{\boldsymbol{\psi}}|\boldsymbol{\theta}_0)] \\
&\quad - 2\left[\int \log p(\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} - \int \log p(\boldsymbol{\psi}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}\right] \\
&\quad - \left[p + q - \mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\vartheta}}_0) V_{22}(\bar{\boldsymbol{\vartheta}})]\right] + o_p(1) \\
&= 2[\log p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}})] + 2\left[\log p(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\psi}})\right. \\
&\quad \left. - \log p(\bar{\boldsymbol{\psi}}|\boldsymbol{\theta}_0)\right] - 2\left[\int \log p(\boldsymbol{\theta}|\boldsymbol{\psi}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}\right] \\
&\quad - \left[p + q - \mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\psi}}) V_{22}(\bar{\boldsymbol{\vartheta}})]\right] + o_p(1).
\end{aligned}$$

For latent variable models, $p(\mathbf{y}|\boldsymbol{\vartheta})$ generally does not have an analytical form. Using the path sampling technique of Gelman and Meng (1998), we get:

$$p(\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b) = \frac{p(\mathbf{z}, \mathbf{y}|\bar{\boldsymbol{\vartheta}}_b)}{p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}_b)} = \frac{p(\mathbf{z}, \mathbf{y}|\bar{\boldsymbol{\vartheta}}_b)}{f(b)},$$

where $f(b) = p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}_b)$ such that $f(1) = p(\mathbf{y}|\bar{\boldsymbol{\vartheta}})$ and $f(0) = p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}_*)$. Then,

$$\begin{aligned}
\frac{\partial \log f(b)}{\partial b} &= \frac{f'(b)}{f(b)} = \frac{1}{f(b)}\int \frac{\partial p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_b)}{\partial b}\mathrm{d}\mathbf{z} \\
&= \frac{1}{f(b)}\int \frac{\partial \log p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_b)}{\partial b} p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_b)\mathrm{d}\mathbf{z} \\
&= \int \frac{\partial \log p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_b)}{\partial b}\frac{p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_b)}{f(b)}\mathrm{d}\mathbf{z} \\
&= \int \frac{\partial \log p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_b)}{\partial b} p(\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b)\mathrm{d}\mathbf{z} \\
&= E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b}\left[\frac{\partial \log p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_b)}{\partial b}\right] \\
&= E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b}\left[\frac{\partial \bar{\boldsymbol{\vartheta}}_b}{\partial b}\frac{\partial \log p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\vartheta}}_b)}{\partial \bar{\boldsymbol{\vartheta}}_b}\right].
\end{aligned}$$

Hence, we get

$$\begin{aligned}
\log p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}) - \log p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}_*) &= \log\frac{f(1)}{f(0)} = \int_0^1 \frac{\partial \log f(b)}{\partial b}\mathrm{d}b \\
&= \int_0^1 \left\{(\bar{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}}_*)' E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b}\left[\left.\frac{\partial \log p(\mathbf{y}, \mathbf{z}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\bar{\boldsymbol{\vartheta}}_b}\right]\right\}\mathrm{d}b \\
&= \int_0^1 \left\{(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b}\left[\left.\frac{\partial \log p(\mathbf{y}, \mathbf{z}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}_b}\right]\right\}\mathrm{d}b \\
&\quad + \int_0^1 \left\{(\bar{\boldsymbol{\psi}} - \bar{\boldsymbol{\psi}})' E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b}\left[\left.\frac{\partial \log p(\mathbf{y}, \mathbf{z}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}_b}\right]\right\}\mathrm{d}b \\
&= \int_0^1 \left\{(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' E_{\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\vartheta}}_b}[S_1(\mathbf{x}|\bar{\boldsymbol{\vartheta}}_b)]\right\}\mathrm{d}b.
\end{aligned}$$

Theorem 3.2 is proven.

Appendix C. Proof of Theorem 3.3

When $n \to \infty$, the prior information is negligible. Hence, we have

$$\frac{\partial \log p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = L_n^{(1)}(\boldsymbol{\theta}), \qquad \frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\boldsymbol{\theta}'} = L_n^{(2)}(\boldsymbol{\theta}),$$

and the ML estimator is asymptotically equivalent to the posterior mode $\hat{\boldsymbol{\theta}}$. Furthermore, according to Theorem 3.2, it can be shown that

$$T(\mathbf{y}, \boldsymbol{\theta}_0) = 2\left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y})\mathrm{d}\boldsymbol{\vartheta}\right.$$

Author's personal copy

$$
\begin{aligned}
&- \int \log p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}_0) p(\boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\psi} \Big] \\
&= 2\left[\log p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}})\right] \\
&\quad - \left[p + q - \mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\vartheta}}) V_{22}(\bar{\boldsymbol{\vartheta}})]\right] + o_p(1).
\end{aligned}
$$

In Theorem 3.2, it is shown that $\log p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}) = \log p(\mathbf{y}|\hat{\boldsymbol{\vartheta}}) + o_p(1)$. Similarly, when H_0 is true, let $\bar{\boldsymbol{\vartheta}}_* = (\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}})$, we can show that

$$
\begin{aligned}
\log p(\mathbf{y}|\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}}) &= \log p(\mathbf{y}|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) + L_n^{(1)}(\hat{\boldsymbol{\vartheta}})(\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}}) \\
&\quad + \frac{1}{2}(\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}})' L_n^{(2)}(\hat{\boldsymbol{\vartheta}})(\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}}) + o_p(1) \\
&= \log p(\mathbf{y}|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) + \frac{1}{2}(\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}})' \\
&\quad \times L_n^{(2)}(\hat{\boldsymbol{\vartheta}})(\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}}) + o_p(1).
\end{aligned}
$$

Furthermore, under the null hypothesis, it is noted that $\bar{\boldsymbol{\psi}} = \hat{\boldsymbol{\psi}} + o_p(n^{-\frac{1}{2}})$, $\frac{1}{n} L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) = O_p(1)$ and

$$
\begin{aligned}
&-\frac{1}{n} L_n^{(2)}(\hat{\boldsymbol{\vartheta}}) = -\frac{1}{n} L_n^{(2)}(\boldsymbol{\vartheta}_0) + o_p(1) = \mathbf{J}(\boldsymbol{\vartheta}_0) + o_p(1), \\
&\left[-\frac{1}{n} L_n^{(2)}(\hat{\boldsymbol{\vartheta}}_0)\right]^{-1} = -\left[\frac{1}{n} L_n^{(2)}(\boldsymbol{\vartheta}_0)\right]^{-1} + o_p(1) = \mathbf{J}^{-1}(\boldsymbol{\vartheta}_0) + o_p(1) \\
&\qquad = \mathbf{IJ}(\boldsymbol{\vartheta}_0) + o_p(1).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&(\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}})' L_n^{(2)}(\hat{\boldsymbol{\vartheta}})(\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}}) \\
&= (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' L_{n,11}^{(2)}(\hat{\boldsymbol{\vartheta}})(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + 2(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' \\
&\quad \times L_{n,12}^{(2)}(\hat{\boldsymbol{\vartheta}})(\bar{\boldsymbol{\psi}} - \hat{\boldsymbol{\psi}}) + (\bar{\boldsymbol{\psi}} - \hat{\boldsymbol{\psi}})' L_{n,22}^{(2)}(\hat{\boldsymbol{\vartheta}})(\bar{\boldsymbol{\psi}} - \hat{\boldsymbol{\psi}}) \\
&= (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' L_{n,11}^{(2)}(\hat{\boldsymbol{\vartheta}})(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) \\
&\quad + 2O_p(n^{-1/2}) O_p(n) o_p(n^{-1/2}) + o_p(n^{-1/2}) O_p(n) o_p(n^{-1/2}) \\
&= (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' L_{n,11}^{(2)}(\hat{\boldsymbol{\vartheta}})(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + o_p(1) \\
&= \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' \left[\frac{1}{n} L_{n,11}^{(2)}(\hat{\boldsymbol{\vartheta}})\right] \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + o_p(1) \\
&= -\sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' \left[\mathbf{J}_{11}(\boldsymbol{\vartheta}_0) + o_p(1)\right] \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + o_p(1) \\
&= -\sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' \left[\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)\right] \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) \\
&\quad + o_p(1) O_p(1) O_p(1) + o_p(1) \\
&= -\sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' \left[\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)\right] \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + o_p(1).
\end{aligned}
$$

According to the ML theory, we know that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim N[\mathbf{0}, \mathbf{IJ}_{11}(\boldsymbol{\vartheta}_0)]$ so that $\boldsymbol{\epsilon} = \sqrt{n}\mathbf{IJ}_{11}^{-1/2}(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim N[\mathbf{0}, \mathbf{I}_p]$. Hence, we have

$$
\begin{aligned}
&2\left[\log p(\mathbf{y}|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}})\right] \\
&= (\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}})' \left[-L_n^{(2)}(\hat{\boldsymbol{\vartheta}})\right] (\bar{\boldsymbol{\vartheta}}_* - \hat{\boldsymbol{\vartheta}}) + o_p(1) \\
&= \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' \left[\mathbf{J}_{11}(\boldsymbol{\vartheta}_0)\right] \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + o_p(1) \\
&= \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' \mathbf{IJ}_{11}^{-1/2}(\boldsymbol{\vartheta}_0) \left[\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0) \mathbf{J}_{11}(\boldsymbol{\vartheta}_0) \mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\right] \\
&\quad \times \mathbf{IJ}_{11}^{-1/2} \sqrt{n}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + o_p(1) \\
&= \boldsymbol{\epsilon}' \left[\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0) \mathbf{J}_{11}(\boldsymbol{\vartheta}_0) \mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\right] \boldsymbol{\epsilon} + o_p(1).
\end{aligned}
$$

Further, when the null hypothesis is true, we can get that

$$
\begin{aligned}
&T(\mathbf{y}, \boldsymbol{\theta}_0) + \left[p + q - \mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\vartheta}}) V_{22}(\bar{\boldsymbol{\vartheta}})]\right] \\
&= 2\left[\int \log p(\mathbf{y}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} - \int \log p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}_0) p(\boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\psi}\right] \\
&\quad + \left[p + q - \mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\vartheta}}) V_{22}(\bar{\boldsymbol{\vartheta}})]\right] \\
&= T_1(\mathbf{y}, \boldsymbol{\theta}_0) + \left[p + q - \mathbf{tr}[-L_{0n}^{(2)}(\bar{\boldsymbol{\vartheta}}) V_{22}(\bar{\boldsymbol{\vartheta}})]\right] + o_p(1) \\
&= 2[\log p(\mathbf{y}|\bar{\boldsymbol{\vartheta}}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}})] + o_p(1) \\
&= 2\left[\log p(\mathbf{y}|\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}})\right] + o_p(1) \\
&\sim \boldsymbol{\epsilon}' \left[\mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0) \mathbf{J}_{11}(\boldsymbol{\vartheta}_0) \mathbf{IJ}_{11}^{1/2}(\boldsymbol{\vartheta}_0)\right] \boldsymbol{\epsilon}.
\end{aligned}
$$

References

Ait-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: disentangling sources of bias at high frequency. Journal of Financial Economics 109, 224–229.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, 2nd ed. In: Springer Series in Statistics, Springer, New York, p. 1.

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

Black, F., 1976. Studies of stock market volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics Section 177–181.

Chen, C., 1985. On asymptotic normality of limiting density function with Bayesian implications. Journal of the Royal Statistical Society. Series B. Statistical Methodology 47, 540–546.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B. Statistical Methodology 39, 1–38.

Edwards, W., Lindman, H., Savage, L.J., 1963. Bayesian statistical inference for psychological research. Psychological Review 70, 193.

Efron, B., Gous, A., Kass, R., Datta, G., Lahiri, P., 2001. Scales of Evidence for Model Selection: Fisher versus Jeffreys. In: Lecture Notes-Monograph Series, pp. 208–256.

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

Gelman, A., Meng, X.-L., 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Statistical Science 13, 163–185.

Geweke, J., 2007. Bayesian model comparison and validation. The American Economic Review 97, 60–64.

Geweke, J., 2005. Contemporary Bayesian Econometrics and Statistics, vol. 537. Wiley-Interscience.

Geweke, J., Koop, G., van Dijk, H., 2011. The Oxford Handbook of Bayesian Econometrics. Oxford University Press.

Good, I., 1956. The surprise index for the multivariate normal distribution. The Annals of Mathematical Statistics 27, 1130–1135.

Jeffreys, H., 1961. Theory of Probability. Clarendon Press, Oxford.

Kan, R., Zhou, G., 2006. Modelling Non-normality using Multivariate t: Implications to Asset Pricing, Working Paper.

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

Kim, J., 1994. Bayesian asymptotic theory in a time series model with a possible nonstationary process. Econometric Theory 10, 764–773.

Kim, J., 1998. Large sample properties of posterior densities, bayesian information criterion and the likelihood principle in nonstationary time series models. Econometrica 66, 359–380.

Li, Y., Yu, J., 2012. Bayesian hypothesis testing in latent variable models. Journal of Econometrics 166, 237–246.

Li, Y., Zeng, T., Yu, J., 2012. Robust Deviation Information Criterion for Latent Variable Models, Working Paper.

Louis, T.A., 1982. Finding the observed information matrix when using the em algorithm. Journal of the Royal Statistical Society. Series B. Statistical Methodology 44, 226–233.

McCulloch, R.E., 1989. Local model influence. Journal of the American Statistical Association 84, 473–478.

Oakes, D., 1999. Direct calculation of the information matrix via the EM. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61, 479–482.

O'Hagan, A., 1995. Fractional Bayes factors for model comparison. Journal of the Royal Statistical Society. Series B. Methodological 57, 99–138.

Poirier, D.J., 1995. Intermediate Statistics and Econometrics: A Comparative Approach. The MIT Press.

Raftery, A.E., 1986. Choosing models for cross-classifications. American Sociological Review 51, 145–146.

Raftery, A., Lewis, S., 1992. How Many Iterations in the Gibbs Sampler? In: Bayesian Statistics, vol. 4.

Robert, C.P., 1993. A note on Jeffreys–Lindley paradox. Statistica Sinica 3, 601–608.

Robert, C., 2001. The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation. Springer.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64, 583–639.

Wooldridge, J.M., 2009. Introductory Econometrics: A Modern Approach. South-Western Pub.

Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.

Yu, J., 2012. A semiparametric stochastic volatility model. Journal of Econometrics 167, 473–482.

Zhou, G., 1993. Asset-pricing tests under alternative distributions. Journal of Finance 48, 1927–1942.