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.

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
http://dx.doi.org/10.1016/j.jeconom.2013.08.035

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

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

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:

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

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.