6-2014

A Bayesian Chi-Squared Test for Hypothesis Testing

Yong LI
Renmin University of China

Xiao-Bin LIU
Singapore Management University, xb.liu.2013@phdecons.smu.edu.sg

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

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

A Bayesian Chi-Squared Test for Hypothesis Testing

Yong Li, Xiao-Bin Liu and Jun Yu

June 2014

Paper No. 03-2014

A Bayesian Chi-Squared Test for Hypothesis Testing*

Yong Li
Renmin University

Xiao-Bin Liu
Singapore Management University

Jun Yu
Singapore Management University

May 16, 2014

Abstract

A new Bayesian test statistic is proposed to test a point null hypothesis based on a quadratic loss. The proposed test statistic may be regarded as the Bayesian version of Lagrange multiplier test. Its asymptotic distribution is obtained based on a set of regular conditions and follows a chi-squared distribution when the null hypothesis is correct. The new statistic has several important advantages that make it appeal in practical applications. First, it is well-defined under improper prior distributions. Second, it avoids Jeffrey-Lindley's paradox. Third, it is relatively easy to compute, even for models with latent variables. Finally, it is pivotal and its threshold value can be easily obtained from the asymptotic chi-squared distribution. The method is illustrated using some real examples in economics and finance.

JEL classification: C11, C12
Keywords: Bayes factor; Decision theory; EM algorithm; Lagrange multiplier; Markov chain Monte Carlo; Latent variable models.

*Li gratefully acknowledges the financial support of the Chinese Natural Science fund (No.71271221) and the hospitality during his research visits to Singapore Management University. Jun 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. We would like to acknowledge helpful comments from the seminar and conference participants at Peking University, Princeton University, the inaugural China Meeting of the Econometric Society. Yong Li, Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing, 100872, P.R. China. Xiao-Bin, Liu, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yujun@smu.ed.sg. URL: http://www.mysmu.edu/faculty/yujun/.

1 Introduction

This paper is concerned with statistical testing of a point null hypothesis after a Bayesian Markov chain Monte Carlo (MCMC) method has been used to estimate the models. The importance of testing a point null hypothesis is well-known in economics. In the recent years, Bayesian MCMC methods have found more and more applications in economics because they make it possible to fit increasingly complex models, including latent variable models (Shephard, 2005), dynamic discrete choice models (Imai, Jain and Ching, 2009) and dynamic general equilibrium models (An and Schorfheide, 2007).

In the Bayesian paradigm, Bayes factor (BF) is the gold standard for the Bayesian model comparison and the Bayesian hypothesis testing (Kass and Raftery 1995; Geweke, 2007). Unfortunately, BF is not problem-free. First, BF is sensitive to the prior and subject to Jeffreys-Lindley's paradox; see for example, Kass and Raftery (1995), Poirier (1995), Robert (1993, 2001). Second, the calculation of BF for hypothesis testing generally requires the evaluation of marginal likelihood which is a marginalization over the unknown quantities. In many cases, the evaluation of marginal likelihood is difficult. Not surprisingly, alternative strategies have been proposed to test a point null hypothesis in the Bayesian literature. These methods can be classified into two classes.

In the first class, refinements are made to BF to overcome the theoretical and computational difficulties. For example, to reduce the influence of BF to the prior, one may split the data into two parts, a training sample and a sample for statistical analysis. The training sample is used to update the non-informative prior and to obtain a new proper informative prior. This idea includes the fractional BF (O'Hagan 1995) and the intrinsic BF (Berger and Perrichi, 1996). In practice, however, it is often not clear how to split the sample and the testing outcome may be sensitive to how the sample is split.

In the second class, instead of refining the BF methodology, several interesting Bayesian approaches have been proposed for hypothesis testing based on the decision theory. For example, Bernardo and Rueda (2002, BR hereafter) showed that BF for the Bayesian hypothesis testing can be regarded as a decision problem with a simple zero-one discrete loss function. However, the zero-one discrete function requires the use of non-regular (not absolutely continuous) prior and this is why BF leads to Jeffreys-Lindley's paradox. BR further suggested using a continuous loss function, based on the well-known continuous Kullback-Leibler (KL) divergence function. As a result, it was shown in BR that their Bayesian test statistic does not depend on any arbitrary constant in the prior. However, BR's approach has some disadvantages. First, the analytical expression of the KL loss function required by BR is not always available, especially for latent variable models. Second, the test statistic is not a pivotal quantity. Consequently, BR had to use subjective threshold values to test the hypothesis.

To deal with the computational problem in BR in latent variable models, Li and Yu (2012, LY hereafter) proposed a new test statistic based on the $\mathcal{Q}$ function in the Expectation-Maximization (EM) algorithm.[1] LY showed that the new statistic is well-defined under improper priors and easy to compute for latent variable models. Following the idea of McCulloch (1989), LY proposed to choose the threshold values based on the Bernoulli distribution. However, like the test statistic proposed by BR, the test statistic proposed by LY is not pivotal. Moreover, it is not clear if the test statistic of LY can resolve Jeffreys-Lindley's paradox.

Based on the difference between the deviances, Li, Zeng and Yu (2014, LZY hereafter) developed another Bayesian test statistic for hypothesis testing. This test statistic is well-defined under improper priors, free of Jeffreys-Lindley's paradox, and not difficult to compute. Moreover, its asymptotic distribution can be derived and one may obtain the threshold values from the asymptotic distribution. Unfortunately, in general the asymptotic distribution depends on some unknown population parameters and hence the test is not pivotal.

In the present paper, we propose a pivotal Bayesian test statistic, based on a quadratic loss function, to test a point null hypothesis within the decision-theoretic framework. The new statistic has the four desirable properties that makes it appeal in practice after the models are estimated by Bayesian MCMC methods. First, it is well-defined under improper prior distributions. Second, it is immune to Jeffreys-Lindley's paradox. Third, it is easy to compute. The main computational effort is to get the first derivative of the likelihood function with respect to the parameters. For latent variable models, the first derivative can be easily evaluated from the MCMC output with the help of the EM algorithm. Finally, the asymptotic distribution of the test statistic follows the chi-squared distribution and hence the test is pivotal. In particular, under a set of regularity conditions, we show that our test statistic is asymptotically equivalent to the Lagrange multiplier (LM) statistic that has been commonly used in the frequentist's paradigm to test a point null hypothesis.

The paper is organized as follows. Section 2 reviews the Bayesian literature on testing a point null hypothesis from the viewpoint of the decision theory. Section 3 develops the new Bayesian test statistic and establishes its asymptotic 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 all the theoretical results.

[1]The EM algorithm was originally proposed to compute maximum likelihood estimates in latent variable models (Dempster, et al., 1977).

2 Bayesian Hypothesis Testing under Decision Theory

2.1 Testing a point null hypothesis

Let the observable data, $\mathbf{y} = (y_1, y_2, \cdots, y_n)' \in \mathbf{Y}$. A probability model $M \equiv \{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})\}$ is used to fit the data. We are concerned with a point null hypothesis testing problem which may arise from the prediction of a particular theory. Let $\boldsymbol{\theta} \in \Theta$ denote a vector of p-dimensional parameters of interest and $\boldsymbol{\psi} \in \Psi$ a vector of q-dimensional nuisance parameters. The problem of testing a point null hypothesis is given by

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

The hypothesis testing may be formulated as a decision problem. It is obvious that the decision space has two statistical decisions, to accept H_0 (name it d_0) or to reject H_0 (name it d_1). Let $\{\mathcal{L}[d_i, (\boldsymbol{\theta}, \boldsymbol{\psi})], i = 0, 1\}$ be the loss function of statistical decision. Hence, a natural statistical decision to reject H_0 can be made when the expected posterior loss of accepting H_0 is sufficiently larger than the expected posterior loss of rejecting H_0, i.e.,

$$\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = \int_\Theta \int_\Psi \{\mathcal{L}[d_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] - \mathcal{L}[d_1, (\boldsymbol{\theta}, \boldsymbol{\psi})]\} \, p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} > c \geq 0,$$

where $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ is a Bayesian test statistic; $p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})$ the posterior distribution with some given prior $p(\boldsymbol{\theta}, \boldsymbol{\psi})$; c a threshold value. Let $\triangle\mathcal{L}[H_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] = \mathcal{L}[d_0, (\boldsymbol{\theta}, \boldsymbol{\psi})] - \mathcal{L}[d_1, (\boldsymbol{\theta}, \boldsymbol{\psi})]$ be the net loss difference function which can generally be used to measure the evidence against H_0 as a function of $(\boldsymbol{\theta}, \boldsymbol{\psi})$. Hence, the Bayesian test statistic can be rewritten as

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

2.2 A literature review

BF is defined as the ratio of the two marginal likelihood functions, namely,

$$BF_{01} = \frac{p(\mathbf{y}|M_0)}{p(\mathbf{y}|M_1)},$$

where $M_0 := \{p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}), \boldsymbol{\psi} \in \boldsymbol{\Psi}\}$ is the model under the null; $M_1 := M$ is the model under the alternative. The two marginal likelihood functions are defined as

$$p(\mathbf{y}|M_0) = \int_\Psi p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}_0) \mathrm{d}\boldsymbol{\psi},$$
$$p(\mathbf{y}|M_1) = \int_\Theta \int_\Psi p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi}.$$

BF corresponds to the use of the zero-one discrete loss function, namely,

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

and in this case, with $c = 0$, we

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

Remark 2.1 *BF has several disadvantages. If the Jeffreys or the reference prior (Jeffreys, 1961) is used to reflect the objectiveness, BF is not well-defined since it depends on an arbitrary constant (BR, 2002). In addition, if a proper prior with a large spread is used to represent the prior ignorance, BF has a tendency to favor the null hypothesis, giving rise to Jeffreys-Lindley's paradox; see Poirier (1995), Robert (1993, 2001). Moreover, for many models in economics, such as latent variable models and the dynamic general equilibrium models, the marginal likelihood and hence BF are very difficult to evaluate; see Han and Carlin (2001) for a good review of methods for calculating BF from the MCMC outputs. Moreover, Jeffreys's scales are often used to interpret BF and logarithmic BF (Kass and Raftery, 1995). However, the interpretation lacks of statistical justification.*

BR (2002) suggested using a continuous loss function based on the KL divergence given by

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

where $p(x)$ and $q(x)$ are any two regular probability density functions. The corresponding Bayesian test statistic is:

$$\mathbf{T}_{BR}(\mathbf{y}, \boldsymbol{\theta}_0) = E_{\boldsymbol{\vartheta}|\mathbf{y}} \left(\min \left\{ KL\left[p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}), p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) \right], KL\left[p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}), p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \right] \right\} \right). \tag{3}$$

Remark 2.2 *It is shown in BR (2002) that $\mathbf{T}_{BR}(\mathbf{y}, \boldsymbol{\theta}_0)$ is well-defined under improper distributions. This is an important advantage over the BF. However, the BR test is not without its problems. First, the KL divergence function often does not have a closed-form expression. Consequently, $\mathbf{T}_{BR}(\mathbf{y}, \boldsymbol{\theta}_0)$ may be difficult to compute. Second, BR suggested using some subjective and arbitrary threshold values to implement the test. Unfortunately, the choice of these threshold values has not been justified.*

To alleviate the computational problems of $\mathbf{T}_{BR}(\mathbf{y}, \boldsymbol{\theta}_0)$ in the context of latent variable models, LY (2012) proposed a new loss difference function, based on the $\mathcal{Q}$ function used in the EM algorithm (Dempster, Laird and Rubin, 1977). Let $\mathbf{z} = (z_1, z_2, \cdots, z_n)'$ denote the latent variables and $\mathbf{x} = (\mathbf{y}', \mathbf{z}')'$. Let $p(\mathbf{y}|\boldsymbol{\vartheta})$ and $p(\mathbf{x}|\boldsymbol{\vartheta})$ $(:= p(\mathbf{y}, \mathbf{z}|\boldsymbol{\vartheta}))$ be the observed data likelihood function and the complete data likelihood function, respectively. The relationship between these two likelihood functions is

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

For any $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$, the $\mathcal{Q}$ function is:

$$\mathcal{Q}\left(\boldsymbol{\vartheta}_1|\boldsymbol{\vartheta}_2\right) = E_{\mathbf{z}|\mathbf{y},\boldsymbol{\vartheta}_2}\left[\log p(\mathbf{y},\mathbf{z}|\boldsymbol{\vartheta}_1)\right].$$

Compared with the observed data likelihood function $p(\mathbf{y}|\boldsymbol{\vartheta})$, the $\mathcal{Q}$ function is easier to evaluate in latent variable models. In particular, when the analytical expression of $p(\mathbf{y}|\boldsymbol{\vartheta})$ is not available, the $\mathcal{Q}$ function can be easily approximated from the MCMC outputs via,

$$\mathcal{Q}\left(\boldsymbol{\vartheta}_1|\boldsymbol{\vartheta}_2\right) \approx \frac{1}{S}\sum_{s=1}^{S}\log p\left(\mathbf{y},\mathbf{z}^{(s)}|\boldsymbol{\vartheta}_1\right),$$

where $\{\mathbf{z}^{(s)}, s = 1, 2, \cdots, S\}$ are the effective MCMC draws from the posterior distribution $p(\mathbf{z}|\mathbf{y},\boldsymbol{\vartheta}_2)$. Let $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0,\boldsymbol{\psi})$. LY (2012) defined a new continuous net loss difference function as:

$$\triangle\mathcal{L}(\boldsymbol{\vartheta},\boldsymbol{\vartheta}_0) = \{\mathcal{Q}(\boldsymbol{\vartheta},\boldsymbol{\vartheta}) - \mathcal{Q}(\boldsymbol{\vartheta}_0,\boldsymbol{\vartheta})\} + \{\mathcal{Q}(\boldsymbol{\vartheta}_0,\boldsymbol{\vartheta}_0) - \mathcal{Q}(\boldsymbol{\vartheta},\boldsymbol{\vartheta}_0)\},$$

and proposed a Bayesian test statistic as:

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

Remark 2.3 *It is shown in LY (2012) that the test statistic, $T_{LY}(\mathbf{y},\boldsymbol{\theta}_0)$, is well-defined under improper priors and also easy to compute. However, this test statistic has some practical disadvantages. First, like the test statistic of BR, some threshold values have to be specified. Following the idea of McCulloch (1989), LY (2012) proposed to choose threshold values based on the Bernoulli distribution. Such a choice is not justified, unfortunately. Second, it is not clear whether this test statistic is immune to Jeffreys-Lindley's paradox.*

Aiming to alleviate Jeffreys-Lindley's paradox, LZY (2014) developed an alternative Bayesian test statistic based on the Bayesian deviance. The net loss function and the test statistic are given, respectively, by

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

$$\mathbf{T}_{LZY}(\mathbf{y},\boldsymbol{\theta_0}) = 2\int\left[\log p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi}) - \log p(\mathbf{y}|\boldsymbol{\theta}_0,\boldsymbol{\psi})\right]p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}. \tag{4}$$

$\mathbf{T}_{LZY}$ can be understood as the Bayesian version of the likelihood ratio test. However, for latent variable models, the likelihood function $p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi})$ generally is not available in closed-form. To achieve computational tractability, under some regularity conditions, LZY (2014) gave an asymptotically equivalent form for $\mathbf{T}_{LZY}(\mathbf{y},\boldsymbol{\theta_0})$, i.e.,

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

where $\bar{\boldsymbol{\vartheta}} = (\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\psi}})'$ is the posterior mean of $\boldsymbol{\vartheta}$ under H_1, $\bar{\boldsymbol{\vartheta}}_* = (\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}})'$, $\bar{\boldsymbol{\vartheta}}_b = (1-b)\bar{\boldsymbol{\vartheta}}_* + b\bar{\boldsymbol{\vartheta}}$, for $b \in [0,1]$, $S(\mathbf{x}|\boldsymbol{\vartheta}) = \partial \log p(\mathbf{x}|\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}$, $D = \int_0^1 \left\{(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \left[E_{\mathbf{z}|\mathbf{y},\bar{\boldsymbol{\vartheta}}_b}\left(S_1(\mathbf{x}|\bar{\boldsymbol{\vartheta}}_b)\right)\right]\right\} \mathrm{d}b$ the subvector of $S(\mathbf{x}|\boldsymbol{\vartheta})$ corresponding to $\boldsymbol{\theta}$, $V_{22}(\bar{\boldsymbol{\vartheta}}) = E[(\boldsymbol{\psi}-\bar{\boldsymbol{\psi}})(\boldsymbol{\psi}-\bar{\boldsymbol{\psi}})'|\mathbf{y}, H_1]$, the submatrix of $V(\bar{\boldsymbol{\vartheta}})$ corresponding to $\boldsymbol{\psi}$, and $L_{0n}^{(2)}(\boldsymbol{\psi}) = \partial^2 \log p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0)/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'$.

Remark 2.4 *As shown in LZY (2014),* $\mathbf{T}^*_{LZY}(\mathbf{y}, \boldsymbol{\theta}_0)$ *appeals in four aspects. First, it is well-defined under improper priors. Second, it does not suffer from Jeffreys-Lindley's paradox and, hence, can be used under noninformative vague priors. Third, it is easy to compute. Furthermore, for latent variable models,* $\mathbf{T}^*_{LZY}(\mathbf{y}, \boldsymbol{\theta}_0)$ *only involves the first and the second derivatives which is easy to evaluate from the MCMC outputs with the help of the EM algorithm. Finally, LZY (2014) derived the asymptotic distribution of* $\mathbf{T}^*_{LZY}(\mathbf{y}, \boldsymbol{\theta}_0)$*. When* $\boldsymbol{\theta}$ *and* $\bar{\boldsymbol{\psi}}$ *are orthogonal, the asymptotic distribution is determined by the chi-squared distribution. In this case the test is pivotal and the thresholds can be obtained form the asymptotic distribution. Unfortunately, in general the test is not pivotal because the asymptotical distribution depends on some unknown population parameters.*

3 Bayesian Hypothesis Testing Based on a Quadratic Loss

3.1 The test statistic

To deal with the non-pivotal problem, in this section, we develop a new Bayesian test statistic for hypothesis testing. The new statistic shares all the nice features of the LZY statistic. First, it is motivated from the decision-theoretic perspective. Second, it is well-defined under improper prior distributions. Second, it is immune to Jeffreys-Lindley's paradox. Fourth, it is easy to compute. However, unlike the LZY statistic, the new statistic is pivotal and hence the threshold can be easily obtained from its asymptotic distribution.

To fix the idea, let

$$s(\boldsymbol{\vartheta}) = \frac{\partial \log p(\mathbf{y}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}, C(\boldsymbol{\vartheta}) = s(\boldsymbol{\vartheta})s(\boldsymbol{\vartheta})',$$

where $s(\boldsymbol{\vartheta})$ is the score function and $\boldsymbol{\vartheta} = (\boldsymbol{\theta}, \boldsymbol{\psi})$. We define a quadratic loss function as:

$$\triangle\mathcal{L}[H_0, \boldsymbol{\vartheta}] = (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}), \tag{5}$$

where $C_{\theta\theta}(\boldsymbol{\vartheta})$ is the submatrix of $C(\boldsymbol{\vartheta})$ corresponding to $\boldsymbol{\theta}$ and is semi-positive definite, $\bar{\boldsymbol{\vartheta}}_0 = (\boldsymbol{\theta}_0, \bar{\boldsymbol{\psi}}_0)$ is the Bayesian estimator of $\boldsymbol{\vartheta}$ under H_0, $\bar{\boldsymbol{\theta}}$ is the Bayesian estimator of $\boldsymbol{\theta}$ under H_1. Based on this quadratic loss, we propose the following Bayesian test statistic:

$$\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = \int \triangle\mathcal{L}[H_0, \boldsymbol{\vartheta}] p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} = \int (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}, \tag{6}$$

where $p(\boldsymbol{\vartheta}|\mathbf{y})$ is the posterior distribution of $\boldsymbol{\vartheta}$ under H_1.

Remark 3.1 *Clearly* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta_0})$ *depends on the posterior distribution directly. The prior information only influences the test statistic via the posterior distribution.*

Remark 3.2 *Since the posterior distribution* $p(\boldsymbol{\vartheta}|\mathbf{y})$ *is independent of an arbitrary constant in the prior distributions, both* $s(\boldsymbol{\vartheta})$ *and* $C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)$ *are independent of the arbitrary constant. As a result,* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta_0})$ *is well-defined under improper priors.*

Remark 3.3 *To see how the new statistic can avoid Jeffreys-Lindley's paradox, consider the example discussed in LZY (2014). Let* $y \sim N(\theta, \sigma^2)$ *with a known* σ^2 *and we test the null hypothesis* $H_0 : \theta = 0$*. Let the prior distribution of* θ *be* $N(\mu, \tau^2)$ *with* $\mu = 0$*. LZY showed that the posterior distribution of* θ *is* $N(\mu(y), \omega^2)$ *with*

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

and BF is

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

As $\tau^2 \to +\infty$, $BF_{10} \to 0$*, suggesting the test always supports* H_0*, whether or not* H_0 *holds true, giving rise to Jeffreys-Lindley's paradox. On the other hand, it is easy to show that*

$$C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0) = \frac{y^2}{\sigma^4}, \text{ and } \mathbf{T}(y, 0) = \frac{y^2}{\sigma^4}\int (\theta - \bar{\theta})^2 p(\theta|y)\, d\theta = \frac{\omega^2 y^2}{\sigma^4}.$$

As $\tau^2 \to +\infty$, $\mu(y) \to y$, $\omega^2 \to \sigma^2$*, and, hence,* $\mathbf{T}(y, 0) \to y^2/\sigma^2$ *which is distributed as* $\chi^2(1)$ *when* H_0 *is true. Consequently, our proposed test statistic is immune to Jeffreys-Lindley's paradox.*

Remark 3.4 *To calculate* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta_0})$*, the first derivatives of the observed-data likelihood function must be evaluated. For most latent variable models, the first derivatives are difficult to evaluate directly because the observed-data likelihood function is not available in closed-form. Fortunately, with the help of the EM algorithm, the first derivatives can be easily approximated from the MCMC outputs in connection with the data augmentation technique. For any* $\boldsymbol{\vartheta}$ *and* $\boldsymbol{\vartheta}^*$ *in the support space of* $\boldsymbol{\vartheta}$*, it was shown in Dempster et al. (1977) that*

$$s(\boldsymbol{\vartheta}) = \frac{\partial \log p(\mathbf{y}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \frac{\partial \mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta}}|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*} = \int \frac{\partial \log p(\mathbf{y}, \mathbf{z}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} p(\boldsymbol{z}|\mathbf{y}, \boldsymbol{\vartheta})\, d\boldsymbol{z}.$$

Hence, based on the MCMC outputs, the first derivative can be approximated by:

$$s(\boldsymbol{\vartheta}) \approx \frac{1}{M}\sum_{m=1}^{M}\left\{\frac{\partial \log p(\mathbf{y}, \boldsymbol{z}^{(m)}|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right\},$$

where $\{\boldsymbol{z}^{(m)}, m = 1, 2, \cdots, M\}$ *are effective MCMC draws from the posterior distribution* $p(\boldsymbol{z}|\mathbf{y}, \boldsymbol{\vartheta})$ *due to the use of data augmentation.*

3.2 The threshold value

To implement the proposed test, a threshold value, c, has to be specified, i.e.,

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

This section obtains the asymptotic distribution of the test statistic under H_0 and establishes the link between the test statistic and the LM test. To do so, following LZY (2014), we first impose a set of regularity conditions.

Assumption 1: There exists a finite sample size n^*, so that, for $n > n^*$, there is a local maximum at $\hat{\boldsymbol{\vartheta}}$ (i.e., posterior mode) 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 2: The largest eigenvalue λ_n of $-L_n^{-(2)}(\hat{\boldsymbol{\vartheta}})$ goes to zero when $n \to \infty$.

Assumption 3: For any $\epsilon > 0$, there exists an integer N and some $\delta > 0$ such that for any $n > \max\{N, n^*\}$ and $\boldsymbol{\vartheta} \in H(\hat{\boldsymbol{\vartheta}}, \delta) = \{\boldsymbol{\vartheta} : ||\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}|| \leq \delta\}$, $L_n^{(2)}(\boldsymbol{\vartheta})$ exists and satisfies

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

where $\mathbf{E}_{p+q}$ is an identity matrix and $A(\epsilon)$ is a positive semidefinite symmetric matrix whose largest eigenvalue goes to zero as $\epsilon \to 0$.

Assumption 4: For any $\delta > 0$, as $n \to \infty$,

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

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

Assumption 5: The likelihood of the models under both the null hypothesis and the alternative hypothesis is regular so that the standard maximum likelihood (ML) theory can be applied. Furthermore, if the null hypothesis is true, let $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\psi}_0)$ be true value of $\boldsymbol{\vartheta}$, as $n \to \infty$, for any null sequence $k_n \to 0$, so that,

$$\sup_{||\vartheta - \vartheta_0|| < k_n} n^{-1} ||\mathbf{I}(\boldsymbol{\vartheta}) - \mathbf{I}(\boldsymbol{\vartheta}_0)|| \xrightarrow{p} 0,$$

where $\mathbf{I}(\boldsymbol{\vartheta}) = \partial^2 \log p(\mathbf{y}|\boldsymbol{\vartheta})/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'$.

Remark 3.5 *In the literature, Assumption 1-4 have been used to develop the Bayesian large sample theory; see, for example, Chen (1985). Assumption 5 is a fundamental regularity condition for developing the standard ML theory. Based on these regularity conditions, LZY (2014) showed that*

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

$$V(\hat{\boldsymbol{\vartheta}}) = E\left[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})' |\mathbf{y}, H_1\right] = -L_n^{-(2)}(\hat{\boldsymbol{\vartheta}}) + o_p(n^{-1}).$$

When the null hypothesis holds, we also have

$$\bar{\boldsymbol{\psi}}_0 = E\left[\boldsymbol{\psi}|\mathbf{y}, H_0\right] = \int \boldsymbol{\psi} p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\theta}_0) d\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}_0 + o_p(n^{-1/2}),$$
$$V_0(\hat{\boldsymbol{\psi}}_0) = E\left[(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_0)(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_0)'|\mathbf{y}, H_0\right] = -L_{0n}^{-(2)}(\hat{\boldsymbol{\psi}}_0) + o_p(n^{-1}),$$

where $L_{0n}^{(2)}(\boldsymbol{\psi}_0) = \log p(\boldsymbol{\psi}|\boldsymbol{\theta}_0, \mathbf{y})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'|_{\boldsymbol{\psi}=\boldsymbol{\psi}_0}$ *and* $\hat{\boldsymbol{\psi}}_0$ *is the local maximum of* $\log p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\theta}_0)$ *under* H_0*.*

Lemma 3.1 *Let*

$$\mathbf{I}(\boldsymbol{\vartheta}) = \frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}, \mathbf{J}(\boldsymbol{\vartheta}) = \mathbf{I}^{-1}(\boldsymbol{\vartheta}).$$

When the null hypothesis is true, and $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\psi}_0)$ *is the true value of* $\boldsymbol{\vartheta}$*, for any consistent estimator* $\tilde{\boldsymbol{\vartheta}}$ *of* $\boldsymbol{\vartheta}$*, we have*

$$\mathbf{I}(\boldsymbol{\vartheta}_0) = O_p(n), \mathbf{I}(\tilde{\boldsymbol{\vartheta}}) = \mathbf{I}(\boldsymbol{\vartheta}_0) + o_p(n) = O_p(n),$$
$$\mathbf{J}(\boldsymbol{\vartheta}_0) = O_p(n^{-1}), \mathbf{J}(\tilde{\boldsymbol{\vartheta}}) = \mathbf{J}(\boldsymbol{\vartheta}_0) + o_p(n^{-1}) = O_p(n^{-1}).$$

Lemma 3.2 *Let* $\hat{\boldsymbol{\vartheta}}_0 = (\boldsymbol{\theta}_0, \hat{\boldsymbol{\psi}}_0)$ *be the posterior mode of* $\boldsymbol{\vartheta}$ *under the null hypothesis. Under Assumptions 1-5 and when the null hypothesis is true, we have*

$$s(\hat{\boldsymbol{\vartheta}}_0) = O_p(n^{1/2}), s(\bar{\boldsymbol{\vartheta}}_0) = O_p(n^{1/2}), C(\hat{\boldsymbol{\vartheta}}_0) = O_p(n),$$
$$C(\bar{\boldsymbol{\vartheta}}_0) = C(\hat{\boldsymbol{\vartheta}}_0) + o_p(n) = O_p(n).$$

Let the LM statistic (Breusch and Pagan, 1980) be

$$\mathbf{LM} = s_\theta(\hat{\boldsymbol{\vartheta}}_{m0})\left[-\mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}}_{m0})\right] s_\theta(\hat{\boldsymbol{\vartheta}}_{m0}),$$

where $\hat{\boldsymbol{\vartheta}}_{m0} = (\boldsymbol{\theta}_0, \hat{\boldsymbol{\psi}}_{m0})$ is the ML estimator of $\boldsymbol{\vartheta}$ under the null hypothesis, $s_\theta(\boldsymbol{\vartheta})$ is the score function corresponding to $\boldsymbol{\theta}$, $\mathbf{J}_{\theta\theta}(\boldsymbol{\vartheta})$ is the submatrix of $\mathbf{J}(\hat{\boldsymbol{\vartheta}})$ corresponding to $\boldsymbol{\theta}$.

Theorem 3.1 *Under Assumptions 1-5, we can show that*

$$\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = s_\theta(\hat{\boldsymbol{\vartheta}}_0)\left[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})\right] s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1), \quad (7)$$

where $L_{n,\theta\theta}^{-(2)}$ *is the submatrix of* $L_n^{-(2)}$ *corresponding to* $\boldsymbol{\theta}$*. Furthermore, when the null hypothesis is true and the likelihood dominates the prior, we have*

$$\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = \mathbf{LM} + o_p(1) \xrightarrow{d} \chi^2(p). \quad (8)$$

Remark 3.6 *From Equation (8),* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ *may be regarded as the Bayesian version of the LM statistic. However, LM is a frequentist test which is based on the ML estimation of the model in the null hypothesis whereas our test is a Bayesian test which is based on the posterior quantities of the models under both the null hypothesis as well as the alternative hypothesis.*

Remark 3.7 *In Theorem 3.1, we can see that under the null hypothesis, the asymptotic distribution of* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ *always follows the* χ^2 *distribution and, hence, is independent of the nuisance parameters. This suggests that the new test is pivotal, a property that compares favorably with the use of the subjective threshold values as in BF (Jeffreys, 1961), BR (2002), LY (2012) or LZY (2014).*

Remark 3.8 *When the likelihood dominates the prior, the posterior mode,* $\hat{\boldsymbol{\vartheta}}$*, reduces to the ML estimator of* $\boldsymbol{\vartheta}$ *under the alternative hypothesis, and the posterior mode,* $\hat{\boldsymbol{\vartheta}}_0 = (\boldsymbol{\theta}_0, \hat{\boldsymbol{\psi}}_0)$*, reduces to the ML estimator of* $\boldsymbol{\vartheta}$ *under the null hypothesis. From Equation (7), we can see that*

$$\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = s_\theta(\hat{\boldsymbol{\vartheta}}_0)\left[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})\right]s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) = -s_\theta(\hat{\boldsymbol{\vartheta}}_0)\left[\mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}})\right]s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1).$$

If the null hypothesis is false, according to the standard ML theory, we get

$$\mathbf{J}(\boldsymbol{\vartheta}_0) = \mathbf{J}(\hat{\boldsymbol{\vartheta}}) + o_p(n^{-1}) \neq \mathbf{J}(\hat{\boldsymbol{\vartheta}}_0) + o_p(n^{-1}).$$

This is because, under the alternative, $\hat{\boldsymbol{\vartheta}}$ *is a consistent estimator of* $\boldsymbol{\vartheta}$ *whereas* $\hat{\boldsymbol{\vartheta}}_0$ *is not. Hence,*

$$\begin{aligned}
\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) &= -s_\theta(\hat{\boldsymbol{\vartheta}}_0)'\mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}})s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
&\neq -s_\theta(\hat{\boldsymbol{\vartheta}}_0)'\mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}}_0)s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
&= \mathbf{LM} + o_p(1).
\end{aligned}$$

The implementation of the LM test requires the ML estimation of the null model. When it is hard to do the ML estimation, it will be difficult to calculate the LM statistic. This is the case for many models that involve latent variables. However, as long as the Bayesian MCMC methods are applicable, our test can be implemented. Moreover, our method offers two additional advantages over the LM test, which we explain below.

Remark 3.9 *We have shown that when the alternative hypothesis is correct, our test statistic is not close to the LM test. In this case, our test continues to take a positive value whereas the LM test can take a negative value. This is because, in our test, the weight matrix* $-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})$ *remains positive definite as* $\hat{\boldsymbol{\vartheta}}$ *is consistent. When* $\boldsymbol{\theta}_0$ *is further* *away from the true value of $\boldsymbol{\theta}$, $s_{\theta}(\hat{\boldsymbol{\vartheta}}_0)$ will be further way from zero. Consequently, $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ will be larger so that it can discriminate H_0 against H_1. Whereas, when $\boldsymbol{\theta}_0$ is further away from the true value of $\boldsymbol{\theta}$, the weight matrix $-\mathbf{I}(\boldsymbol{\vartheta}_0)$ in the LM statistic may not be positive definite. This may cause some difficulties in the use of the LM test.*

To illustrate the remark, consider the following example where $y_t \sim N(0, \sigma^2), t = 1, 2 \cdots, n$, and the true value of σ^2 is 0.1. We would like to test

$$H_0 : \sigma^2 = 1, H_1 : \sigma^2 \neq 1.$$

In this case, we have

$$\mathbf{I}(\boldsymbol{\vartheta}) = \mathbf{I}(\sigma^2) = \frac{\partial^2 \log p(\mathbf{y}|\sigma^2)}{\partial\sigma^2\partial\sigma^2} = \frac{n}{2\sigma^4} - \frac{\sum_{t=1}^n y_t^2}{\sigma^6}.$$

When n is large enough, we know that $\sum_{t=1}^n y_t^2/n \approx 0.1$ and, hence,

$$\mathbf{I}(\hat{\boldsymbol{\vartheta}}_0) = \mathbf{I}(\sigma^2 = 1) = \frac{n}{2} - \sum_{t=1}^n y_t^2 = \frac{n}{2}\left(1 - 2\frac{\sum_{t=1}^n y_t^2}{n}\right) \approx 0.8n > 0,$$
$$-\mathbf{J}(\hat{\boldsymbol{\vartheta}}_0) = \frac{1}{-\mathbf{I}(\hat{\boldsymbol{\vartheta}}_0)} = -\frac{1}{0.8n} < 0.$$

Consequently, the LM test statistic is negative. Whereas, for our statistic, we have

$$\hat{\sigma}^2 = \sum_{t=1}^n \frac{y_t^2}{n}, -\mathbf{I}(\hat{\boldsymbol{\vartheta}}) = \frac{n}{2\left(\hat{\sigma}^2\right)^2}, -\mathbf{J}\left(\boldsymbol{\vartheta}\right) = \frac{2\left(\hat{\sigma}^2\right)^2}{n}.$$

Hence, out test statistic does not suffer from the problem of taking a negative value.

Remark 3.10 *The implementation of the LM test requires the inversion of $-\mathbf{I}(\boldsymbol{\vartheta}_0)$. When the dimension of $\boldsymbol{\vartheta}$ is high, such an inversion may be numerically challenging. Whereas, to calculate $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$, one does not need to invert any matrix.*

4 Empirical Illustrations

In this section, we illustrate the proposed test using three popular examples in economics and finance. The first example is a simple linear regression model where BF is easy to calculate. We hope to compare BF and $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$, paying particular attentions to the sensitivity of the statistics to the choice of prior. The second example is a probit model where there are latent variable models. However, the observed data likelihood is available in closed-form in the probit model, facilitating the implementation of the LM test. In this example, we will test a point null hypothesis using both the LM test and the proposed Bayesian test. The third example is a stochastic conditional duration (SCD) model, where latent variable models are also in presence. However, in this example the observed data likelihood is not available in closed-form even for the model under the null, making the implementation of the LM test difficult. Since the complete data likelihood is available in closed-form, we can use MCMC to estimate the models. We will show how to implement our Bayesian test in this case.

4.1 Hypothesis testing in linear regression models

The first example is the simple linear regression model:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim i.i.d.\ N\left(0, \sigma^2\right), i = 1, ..., n. \tag{9}$$

We would like to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$. Assume the prior distributions for (α, β) and σ^2 are normal and inverse gamma, respectively,

$$(\alpha, \beta)' \sim N\left(\tilde{\mu}, \sigma^2 \tilde{V}\right), \quad \sigma^2 \sim IG(a, b),$$

where $\tilde{\mu} = (\mu_\alpha, \mu_\beta)'$, $\tilde{V} = diag\left(V_\alpha, V_\beta\right)$.

The marginal likelihood for the model under H_0 is

$$p_0\left(\mathbf{y}\right) = \frac{b^a \Gamma\left(a + \frac{n}{2}\right)}{(2\pi)^{\frac{n}{2}} \Gamma\left(a\right)} \sqrt{\frac{1}{nV_\alpha + 1}} \left[b + \frac{1}{2}\left(\left(\mathbf{y} - \beta_0 \mathbf{x}\right)' \left(\mathbf{y} - \beta_0 \mathbf{x}\right) + \frac{\mu_\alpha^2}{V_\alpha} - \frac{\mu_\alpha^*}{V_\alpha^*}\right)\right]^{-\left(a + \frac{n}{2}\right)}, \tag{10}$$

where $V_\alpha^* = \frac{V_\alpha}{nV_\alpha + 1}$, $\mu_\alpha^* = V_\alpha^* \left(\sum_{i=1}^n \left(y_i - \beta_0 x_i\right) + \frac{\mu_\alpha}{V_\alpha}\right) = V_\alpha^* \left(\iota' \left(\mathbf{y} - \beta_0 \mathbf{x}\right) + \frac{\mu_\alpha}{V_\alpha}\right)$ with $n \times 1$ vector $\iota = (1, ..., 1)'$. The marginal likelihood for the model under H_1 is:

$$p_1\left(\mathbf{y}\right) = \frac{b^a \sqrt{|V^*|} \Gamma\left(a + \frac{n}{2}\right)}{(2\pi)^{\frac{n}{2}} \Gamma\left(a\right) \sqrt{|\tilde{V}|}} \left[b + \frac{1}{2}\left((\tilde{\mu})' \tilde{V}^{-1} \tilde{\mu} + \mathbf{y}'\mathbf{y} - (\mu^*)' V^{*-1} \mu^*\right)\right]^{-\left(a + \frac{n}{2}\right)}, \tag{11}$$

where $V^* = \left(\tilde{V}^{-1} + X'X\right)^{-1}$, $\mu^* = V^* \left(\tilde{V}^{-1} \tilde{\mu} + X'\mathbf{y}\right)$, $X = (\iota, \mathbf{x})$. Equations (10) and (11) are derived in Appendix. In this simple model, $BF_{01} = p_0\left(\mathbf{y}\right) / p_1\left(\mathbf{y}\right)$ has an analytical expression.

To calculate our proposed statistic, note that $\boldsymbol{\vartheta} = (\alpha, \beta, \sigma^2)'$ and $\boldsymbol{\theta} = \beta$. Given the posterior sample $\left\{\boldsymbol{\vartheta}^{(j)}\right\}_{j=1}^M = \left\{\alpha^{(j)}, \beta^{(j)}, \sigma^{2(j)}\right\}_{j=1}^M$ under H_1, the approximation of the statistic is

$$\hat{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right) = \frac{1}{M} C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right) \sum_{j=1}^M \left(\beta^{(j)} - \overline{\beta}\right)^2, \tag{12}$$

where $\overline{\beta} = \frac{1}{M} \sum_{j=1}^M \beta^{(j)}$ and

$$C_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}_0\right) = \frac{1}{(\overline{\sigma_0^2})^2} \left[\mathbf{x}' \left(\mathbf{y} - \overline{\alpha}_0 \iota - \beta_0 \mathbf{x}\right)\right]^2.$$

where $\overline{\alpha}_0$ and $\overline{\sigma_0^2}$ are the posterior means of α and σ^2 under H_0. The derivation of Equation (12) is given in Appendix.

We now analyze a model in Brooks (2008, Page 40) where the return on a spot price is linked to the return on a futures price, i.e.,

$$\Delta \log(s_t) = \alpha + \beta \Delta \log(f_t) + \varepsilon_t,\ \varepsilon_t \sim i.i.d.N(0, \sigma^2),$$

where $\Delta \log(s_t)$ is the log-difference of the spot S&P500 index and $\Delta \log(f_t)$ is the log-difference of the S&P500 futures price, and β captures the optimal hedge ratio. We would like to test if $\beta = 1$.

The hyperparameters are set at

$$\mu_a = 0,\ V_a = 10^3,\ \mu_\beta = 0,\ a = 0.001,\ b = 0.001.$$

In addition, we allow the prior variance of β, V_β, to vary so that we can examine how the prior influences BF and $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$. Since both the priors and the likelihood function are in the Normal-Gamma form, we can directly draw MCMC samples from their posterior joint distributions under H_0 and H_1. In particular, 50,000 random draws are sampled from the posterior distributions.

Table 1 reports $\log BF_{01}$, $\hat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0)$, the posterior means and the posterior standard errors of all the parameters under H_1 for different values of V_β. From Table 1, we observe that the posterior quantities of all three parameters are robust to V_β. However, $\log BF_{01}$ is sensitive to V_β. In particular, $\log BF_{01}$ increases as V_β increases. When the prior variance V_β is moderate, $\log BF_{01}$ is less than 0 and tends to reject the null hypothesis. When V_β is sufficiently large, $\log BF_{01}$ is larger than 0 and does not reject the null hypothesis. This observation clearly demonstrates that BF is subject to Jeffreys-Lindley's paradox. On the contrary, $\hat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0)$ takes nearly identical values with different V_β. Therefore, $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ is immune to Jeffreys-Lindley's paradox. The asymptotic distribution of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ under H_0 is $\chi^2(1)$, and the 99.9% percentile of $\chi^2(1)$ is 10.83. $\hat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0)$ is much larger than 10.83 in all cases, suggesting that the null hypothesis is rejected under the 99.9% probability level.

4.2 Hypothesis testing in discrete choice models

The probit model is widely used to analyze binary choice data. In this section, we fit the probit model to a dataset originally used in Mroz (1987). Since the observed data likelihood in the probit model is available in closed-form, we can directly compute the proposed Bayesian test statistic $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_\mathbf{0})$ based on the MCMC outputs. Also, the LM test can be easily obtained.

In the probit model, we take the married women's labor force participation ($inlf$) as the binary dependent variable and $nwifeinc, educ, exper, expersq, age, kedslt6$, and $kidsge6$

Table 1: Bayesian estimates of $\boldsymbol{\vartheta}$ under H_1, BF and the proposed test statistic.

	$V_\beta = 0.1$	$V_\beta = 100$	$V_\beta = 10^5$	$V_\beta = 10^{22}$	$V_\beta = 10^{25}$	$V_\beta = 10^{35}$
$\log BF_{01}$	-14.7354	-11.2948	-7.8409	11.7311	15.1849	26.6979
$\hat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0)$	14.9558	15.2747	15.1514	15.1569	15.1823	15.1826
$\overline{\beta}$	0.1229	0.1231	0.1246	0.1231	0.1236	0.1242
$SE(\beta)$	0.1329	0.1343	0.1337	0.1338	0.1339	0.1339
$\overline{\alpha}$	0.3622	0.3633	0.3669	0.3626	0.3639	0.3607
$SE(\alpha)$	0.4430	0.4427	0.4459	0.4458	0.4438	0.4429
$\overline{\sigma^2}$	12.6056	12.5683	12.5688	12.5965	12.5802	12.5821
$SE(\sigma^2)$	2.2716	2.2701	2.2779	2.2922	2.2649	2.2840

Table 2: The Bayesian and ML estimates

	Bayesian Method		ML Method	
	Posterior Mean	SE	Estimate	SD
ϑ_0	0.2675	0.5108	0.2701	0.5086
ϑ_1	-1.2186×10^{-2}	4.85×10^{-3}	-1.2024×10^{-2}	4.8398×10^{-3}
ϑ_2	0.132	2.5438×10^{-2}	0.1309	2.5254×10^{-2}
ϑ_3	0.124	1.8755×10^{-2}	0.1233	1.8716×10^{-2}
ϑ_4	-1.8953×10^{-3}	6.0381×10^{-4}	-1.8871×10^{-3}	6×10^{-4}
ϑ_5	-5.3107×10^{-2}	8.4833×10^{-3}	-5.2853×10^{-2}	8.4772×10^{-3}
ϑ_6	-0.8741	0.1182	-0.8683	0.1185
ϑ_7	3.6055×10^{-2}	4.3472×10^{-2}	3.6005×10^{-2}	4.3477×10^{-2}

are taken as independent variables; see Wooldridge (2002) for detailed explanation of these variables. The latent variable representation of the model is given by

$$z = \vartheta_0 + \vartheta_1 nwifeinc + \vartheta_2 educ + \vartheta_3 exper + \vartheta_4 expersq + \vartheta_5 age + \vartheta_6 kedslt6 + \vartheta_7 kidsge6 + e,$$

where z is the latent variable, e follows a standard normal distribution, and $inlf$ takes value 1 if $z > 0$, and 0 otherwise.

To compute the Bayesian estimates and the proposed test statistic, a proper but vague prior is used for all the regression coefficients. Specifically, each element of $\boldsymbol{\vartheta}$ is assumed to follow the normal distribution with mean 0 and variance 10^8. In this example, we test if $exper$ and $expersq$ have explanatory power for y_i. Hence, the parameters of interest are $\boldsymbol{\theta} = (\vartheta_3, \vartheta_4)'$. The null hypothesis is $\boldsymbol{\theta} = \mathbf{0}$.

For the Bayesian analysis, based on Koop (2003), 220,000 draws are obtained using the Gibbs sampler under H_0 and H_1 with the first 20,000 samples being discarded. The

Table 3: The Bayesian test statistic and the LM test statistic

Statistic	$\hat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta_0})$	LM
Value	127.121	99.088

parameter estimates under H_1 for both the Bayesian method and the ML method are reported in Table 2. For the Bayesian method, we report the posterior means and the posterior standard errors. For the ML method, we report the ML estimates and the asymptotic standard deviations. Since the likelihood of the probit model has an analytical form, the approximation of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta_0})$ can be easily obtained and is reported in Table 3. The derivation of $\hat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta_0})$ is given in Appendix. For the same reason, the LM test can be easily obtained and also reported in Table 3.

The asymptotic distribution of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ under H_0 is $\chi^2(2)$ and the 99.99% percentiles of $\chi^2(2)$ is 18.42. The approximated value $\hat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta_0})$ is much larger than 18.42 and suggests that the null hypothesis is rejected under the 99.99% probability level. Similarly, the LM statistic is 99.088, which is also much larger than the 99.99% percentile of $\chi^2(2)$, and rejects the null hypothesis.

4.3 Hypothesis testing in stochastic conditional duration models

The third example is the stochastic conditional duration (SCD) model of Bauwens and Veredas (2004) given by

$$\begin{cases} d_t = \exp(\varphi_t)\,\varepsilon_t & \varepsilon_t \sim \text{Exp}(1) \\ \varphi_t = \mu + \phi(\varphi_{t-1} - \mu) + \sigma\epsilon_t & \epsilon_t \sim N(0,1) \\ \varphi_1 \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right) & \end{cases},$$

for $t = 1, ..., T$, where $|\phi| < 1$, d_t is the adjusted duration between transactions, φ_t is the latent variable which is potentially serially correlated. ε_t and ϵ_t are assumed to be independent. The distribution of ε_t is assumed to follow the exponential distribution with the rate parameter of 1.

The data, collected from the TAQ database, are the time intervals (durations) between transactions for IBM between September 3, 1996 and September 30, 1996 and have 17,237 observations. Following Bauwens and Veredas (2004), the transaction data before 9:30 and after 16:00 are excluded and the simultaneous trades are treated as one single transaction. Consequently, we are left with 17,157 raw durations.

Following Engle and Russell (1998), we adjust the raw durations using the daily season factor $\Psi(t_i)$ which is assumed to be a cubic spline with each node being the average

Table 4: The estimation results of SCD under H_0 and H_1

	H_0		H_1	
	Posterior Mean	Standard Error	Posterior Mean	Standard Error
μ	-0.1100	9.3943×10^{-3}	-0.1071	2.502×10^{-2}
ϕ	-	-	0.9549	4.791×10^{-3}
σ^2	0.2122	1.0776×10^{-2}	1.912×10^{-2}	2.153×10^{-3}

duration on each half hour from 9:30 to 16:00, i.e.,

$$d_{t_i} = \frac{D_{t_i}}{\Psi\left(t_i\right)},$$

where D_{t_i} is the raw durations. We are interested in testing whether or not ψ_t is serially correlated, i.e., $\phi = 0$. Hence, $\boldsymbol{\theta} = \phi$, $\boldsymbol{\vartheta} = \left(\mu, \sigma^2\right)$.

Because the observed-data likelihood function is not available in closed-form, it is very hard to calculate the LM statistic even for the model under the null hypothesis. However, since the complete-data likelihood function has an analytical expression, the data augmentation facilitates the Bayesian MCMC estimation of the models. As a result, the proposed test statistic is easy to calculate and the detailed derivation of $\hat{\mathbf{T}}\left(\mathbf{d}, \boldsymbol{\theta}_0\right)$ is reported in Appendix. For H_1, we use OpenBUGS to obtain the posterior samples of the parameters. 60,000 MCMC draws are obtained with the first 10,000 being treated as burn-in samples. For H_0, we use the Gibbs sampler to draw the parameters and the adaptive rejection Metropolis sampling method (Gilks et.al, 1995) to draw the latent variables. The estimation results under H_0 and H_1 are reported in Table 4.

Our test statistic, $\hat{\mathbf{T}}\left(\mathbf{d}, \boldsymbol{\theta}_0\right)$, takes the value of 4.2353. Under the null, its asymptotic distribution is $\chi^2\left(1\right)$ and the 95% percentiles of $\chi^2\left(1\right)$ is 3.8415. Therefore, at the 95% significant level, the proposed test rejects the null hypothesis that the latent variables are serially uncorrelated.

5 Conclusion

In this paper, we have proposed a new Bayesian test statistic to test a point null hypothesis based on a quadratic loss function. Under the null hypothesis and a set of regularity conditions, we show that our test is asymptotically equivalent to frequentist's LM test and follows a chi-squared distribution asymptotically.

The main advantages of the new statistic can be summarized as follows: (1) it is well-defined under improper prior distributions. (2) it is immune to Jeffreys-Lindley's paradox; (2) it is easy to compute, even for the latent variable models; (3) the asymptotic distribution is pivotal. The proposed method is illustrated using a simple linear regression model, a discrete choice model and a stochastic conditional duration model.

6 Appendix

6.1 Appendix 1: Proof of Lemma 3.1

When the likelihood information dominates the prior information, the posterior mean $\bar{\boldsymbol{\vartheta}}$ reduces to the ML estimator $\hat{\boldsymbol{\vartheta}}$, under the alternative hypothesis. When the null hypothesis is true, let $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\psi}_0)$ be the true value of $\boldsymbol{\vartheta}$. According to the standard ML theory and the central limit theorem, it can be shown that

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \xrightarrow{d} N\left[0, F(\boldsymbol{\vartheta}_0)\right],$$

where $F(\boldsymbol{\vartheta}_0) = n\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\vartheta}_0)$, $\boldsymbol{\mathcal{I}}(\boldsymbol{\vartheta}_0) = -E\left[\mathbf{I}\left(\boldsymbol{\vartheta}_{\mathbf{0}}\right)\right]$ is the Fisher information matrix, and

$$\mathbf{I}(\boldsymbol{\vartheta}) = \frac{\partial^2 \log p\left(\mathbf{y}|\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} = L_n^{(2)}(\boldsymbol{\vartheta}).$$

Under the standard regularity conditions, as $n \to \infty$, we have

$$-n\mathbf{J}(\boldsymbol{\vartheta}_0) \xrightarrow{p} F(\boldsymbol{\vartheta}_0),$$

where $\mathbf{J}(\boldsymbol{\vartheta}_0)$ is the inverse matrix of $\mathbf{I}(\boldsymbol{\vartheta}_0)$. Therefore, it can be shown that

$$\begin{aligned} &\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = O_p(n^{-\frac{1}{2}}), \\ &\mathbf{J}(\boldsymbol{\vartheta}_0) = O_p(n^{-1}), \mathbf{I}(\boldsymbol{\vartheta}_0) = O_p(n). \end{aligned}$$

For any consistent estimator of $\boldsymbol{\vartheta}$, say $\tilde{\boldsymbol{\vartheta}}$, there exists a positive sequence $k_n^* \to 0$ such that $p(||\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0|| \le k_n^*) \ge 1 - k_n^*$. Hence, when n is large enough, we can find some $N > 0$, and $n > N$ to make $||\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0|| \le k_n^*$. Under Assumption 5, we have

$$\frac{1}{n}||\mathbf{I}(\tilde{\boldsymbol{\vartheta}}) - \mathbf{I}(\boldsymbol{\vartheta}_0)|| \le \sup_{||\vartheta - \vartheta_0|| < k_n} \frac{1}{n}||\mathbf{I}(\boldsymbol{\vartheta}) - \mathbf{I}(\boldsymbol{\vartheta}_0)|| \xrightarrow{p} 0.$$

Hence, for any consistent estimator $\tilde{\boldsymbol{\vartheta}}$, $\frac{1}{n}\left[\mathbf{I}(\tilde{\boldsymbol{\vartheta}}) - \mathbf{I}(\boldsymbol{\vartheta}_0)\right] = o_p(1)$ so that $\mathbf{I}(\tilde{\boldsymbol{\vartheta}}) = \mathbf{I}(\boldsymbol{\vartheta}_0) + o_p(n)$ and that $\mathbf{I}(\tilde{\boldsymbol{\vartheta}}) = O_p(n)$. Similarly, $\mathbf{J}(\tilde{\boldsymbol{\vartheta}}) = \mathbf{J}(\boldsymbol{\vartheta}_0) + o_p(n^{-1})$ and $\mathbf{J}(\tilde{\boldsymbol{\vartheta}}) = O_p(n^{-1})$.

6.2 Appendix 2: Proof of Lemma 3.2

When the likelihood information dominates the prior information, the posterior mode $\hat{\boldsymbol{\vartheta}}_0$ of $\boldsymbol{\vartheta}$ under the null hypothesis reduces to the ML estimator of $\boldsymbol{\vartheta}$ under the null hypothesis.

Similar to Lemma 3.1, when the null hypothesis is true, according to the standard ML theory, it can be shown that

$$\frac{1}{\sqrt{n}}s(\boldsymbol{\vartheta}_0) \sim N[0, F(\boldsymbol{\vartheta}_0)],$$
$$\sqrt{n}(\hat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0) \sim N[0, F_{\psi\psi}(\boldsymbol{\vartheta}_0)],$$

where $F_{\psi\psi}(\boldsymbol{\vartheta}_0)$ is the submatrix of $F(\boldsymbol{\vartheta}_0)$ corresponding to $\boldsymbol{\psi}$. Hence, we have

$$s(\boldsymbol{\vartheta}_0) = O_p(n^{1/2}), \hat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0 = O_p(n^{-1/2}), \hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0 = O_p(n^{-1/2}).$$

Furthermore, based on Remark 3.5, it can be shown that

$$\bar{\boldsymbol{\psi}}_0 - \hat{\boldsymbol{\psi}}_0 = o_p(n^{-1/2}), \bar{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_0 = o_p(n^{-1/2}),$$
$$\bar{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0 = \bar{\boldsymbol{\psi}}_0 - \hat{\boldsymbol{\psi}}_0 + \hat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0 = o_p(n^{-1/2}) + O_p(n^{-1/2}) = O_p(n^{-1/2}),$$
$$\bar{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0 = O_p(n^{-1/2}).$$

Using the first-order Taylor expansion, we have

$$s(\hat{\boldsymbol{\vartheta}}_0) = s(\boldsymbol{\vartheta}_0) + \mathbf{I}(\tilde{\boldsymbol{\vartheta}}_0)(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0),$$

where $\tilde{\boldsymbol{\theta}}_0$ lies on the segment between $\hat{\boldsymbol{\vartheta}}_0$ and $\boldsymbol{\vartheta}_0$. Since $\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0 = O_p(n^{-1/2})$, it means that $\hat{\boldsymbol{\vartheta}}_0$ is a consistent estimator of $\boldsymbol{\vartheta}_0$ so that $\tilde{\boldsymbol{\vartheta}}_0$ is also a consistent estimator of $\boldsymbol{\vartheta}_0$. Hence, we get

$$\begin{aligned}
s(\hat{\boldsymbol{\vartheta}}_0) &= s(\boldsymbol{\vartheta}_0) + \mathbf{I}(\tilde{\boldsymbol{\vartheta}}_0)(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0) \\
&= s(\boldsymbol{\vartheta}_0) + [\mathbf{I}(\boldsymbol{\vartheta}_0) + o_p(n)](\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0) \\
&= s(\boldsymbol{\vartheta}_0) + \mathbf{I}(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0) + o_p(n)(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0) \\
&= s(\boldsymbol{\vartheta}_0) + \mathbf{I}(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0) + o_p(n)O_p(n^{-1/2}) \\
&= s(\boldsymbol{\vartheta}_0) + \mathbf{I}(\boldsymbol{\vartheta}_0)(\hat{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0) + o_p(n^{1/2}) \\
&= O_p(n^{1/2}) + O_p(n)O_p(n^{-1/2}) + o_p(n^{1/2}) = O_p(n^{1/2}),
\end{aligned}$$
$$C(\hat{\boldsymbol{\vartheta}}_0) = s(\hat{\boldsymbol{\vartheta}}_0)s(\hat{\boldsymbol{\vartheta}}_0)' = O_p(n^{1/2})O_p(n^{1/2}) = O_p(n).$$

Similarly, since $\bar{\boldsymbol{\vartheta}}_0 - \boldsymbol{\vartheta}_0 = O_p(n^{-1/2})$, it means that $\bar{\boldsymbol{\vartheta}}_0$ is a consistent estimator of $\boldsymbol{\vartheta}_0$ so that $\tilde{\boldsymbol{\vartheta}}_0$ is also a consistent estimator of $\boldsymbol{\vartheta}_0$. Hence, we can get

$$s(\bar{\boldsymbol{\vartheta}}_0) = O_p(n^{1/2}),$$
$$C(\bar{\boldsymbol{\vartheta}}_0) = s(\bar{\boldsymbol{\vartheta}}_0)s(\bar{\boldsymbol{\vartheta}}_0)' = O_p(n).$$

Furthermore, we can show that

$$s(\bar{\boldsymbol{\vartheta}}_0) = s(\hat{\boldsymbol{\vartheta}}_0) + \mathbf{I}(\tilde{\boldsymbol{\vartheta}}_0)(\bar{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_0),$$

where $\tilde{\boldsymbol{\vartheta}}_0$ lies on the segment between $\bar{\boldsymbol{\vartheta}}_0$ and $\hat{\boldsymbol{\vartheta}}_0$. Because both $\hat{\boldsymbol{\vartheta}}_0$ and $\bar{\boldsymbol{\vartheta}}_0$ are consistent estimators of $\boldsymbol{\vartheta}_0$, $\tilde{\boldsymbol{\vartheta}}_0$ is also a consistent estimator of $\boldsymbol{\vartheta}_0$. Using Lemma 3.1, we get

$$\begin{aligned}
&C(\bar{\boldsymbol{\vartheta}}_0) = s(\bar{\boldsymbol{\vartheta}}_0)s(\bar{\boldsymbol{\vartheta}}_0)' = [s(\hat{\boldsymbol{\vartheta}}_0) + \mathbf{I}(\tilde{\boldsymbol{\vartheta}}_0)(\bar{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_0)][s(\hat{\boldsymbol{\vartheta}}_0) + \mathbf{I}(\tilde{\boldsymbol{\vartheta}}_0)(\bar{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_0)]' \\
&= s(\hat{\boldsymbol{\vartheta}}_0)s(\hat{\boldsymbol{\vartheta}}_0)' + 2\mathbf{I}(\tilde{\boldsymbol{\vartheta}}_0)(\bar{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_0)s(\hat{\boldsymbol{\vartheta}}_0) + \mathbf{I}(\tilde{\boldsymbol{\vartheta}}_0)(\bar{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_0)(\bar{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_0)'\mathbf{I}(\tilde{\boldsymbol{\vartheta}}_0) \\
&= s(\hat{\boldsymbol{\vartheta}}_0)s(\hat{\boldsymbol{\vartheta}}_0)' + 2O_p(n)o_p(n^{-1/2})O_p(n^{1/2}) + O_p(n)o_p(n^{-1/2})o_p(n^{-1/2})O_p(n) \\
&= s(\hat{\boldsymbol{\vartheta}}_0)s(\hat{\boldsymbol{\vartheta}}_0)' + o_p(n) = C(\hat{\boldsymbol{\vartheta}}_0) + o_p(n).
\end{aligned}$$

6.3 Appendix 3: Proof of Theorem 3.1

Using the Bayesian large sample theory, we have

$$\begin{aligned}
& E\left[(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})'|\mathbf{y}\right] = E\left[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}} + \hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}} + \hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})'|\mathbf{y}\right] \\
= \;& E\left[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})'|\mathbf{y}\right] + 2E\left[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})|\mathbf{y}\right](\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}}) + (\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})(\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})' \\
= \;& E\left[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})'|\mathbf{y}\right] - 2(\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})(\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}}) + (\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_{\mathbf{0}})' \\
= \;& E\left[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})'|\mathbf{y}\right] - (\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}})(\hat{\boldsymbol{\vartheta}} - \bar{\boldsymbol{\vartheta}}) \\
= \;& -L_n^{-(2)}(\hat{\boldsymbol{\vartheta}}) + o_p(n^{-1}) + o_p(n^{-1/2})o_p(n^{-1/2}).
\end{aligned}$$

The last equality $E\left[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})'|\mathbf{y}\right] = -L_n^{-(2)}(\hat{\boldsymbol{\vartheta}}) + o_p(n^{-1})$ follows Li, Zeng and Yu (2012) based on the assumptions listed in Section 3.2. Hence, we have

$$\begin{aligned}
& \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = \int (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
= \;& \mathbf{tr}\left[C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0) E[(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})'|\mathbf{y}]\right] \\
= \;& \mathbf{tr}\left[C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}}) + o_p(n^{-1})]\right] \\
= \;& \mathbf{tr}\left[\left(C_{\theta\theta}(\hat{\boldsymbol{\vartheta}}_0) + o_p(n)\right)[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})]\right] + \mathbf{tr}\left[C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0) o_p(n^{-1})\right] \\
= \;& \mathbf{tr}\left[C_{\theta\theta}(\hat{\boldsymbol{\vartheta}}_0)[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})]\right] + o_p(n)[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})] + O_p(n)o_p(n^{-1}) \\
= \;& \mathbf{tr}\left[s_\theta(\hat{\boldsymbol{\vartheta}}_0)s_\theta(\hat{\boldsymbol{\vartheta}}_0)'[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})]\right] + o_p(n)O_p(n^{-1}) + o_p(1) \\
= \;& \mathbf{tr}\left[s_\theta(\hat{\boldsymbol{\vartheta}}_0)s_\theta(\hat{\boldsymbol{\vartheta}}_0)'[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})]\right] + o_p(1) \\
= \;& s_\theta(\hat{\boldsymbol{\vartheta}}_0)'[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})]s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1).
\end{aligned}$$

This proves Equation (7) in the theorem.

When the likelihood information dominates the prior information, the posterior mode $\hat{\boldsymbol{\vartheta}}$ reduces to the ML estimator of $\boldsymbol{\vartheta}$ under the alternative hypothesis, the posterior mode $\hat{\boldsymbol{\psi}}_0$ to the ML estimator of $\boldsymbol{\psi}$ under the null hypothesis, and $L_n^{(2)}(\boldsymbol{\vartheta})$ to $\mathbf{I}(\boldsymbol{\vartheta})$. under the null hypothesis, let $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\psi}_0)$ be the true value of $\boldsymbol{\vartheta}$, and $\hat{\boldsymbol{\vartheta}}_0 = (\boldsymbol{\theta}_0, \hat{\boldsymbol{\psi}})$ be the ML estimator of $\boldsymbol{\vartheta}$. Then, when the null hypothesis is true, $\hat{\boldsymbol{\vartheta}}$ and $\hat{\boldsymbol{\vartheta}}_0$ are both consistent estimators of $\boldsymbol{\vartheta}$. Hence, based on Lemma 3.1 and Lemma 3.2, we get

$$\begin{aligned}
\mathbf{J}(\hat{\boldsymbol{\vartheta}}) &= \mathbf{I}^{-1}(\hat{\boldsymbol{\vartheta}}) = [\mathbf{I}(\boldsymbol{\vartheta}_0) + o_p(n)]^{-1} + o_p(n^{-1}) = \mathbf{J}(\boldsymbol{\vartheta}_\mathbf{0}) + o_p(n^{-1}), \\
\mathbf{J}(\hat{\boldsymbol{\vartheta}}_0) &= \mathbf{I}^{-1}(\boldsymbol{\vartheta}_\mathbf{0}) = [\mathbf{I}(\boldsymbol{\vartheta}_0) + o_p(n)]^{-1} + o_p(n^{-1}) = \mathbf{J}(\boldsymbol{\vartheta}_\mathbf{0}) + o_p(n^{-1}).
\end{aligned}$$

Then, we can further derive that

$$\begin{aligned}
&\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = \int (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} \\
=\ & s_\theta(\hat{\boldsymbol{\vartheta}}_0)'[-L_{n,\theta\theta}^{-(2)}(\hat{\boldsymbol{\vartheta}})] s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}}) s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}}) s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \left[\mathbf{J}_{\theta\theta}(\boldsymbol{\vartheta}_0) + o_p(n^{-1})\right] s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \left[\mathbf{J}_{\theta\theta}(\boldsymbol{\vartheta}_0)\right] s_\theta(\hat{\boldsymbol{\vartheta}}_0) + s_\theta(\hat{\boldsymbol{\vartheta}}_0)' o_p(n^{-1}) s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \left[\mathbf{J}_{\theta\theta}(\boldsymbol{\vartheta}_0)\right] s_\theta(\hat{\boldsymbol{\vartheta}}_0) + O_p(n^{1/2}) o_p(n^{-1}) O_p(n^{1/2}) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \left[\mathbf{J}_{\theta\theta}(\boldsymbol{\vartheta}_0)\right] s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \left[\mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}}_0) + o_p(n^{1/2})\right] s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}}_0) s_\theta(\hat{\boldsymbol{\vartheta}}_0) + O_p(n^{1/2}) o_p(n^{-1}) O_p(n^{1/2}) + o_p(1) \\
=\ & -s_\theta(\hat{\boldsymbol{\vartheta}}_0)' \mathbf{J}_{\theta\theta}(\hat{\boldsymbol{\vartheta}}_0) s_\theta(\hat{\boldsymbol{\vartheta}}_0) + o_p(1) \\
=\ & LM + o_p(1).
\end{aligned}$$

According to the standard ML theory, under the null hypothesis, $LM \xrightarrow{d} \chi^2(p)$. Therefore, $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(p)$ and the theorem is proved.

6.4 Appendix 4: Derivation of BF and the proposed test statistic in a simple linear model

In the simple linear regression model, under the null hypothesis, the marginal likelihood is

$$
\begin{aligned}
p_0(\mathbf{y}) &= \int\int p(\mathbf{y}|\alpha,\beta_0)\,p\left(\alpha|\sigma^2\right)p\left(\sigma^2\right)d\alpha d\sigma^2 \\
&= \frac{b^a}{(2\pi)^{\frac{n}{2}}\Gamma(a)}\int\int \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta_0x_i)^2\right)\frac{1}{\sqrt{2\pi V_\alpha}\sigma}\exp\left(-\frac{(\alpha-\mu_\alpha)^2}{2\sigma^2V_\alpha}\right) \\
&\quad\times\left(\sigma^2\right)^{-a-\frac{n}{2}-1}\exp\left(-\frac{b}{\sigma^2}\right)d\alpha d\sigma^2 \\
&= \frac{b^a}{(2\pi)^{\frac{n}{2}}\Gamma(a)}\int\int\frac{1}{\sqrt{2\pi V_\alpha}\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[-2\alpha\sum_{i=1}^{n}(y_i-\beta_0x_i)+n\alpha^2\right]\right\} \\
&\quad\times\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0x_i)^2\right)\exp\left[-\frac{1}{2\sigma^2V_a}\left(\alpha^2-2\mu_\alpha\alpha\right)\right]\exp\left(-\frac{\mu_\alpha^2}{2\sigma^2V_\alpha}\right)d\alpha d\sigma^2 \\
&= \frac{b^a}{(2\pi)^{\frac{n}{2}}\Gamma(a)}\int\int\frac{1}{\sqrt{2\pi V_\alpha}\sigma}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0x_i)^2\right)\exp\left(-\frac{\mu_\alpha^2}{2\sigma^2V_\alpha}\right) \\
&\quad\times\exp\left\{-\frac{1}{2\sigma^2}\left[\left(n+\frac{1}{V_\alpha}\right)\alpha^2-2\alpha\left(\sum_{i=1}^{n}(y_i-\beta_0x_i)+\frac{\mu_\alpha}{V_\alpha}\right)\right]\right\}d\alpha d\sigma^2 \\
&= \frac{b^a}{(2\pi)^{\frac{n}{2}}\Gamma(a)}\sqrt{\frac{1}{nV_\alpha+1}}, \\
&\quad\times\int_0^{+\infty}\left(\sigma^2\right)^{-a-\frac{n}{2}-1}\exp\left\{-\frac{1}{\sigma^2}\left[b+\frac{1}{2}\left(\sum_{i=1}^{n}(y_i-\beta_0x_i)^2+\frac{\mu_\alpha^2}{V_\alpha}-\frac{\mu_\alpha^{*2}}{V_\alpha^*}\right)\right]\right\}d\sigma^2 \\
&= \frac{b^a\Gamma\left(a+\frac{n}{2}\right)}{(2\pi)^{\frac{n}{2}}\Gamma(a)}\sqrt{\frac{1}{nV_\alpha+1}}\left[b+\frac{1}{2}\left(\sum_{i=1}^{n}(y_i-\beta_0x_i)^2+\frac{\mu_\alpha^2}{V_\alpha}-\frac{\mu_\alpha^*}{V_\alpha^*}\right)\right]^{-\left(a+\frac{n}{2}\right)} \\
&= \frac{b^a\Gamma\left(a+\frac{n}{2}\right)}{(2\pi)^{\frac{n}{2}}\Gamma(a)}\sqrt{\frac{1}{nV_\alpha+1}}\left[b+\frac{1}{2}\left((\mathbf{y}-\beta_0\mathbf{x})'(\mathbf{y}-\beta_0\mathbf{x})+\frac{\mu_\alpha^2}{V_\alpha}-\frac{\mu_\alpha^*}{V_\alpha^*}\right)\right]^{-\left(a+\frac{n}{2}\right)},
\end{aligned}
$$

where $V_\alpha^* = \frac{V_\alpha}{nV_\alpha+1}$, $\mu_\alpha^* = V_\alpha^*\left(\sum_{i=1}^n(y_i-\beta_0x_i)+\frac{\mu_\alpha}{V_\alpha}\right) = V_\alpha^*\left(\iota'\left(\mathbf{y}-\beta_0\mathbf{x}+\frac{\mu_\alpha}{V_\alpha}\right)\right)$ with $\iota=(1,1,...,1)'$. Under H_1, we rewrite the equation in a matrix form:

$$
\mathbf{y} = X\gamma+\varepsilon,
$$

where $\gamma = (\alpha,\beta)'$, $X = (\iota,\mathbf{x})$. The prior for γ is $N\left(\tilde{\mu},\sigma^2\tilde{V}\right)$, where $\tilde{\mu} = (\mu_\alpha,\mu_\beta)'$, $\tilde{V} = diag(V_\alpha,V_\beta)$.

Similarly, the marginal likelihood for the model under the alternative is:

$$\begin{aligned}
p_1(\mathbf{y}) &= \int\int p(\mathbf{y}|\beta,\alpha)\, p\left(\gamma|\sigma^2\right) p\left(\sigma^2\right) d\gamma d\sigma^2 \\
&= \frac{b^a}{(2\pi)^{\frac{n}{2}}\Gamma(a)} \int\int \left(\sigma^2\right)^{-a-\frac{n}{2}-1} \exp\left(-\frac{b}{\sigma^2}\right) \\
&\quad \times \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-X\gamma)'(\mathbf{y}-X\gamma)\right) \frac{1}{2\pi|\tilde{V}|^{\frac{1}{2}}\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(\gamma-\tilde{\mu})'\,\tilde{V}^{-1}(\gamma-\tilde{\mu})\right) d\gamma d\sigma^2 \\
&= \frac{b^a}{(2\pi)^{\frac{n}{2}}\Gamma(a)\sqrt{|\tilde{V}|}} \int\int \frac{1}{2\pi\sigma^2}\left(\sigma^2\right)^{-a-\frac{n}{2}-1} \left\{\left(-\frac{1}{\sigma^2}\left[b+\frac{1}{2}\left(\mathbf{y}'\mathbf{y}+(\tilde{\mu})'\tilde{V}^{-1}\tilde{\mu}\right)\right]\right)\right\} \\
&\quad \times \exp\left\{-\frac{1}{2\sigma^2}\left(\gamma'\left(X'X+\tilde{V}^{-1}\right)\gamma-\gamma'\left(X'\mathbf{y}+\tilde{V}^{-1}\tilde{\mu}\right)-\left(X'\mathbf{y}+\tilde{V}^{-1}\tilde{\mu}\right)'\gamma\right)\right\} d\gamma d\sigma^2 \\
&= \frac{b^a}{(2\pi)^{\frac{n}{2}}\Gamma(a)\sqrt{|\tilde{V}|}} \int\int \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(\gamma-\mu^*)'\,V^{*-1}(\gamma-\mu^*)\right\} \\
&\quad \times \exp\left(-\frac{1}{2\sigma^2}\left((\tilde{\mu})'\tilde{V}^{-1}\tilde{\mu}+\mathbf{y}'\mathbf{y}-(\mu^*)'V^{*-1}\mu^*\right)\right)\left(\sigma^2\right)^{-a-\frac{n}{2}-1} \exp\left(-\frac{b}{\sigma^2}\right) d\gamma d\sigma^2 \\
&= \frac{b^a\sqrt{|V^*|}}{(2\pi)^{\frac{n}{2}}\Gamma(a)\sqrt{|\tilde{V}|}} \int \left(\sigma^2\right)^{-a-\frac{n}{2}-1} \exp\left\{-\frac{1}{\sigma^2}\left[b+\frac{1}{2}\left((\tilde{\mu})'\tilde{V}^{-1}\tilde{\mu}+\mathbf{y}'\mathbf{y}-(\mu^*)'V^{*-1}\mu^*\right)\right]\right\} d\sigma^2 \\
&= \frac{b^a\sqrt{|V^*|}\Gamma\left(a+\frac{n}{2}\right)}{(2\pi)^{\frac{n}{2}}\Gamma(a)\sqrt{|\tilde{V}|}} \left[b+\frac{1}{2}\left((\tilde{\mu})'\tilde{V}^{-1}\tilde{\mu}+\mathbf{y}'\mathbf{y}-(\mu^*)'V^{*-1}\mu^*\right)\right]^{-\left(a+\frac{n}{2}\right)},
\end{aligned}$$

where $V^* = \left(\tilde{V}^{-1}+X'X\right)^{-1}$, $\mu^* = V^*\left(\tilde{V}^{-1}\tilde{\mu}+X'\mathbf{y}\right)$.

Note that the log-likelihood function is:

$$\log p(\mathbf{y}|\boldsymbol{\vartheta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2.$$

Hence, given $\boldsymbol{\vartheta} = \left(\alpha,\beta,\sigma^2\right)'$, $\boldsymbol{\theta} = \beta$, we have

$$s(\boldsymbol{\vartheta}) = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i), \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i(y_i-\alpha-\beta x_i), -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2\right)',$$

and

$$C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right) = \frac{1}{\overline{\sigma_0^2}}\left[\sum_{i=1}^{n}x_i\left(y_i-\overline{\alpha}_0-\beta_0 x_i\right)\right]^2 = \frac{1}{\overline{\sigma_0^2}}\left[\mathbf{x}'\left(\mathbf{y}-\overline{\alpha}_0\iota-\beta_0\mathbf{x}\right)\right]^2,$$

where $\overline{\alpha}_0$ and $\overline{\sigma_0^2}$ are the posterior means of α and σ^2 under H_0.

To sum up, to compute the $\boldsymbol{T}(\mathbf{y},\boldsymbol{\theta}_0)$, we first draw MCMC samples for the model corresponding to H_0 and calculate $C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right)$. And then we draw MCMC samples for the model corresponding to H_1 to obtain $\left\{\boldsymbol{\vartheta}^{(j)}\right\}_{j=1}^{M}=\left\{\alpha^{(j)},\beta^{(j)},\sigma^{(j)}\right\}_{j=1}^{M}$. The approximation of the test statistic is given by

$$\hat{\boldsymbol{T}}\left(\mathbf{y},\boldsymbol{\theta}_0\right)=\frac{1}{M}C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right)\sum_{j=1}^{M}\left(\beta^{(j)}-\overline{\beta}\right)^2,$$

where $\overline{\beta}=\frac{1}{M}\sum_{j=1}^{M}\beta^{(j)}$.

6.5 Appendix 5: Derivation of the proposed test statistic in a binary probit model

In the binary probit model, for each y_i, $i=1,2,...,n$, there is a corresponding latent variable z_i that satisfies:

$$\begin{cases} y_i=1 & \text{if } \ z_i\geq 0 \\ y_i=0 & \text{if } \ z_i<0 \end{cases},$$

and

$$z_i=\mathbf{x}_i'\boldsymbol{\vartheta}+e_i,$$

where $\boldsymbol{\vartheta}$ is the $(p+q)\times 1$ parameter vector measuring the marginal effects and $e_i\sim N\left(0,1\right)$ for $i=1,...,n$.

Rewrite the above equation as:

$$z_i=\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}+e_i.$$

For each i, we have

$$\begin{cases} p\left(y_i=1|\boldsymbol{\vartheta}\right)=p\left(z_i\geq 0|\boldsymbol{\vartheta}\right)=p\left(e_i\geq -\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)|\boldsymbol{\vartheta}\right)=\Phi\left[\left(2y_i-1\right)\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right] \\ p\left(y_i=0|\boldsymbol{\vartheta}\right)=p\left(z_i<0|\boldsymbol{\vartheta}\right)=p\left(e_i<-\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)|\boldsymbol{\vartheta}\right)=\Phi\left[\left(2y_i-1\right)\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right] \end{cases},$$

where the $\Phi\left(\cdot\right)$ is the standard normal cumulative distribution function. Note that the log-likelihood function is:

$$\log p\left(\mathbf{y}|\boldsymbol{\vartheta}\right)=\sum_{i=1}^{n}\log\Phi\left[q_i\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right],$$

where $q_i=2y_i-1$.

In our example, the null hypothesis is: $H_0:\boldsymbol{\theta}=\mathbf{0}$. Note that,

$$\frac{\partial\log p\left(\mathbf{y}|\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\theta}} = \sum_{i=1}^{n}q_i\frac{\phi\left[q_i\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right]\mathbf{x}_{i2}}{\Phi\left[q_i\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right]},$$

where $\phi(\cdot)$ is the density function of the standard normal distribution. The test statistic is

$$\boldsymbol{T}(\mathbf{y},\boldsymbol{\theta_0}) = \int \left(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}\right)' C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right)\left(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}\right) p\left(\boldsymbol{\vartheta}|\mathbf{y}\right) d\boldsymbol{\vartheta}.$$

where

$$\begin{aligned} C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right) &= \left.\left(\frac{\partial \log p\left(\mathbf{y}|\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log p\left(\mathbf{y}|\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\theta}}\right)'\right|_{\boldsymbol{\vartheta}=\overline{\boldsymbol{\vartheta}}_0} \\ &= \left(\sum_{i=1}^{n}\frac{\phi\left[q_i\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right]q_i\mathbf{x}_{i2}}{\Phi\left[q_i\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right]}\right)\times\left(\sum_{i=1}^{n}\frac{\phi\left[q_i\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right]q_i\mathbf{x}_{i2}}{\Phi\left[q_i\left(\mathbf{x}_{i1}'\boldsymbol{\psi}+\mathbf{x}_{i2}'\boldsymbol{\theta}\right)\right]}\right)', \end{aligned}$$

where $\overline{\boldsymbol{\vartheta}}_0 = \left(\boldsymbol{\theta}_0, \overline{\boldsymbol{\psi}}_0\right)$ and $\overline{\boldsymbol{\psi}}_0$ is the posterior mean of $\boldsymbol{\psi}$ under H_0.

To sum up, to compute the $\boldsymbol{T}(\mathbf{y},\boldsymbol{\theta}_0)$, we firstly draw MCMC samples for the model under H_0 and calculate $C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right)$. We then draw MCMC samples for the model under H_1 to obtain $\left\{\boldsymbol{\vartheta}^{(j)}\right\}_{j=1}^{M} = \left\{\boldsymbol{\theta}^{(j)},\boldsymbol{\psi}^{(j)}\right\}_{j=1}^{M}$. The approximation of the statistic is

$$\hat{\boldsymbol{T}}(\mathbf{y},\boldsymbol{\theta_0}) = \frac{1}{M}\sum_{j=1}^{M}\left(\boldsymbol{\theta}^{(j)}-\overline{\boldsymbol{\theta}}\right)' C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right)\left(\boldsymbol{\theta}^{(j)}-\overline{\boldsymbol{\theta}}\right), \tag{13}$$

where $\overline{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$ for the model under H_1.

6.6 Appendix 6: Derivation of the test statistic in the stochastic conditional duration model

In the third example, we choose the SCD model, which is defined as:

$$\begin{cases} d_t = \exp\left(\varphi_t\right)\varepsilon_t & \varepsilon_t \sim \text{Exp}(1) \\ \varphi_t = \mu + \phi\left(\varphi_{t-1}-\mu\right)+\sigma\epsilon_t & \epsilon_t \sim N(0,1) \\ \varphi_1 \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right) & \end{cases},$$

for $t = 1,...,T$, where d_t is the adjusted duration data, $|\phi| < 1$, φ_t is the latent variable. ε_t and ϵ_t are assumed to be independent. We would like to test if $\phi = 0$. Hence, $\boldsymbol{\theta} = \phi$, $\boldsymbol{\psi} = \left(\mu, \sigma^2\right)$, $\boldsymbol{\vartheta} = (\boldsymbol{\theta},\boldsymbol{\psi})$. The proposed test statistic is:

$$\boldsymbol{T}(\mathbf{d},\boldsymbol{\theta}_0) = \int \left(\phi-\overline{\phi}\right) C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right)\left(\phi-\overline{\phi}\right) p\left(\phi|\mathbf{d}\right) d\phi.$$

where $\mathbf{d} = \{d_t\}_{t=1}^{T}$, $\overline{\phi}$ is the posterior mean of ϕ under H_1, and $C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right)$ can be approximated by

$$\begin{aligned} C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right) &= \left.\left[\frac{\partial \log p\left(\boldsymbol{d}|\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\vartheta}}\left(\frac{\partial \log p\left(\boldsymbol{d}|\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\vartheta}}\right)'\right]_{\theta\theta}\right|_{\boldsymbol{\vartheta}=\overline{\boldsymbol{\vartheta}}_0} \\ &= \left.\left[\frac{\partial \log p\left(\boldsymbol{d}|\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\theta}}\left(\frac{\partial \log p\left(\boldsymbol{d}|\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\theta}}\right)'\right]\right|_{\boldsymbol{\vartheta}=\overline{\boldsymbol{\vartheta}}_0} \end{aligned}$$

According to Remark 3.4, the partial derivative of log-likelihood function with respect to ϕ can be approximated by the $\mathcal{Q}$-function. Under H_1, the log-likelihood function given $\boldsymbol{\vartheta}$ and $\boldsymbol{\varphi} = \{\varphi_t\}_{t=1}^{T}$ is

$$\begin{aligned}\log p\left(\boldsymbol{d}|\boldsymbol{\varphi},\boldsymbol{\vartheta}\right) &= -\sum_{t=2}^{T}\frac{1}{2\sigma^2}\left(\varphi_t-\mu-\phi\left(\varphi_{t-1}-\mu\right)\right)^2-\frac{1-\phi^2}{2\sigma^2}\left(\varphi_1-\mu\right)^2\\ &\quad +\frac{1}{2}\log\left(1-\phi^2\right)-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma).\end{aligned}$$

Therefore, the partial derivative of log-likelihood function with respect to ϕ given $\boldsymbol{\varphi}$ and $\boldsymbol{\vartheta}$ is

$$\frac{\partial \log p\left(\boldsymbol{d}|\boldsymbol{\varphi},\boldsymbol{\vartheta}\right)}{\partial \boldsymbol{\theta}} = \sum_{t=2}^{T}\frac{1}{\sigma^2}\left(\varphi_t-\mu-\phi\left(\varphi_{t-1}-\mu\right)\right)\left(\varphi_{t-1}-\mu\right)+\frac{\phi}{\sigma^2}\left(\varphi_1-\mu\right)^2-\frac{\phi}{1-\phi^2}.$$

Then, under H_1, we generate MCMC samples and denote them by $\left\{\mu_1^{(j)},\phi^{(j)},\sigma_1^{2(j)}\right\}_{j=1}^{M}$. Under H_0, the MCMC samples is denoted as $\left\{\mu_0^{(j)},\sigma_0^{2(j)},\boldsymbol{\varphi}^{(j)}\right\}_{j=1}^{M}$, where $\boldsymbol{\varphi}^{(j)}=\left\{\varphi_t^{(j)}\right\}_{t=1}^{T}$ is the set of the draws of the latent variables at j-th iteration. Then $\partial \log p\left(\boldsymbol{d}|\boldsymbol{\vartheta}\right)/\partial\boldsymbol{\theta}$ can be approximated by

$$\left.\frac{\partial \log p\left(\boldsymbol{d}|\boldsymbol{\vartheta}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\vartheta}=\overline{\boldsymbol{\vartheta}}_0} \approx \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{\overline{\sigma_0^2}}\sum_{t=2}^{T}\left(\varphi_t^{(j)}-\overline{\mu}_0\right)\left(\varphi_{t-1}^{(j)}-\overline{\mu}_0\right)\right),$$

where $\overline{\mu}_0=\frac{1}{M}\sum_{j=1}^{M}\mu_0^{(j)}$, $\overline{\sigma_0^2}=\frac{1}{M}\sum_{j=1}^{M}\sigma_0^{2(j)}$, $\overline{\boldsymbol{\vartheta}}_0=\left(\boldsymbol{\theta}_0,\overline{\boldsymbol{\psi}}_0\right)$, $\boldsymbol{\theta}_0=0$, $\overline{\boldsymbol{\psi}}_0=\left(\overline{\mu}_0,\overline{\sigma_0^2}\right)$. The test statistic can be approximated by

$$\hat{\boldsymbol{T}}\left(\mathbf{d},\boldsymbol{\theta}_0\right)=\frac{1}{M}C_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}_0\right)\sum_{j=1}^{M}\left(\phi^{(j)}-\overline{\phi}\right)^2.$$

References

An, S. and Schorfheide, F. (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews*, **26**, 211-219.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent variable model for the analysis of financial durations.*Journal of econometrics*, **119**, 381-412.

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

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

Breusch, T.S. and Pagan, A.R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, **47(1)**, 239-253.

Brooks, C. (2008). *Introductory Econometrics for Finance.* Cambridge university press.

Chen, C.F. (1985). On asymptotic normality of limiting density function with Bayesian implications. *Journal of the Royal Statistical Society, Ser. B*, **47(3)**, 540-546.

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

Engle, R. F., and Russell, J.R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127-1162.

Geweke, J. (2005), *Contemporary Bayesian Econometrics and Statistics*, Wiley.

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

Gilks, Wally R., N. G. Best, and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* , **44**, 455-472.

Han, C. and Carlin, B.P. (2001). Markov chain Monte Carlo methods for computing Bayes factor: a comparative review. *Journal of the American Statistical Association*, **96(455)**, 1122-1132.

Imai,S., JAIN, N., and Ching A. (2009). Bayesian Estimation of Dynamic Discrete Choice Models, *Econometrica*, **77**, 1865-1899.

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

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

Kim, J. (1994), Bayesian asymptotic theory in a time series model with a possible non-stationary process. *Econometric Theory*, 10(3-4): 764-773.

Kim, J. Y. (1998)., Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models. *Econometrica*, 359-380.

Koop, G. (2003). *Bayesian Econometrics.* Wiley.

Li, Y. and Yu, J. (2012). Bayesian hypothesis testing in latent variable models. *Journal of Econometrics*, 166(2), 237-246.

Li, Y., Zeng, T. and Yu, J. (2012). Robust deviation information criterion for latent variable models. *Working paper*, Singapore Management University.

Li, Y., Zeng, T. and Yu, J. (2014). A new approach to Bayesian hypothesis testing. *Journal of Econometrics*, **178(3)**, 602-612.

McCulloch, R.E. (1989). Local model influence. *Journal of the American Statistical Association*, **84**, 473-478.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55(4), 765-799.

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

Poirier, D.J. (1995). Intermediate Statistics and Econometrics: A Comparative Approach. MIT Press, Cambridge, MA.

Poirier, D.J. (1997). A predictive motivation for loss function specification in parametric hypothesis testing. *Economic letters*, **56**, 1-3.

Robert, C. (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 Texts in Statistics, 2nd version.

Shephard, N., (1995). Stochastic volatility: Selective readings. Oxford University Press.

Tanner, T.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528-540.

Wooldridge, J.M. (2002). *Introductory Econometrics: A Modern Approach.* Cengage Learning, 3 edition.