Hypothesis Testing Statistics Based on Posterior Output with Applications in Financial Econometrics

YONG LI *Renmin University of China*

This chapter reviews three recently developed posterior test statistics for hypothesis testing based on posterior output. These three statistics can be viewed as the posterior version of the "trinity" of test statistics based on maximum likelihood (ML), the likelihood ratio (LR) test, the Lagrange multiplier (LM) test, and the Wald test. The asymptotic distributions of the test statistics are discussed under repeated sampling. Also, based on the Bernstein-von Mises theorem, the equivalence of the confidence interval construction between the set of posterior tests and their frequentist counterparts is developed, giving the posterior tests a frequentist asymptotic justification. The three statistics are applicable to many popular financial econometric models, including asset pricing models, copula models, etc. Studies based on simulated data and real data in the context of several financial econometric models are performed to demonstrate the finite sample behavior and usefulness of the test statistics.

10.1 Introduction

In economics and finance, hypothesis testing is a primary concern in statistical inference. For example, an economic theory often corresponds to a testable hypothesis in empirical analysis. Thus, testing a point null hypothesis is important when checking statistical evidence from data to support or to be against a particular economic theory. The central question we ask in this chapter is how to test a point null hypothesis when posterior output is available.

Broadly speaking, there are three posterior-based methods available in the Bayesian literature for hypothesis testing. The first one is the well-known Bayes factor (BF) that compares the marginal likelihoods of the two competing models corresponding to the null and alternative hypotheses (Kass and Raftery, 1995). Unfortunately, BFs are subject to a few theoretical and practical problems. First, BFs are not well-defined under improper priors. Second, BFs are subject to Jeffreys-Lindley-Bartlett's paradox; they tend to choose the null hypothesis when a very vague prior is used for parameters in the null hypothesis (see Kass and Raftery (1995), Poirier (1995), Chapter 4 in Wakefield (2013)). Third, in many cases, the evaluation of marginal likelihood is difficult. Several strategies have been proposed in the literature to address some of these difficulties. For example, to solve the first two problems, one may use a prior that is data-dependent when calculating BFs. To make the prior data-dependent, one may split the data into two parts: one as a training set and the other for statistical analysis. The training data can be used to update a prior, which can be improper, to generate a proper prior to analyze the remaining data (see the fractional BF of O'Hagan (1995) and the intrinsic BF of Berger (1985)). To address the computational problem of BFs,the methods of Friel and Pettitt (2008), Li et al. (2021), and Chib (1995) can be used.

The second posterior-based method uses credible intervals or sets. This line of research has drawn a lot of attention among econometricians and statisticians in recent years (see Chernozhukov and Hong (2003), Moon and Schorfheide (2012), Kline and Tamer (2016), Liao and Simoni (2019) and Chen et al. (2018)).

The third method is based on statistical decision theory. The idea begins with Bernardo and Rueda (2002, BR), where it is demonstrated that the BF can be regarded as a decision problem with a simple zero-one loss function when it is used for point hypothesis testing. It is this zero-one loss that leads to Jeffreys-Lindley-Bartlett's paradox. BR also suggested using the continuous Kullback-Leibler (KL) divergence function as the loss function to replace the zero-one loss. Subsequent extensions include Li and Yu (2012), Li et al. (2014), Li et al. (2015), Liu et al. (2021) and Li et al. (2022), where alternative loss functions are used.

By focusing on the third line of approaches, the goal of this chapter is to review the literature on hypothesis testing based on posterior output. Posterior output can be achieved via some advanced posterior simulation techniques such as Markov chain Monte Carlo (MCMC) or sequential Monte Carlo (SMC). The posterior test statistics developed for hypothesis testing when posterior output is available can be justified in a frequentist set-up in the same way as testing methods based on the maximum likelihood (ML) estimator are justified.

With posterior output, it is not immediately obvious how to make statistical inference in the frequentist framework. For the point-null hypothesis testing problem, this chapter describes how the posterior test statistics for hypothesis testing are developed under the decisional framework. In this chapter, three posterior test statistics for hypothesis testing that were developed in recent years are reviewed. They can be viewed as the Bayesian version of the "trinity" of test statistics based on ML, namely, the likelihood ratio (LR) test, the Lagrange multiplier (LM) test, and the Wald test. Under repeated sampling, the asymptotic distributions of these posterior test statistics are discussed. Simulated data are also used to examine the finite sample properties of the test statistics and real data to show of their usefulness.

This chapter is organized as follows. Section 2 reviews the Bayesian inference based on posterior output. Inferential approaches that are typically used in the Bayesian literature are also briefly explained. Section 3 reviews several statistics for hypothesis testing based on decision theory, and Section 4 reviews the posterior statistics. Section 5 describes various simulation studies, and Section 6 provides the empirical illustrations. Finally, Section 7 concludes the chapter.

10.2 Bayesian Inference based on Posterior

Without loss of generality, let $\mathbf{y} = (y_1, \ldots, y_n)'$ denote the data generated from a probability measure P_0 on the probability space (Ω, F, P_0), and $p(\mathbf{y}|\boldsymbol{\vartheta})$ is denoted as the likelihood function. To perform Bayesian inference about $\boldsymbol{\vartheta}$, let $p(\boldsymbol{\vartheta})$ be the prior distribution of $\boldsymbol{\vartheta}$. Then, the posterior distribution is obtained via the Bayesian theorem:

$$p(\boldsymbol{\vartheta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})}{p(\mathbf{y})} \propto p(\mathbf{y}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}), \tag{10.2.1}$$

where $p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})d\boldsymbol{\vartheta}$ is the marginal likelihood.

Frequentist inference is based on the likelihood function $p(\mathbf{y}|\boldsymbol{\vartheta})$, and the Bayesian statistical inference is based on the posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$. The posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$, when it is not analytically tractable, can be obtained via some advanced posterior simulation techniques such as MCMC or SMC. Gamerman and Lopes (2006) and Chopin and Papaspiliopoulos (2020) provide details about these simulation techniques.

Samples obtained from the posterior simulation can be used for statistical inference. Bayesian estimates of $\boldsymbol{\vartheta}$ can be obtained easily via sampling means of random samples. Specifically, let $\{\boldsymbol{\vartheta}^{(j)}, j = 1, 2, \cdots, J\}$ be the effective random samples generated from the joint posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$. Then, the Bayesian estimates of $\boldsymbol{\vartheta}$ can be obtained as follows:

$$\bar{\boldsymbol{\vartheta}} = E[\boldsymbol{\vartheta}|\mathbf{y}] = \int \boldsymbol{\vartheta} p(\boldsymbol{\vartheta}|\mathbf{y})\mathrm{d}\boldsymbol{\vartheta} \approx \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\vartheta}^{(j)}. \tag{10.2.2}$$

Clearly, these Bayesian estimates are consistent estimates of the corresponding posterior means (Geyer, 1992). A consistent estimate of $Var(\boldsymbol{\vartheta}|\boldsymbol{y})$ can be described by follows:

$$\widehat{Var(\boldsymbol{\vartheta}|\mathbf{y})} = \frac{1}{J}\sum_{j=1}^{J}(\boldsymbol{\vartheta}^{(j)} - \bar{\boldsymbol{\vartheta}}_J)(\boldsymbol{\vartheta}^{(j)} - \bar{\boldsymbol{\vartheta}}_J)'.$$

where $\bar{\boldsymbol{\vartheta}}_J = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\vartheta}^{(j)}$.

Under some regularity conditions, when $p(\boldsymbol{\vartheta}) = O_p(1)$, Li et al. (2017) showed that the relationship between the posterior mean $\bar{\boldsymbol{\vartheta}}$ and the posterior mode $\hat{\boldsymbol{\vartheta}}$ can be expressed as:

$$\bar{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}} + O_p(n^{-1}), \tag{10.2.3}$$

$$Var(\boldsymbol{\vartheta}|\mathbf{y}) = \left[-\frac{\partial^2 \ln p(\mathbf{y}|\hat{\boldsymbol{\vartheta}})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right]^{-1} + O_p(n^{-2}). \tag{10.2.4}$$

The large sample properties in (10.2.3) and (10.2.4) provide the fountainhead from which all the methods reviewed in this chapter springs.

10.3 Hypothesis Testing based on Posterior Output

10.3.1 Hypothesis testing under decision theory

It is assumed that probability model $M \equiv \{p(\mathbf{y}|\boldsymbol{\vartheta})\}$ is used to fit data $\mathbf{y}$ where $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \boldsymbol{\psi}')' \in \boldsymbol{\Theta}$. We are concerned with testing a point null hypothesis that may arise from the prediction of a particular theory. Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}_\theta$ denote a vector of q_θ-dimensional parameters of interest and $\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi$ a vector of q_ψ-dimensional nuisance parameters so that $\boldsymbol{\Theta} = \boldsymbol{\Theta}_\theta \times \boldsymbol{\Theta}_\psi$. The testing problem is given by:

$$\begin{cases} H_0: & \boldsymbol{\theta} = \boldsymbol{\theta}_0, \\ H_1: & \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \end{cases} \tag{10.3.1}$$

In the statistical decision framework, hypothesis testing may be understood as follows. There are two statistical decisions in the decision space, accepting H_0 (name it d_0) or rejecting H_0 (name it d_1). Let $\{\mathcal{L}(d_i, \boldsymbol{\theta}, \boldsymbol{\psi}), i = 0, 1\}$ be the loss function of the statistical decision associated with d_i. For the decision problem, the loss functions $\{\mathcal{L}(d_i, \boldsymbol{\theta}, \boldsymbol{\psi}), i = 0, 1\}$ can be used to measure the loss of accepting H_0 or rejecting H_0 as a function of the real value of the parameters $(\boldsymbol{\theta}, \boldsymbol{\psi})$.

Given the loss function and data $\mathbf{y}$, the optimal action is to reject H_0, if and only if (iff) the expected posterior loss of accepting H_0 is larger than the expected posterior loss of rejecting H_0:

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

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

$$\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) - \mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}). \tag{10.3.2}$$

This equation measures the evidence against H_0 as a function of $(\boldsymbol{\theta}, \boldsymbol{\psi})$.

For this prespecified net loss difference function $\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})$, we can define a posterior-based statistic as:

$$\mathbf{T}_0(\mathbf{y}, \boldsymbol{\theta}_0) = \int_{\boldsymbol{\Theta}_\theta}\int_{\boldsymbol{\Theta}_\psi} \triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} = E_{\boldsymbol{\vartheta}|\mathbf{y}}\left(\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})\right) \tag{10.3.3}$$

Following Berger (1985), any Bayesian admissible solution to the decision problem must satisfy:

$$\text{Reject } H_0 \text{ iff } \mathbf{T}_0(\mathbf{y}, \boldsymbol{\theta}_0) > 0, \tag{10.3.4}$$

BR shows that the net loss function can generally take the form:

$$\Delta\mathcal{L}\left(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = m\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) - c,$$

where $m\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right)$ is a non-negative discrepancy measure between model $p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi})$ and model $p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$, $c > 0$ is a context-dependent utility value that measures the advantage of being able to work with the simpler model when it is true. For this type of net loss function, a possible loss function for the decision can be given as follows:

$$\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_0 + \frac{1}{2} m\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases} \tag{10.3.5}$$

$$\mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_1 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_1 - \frac{1}{2} m\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases} \tag{10.3.6}$$

where c_i measures the cost of action d_i and $c = c_1 - c_0 > 0$. Then, let $\mathbf{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ be defined as:

$$\mathbf{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right) = E_{\boldsymbol{\vartheta}|\mathbf{y}}\left(m\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right)\right) = E_{\boldsymbol{\vartheta}|\mathbf{y}}\left(\Delta\mathcal{L}\left(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right)\right) - c. \tag{10.3.7}$$

Then, $\mathbf{T}_0\left(\mathbf{y}, \boldsymbol{\theta}_0\right) > 0$ is equivalent to $\mathbf{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right) > c$. In this case, H_0 is rejected.

10.3.2 The choice of loss function for hypothesis testing

In this subsection, we review the loss functions that are used to construct hypothesis test statistics. We show that the BFs correspond to the discrete loss function that equals of 0 and 1. To overcome the shortcomings of BFs, alternative continuous loss functions have been proposed in the literature to construct new test statistics based on the MCMC output. There is a more fundamental difference between these new test statistics and the BFs. The new test statistics are justified in a frequentist setup by assuming that $\mathbf{y}$ comes out of the data generating process in a repeated experiment, while BFs are justified in a Bayesian setup (i.e., the decision is made conditional on $\mathbf{y}$).

BFs and zero-one loss function

If one uses the following zero-one loss functions:

$$\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} 0 & if\ \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ 1 & if\ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \qquad \mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} 1 & if\ \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ 0 & if\ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases},$$

then, the net loss function $\triangle\mathcal{L}(H_0, \boldsymbol{\vartheta}, \boldsymbol{\psi})$ is:

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

Thus, the expected posterior loss is given by:

$$\int \triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})\, p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} = \int \triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})\, p(\boldsymbol{\theta}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}.$$

For point hypothesis testing, a reasonable prior for $\boldsymbol{\theta}$ requires a positive probability w being assigned to H_0. Let the prior $\boldsymbol{\theta}$ be a mixed random variable whose density is:

$$p(\boldsymbol{\theta}) = w\delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + (1-w)\, \pi(\boldsymbol{\theta}),$$

where $\delta(\cdot)$ denotes the Dirac delta function and $\pi(\boldsymbol{\theta})$ is a proper density function. For more details about the mixed random variable and Dirac delta function, see Pishro-Nik (2016).

Thus, the joint posterior density of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ is:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) &= \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{y})} \\ &= \frac{p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta} = \boldsymbol{\theta}_0) w\delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{p(\mathbf{y})} + \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta})(1-w)\pi(\boldsymbol{\theta})}{p(\mathbf{y})}, \end{aligned}$$

where $p(y) = \int\int p(y|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi}$ is the marginal likelihood that can be expressed as:

$$p(\mathbf{y}) = w \int p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta} = \boldsymbol{\theta}_0) d\boldsymbol{\psi} + (1-w) \int \pi(\boldsymbol{\theta}) \int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) d\boldsymbol{\psi} d\boldsymbol{\theta}. \tag{10.3.8}$$

Then, the posterior density of $\boldsymbol{\theta}$ is:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) &= \int p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) d\boldsymbol{\psi} \\ &:= w^*\delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + (1-w^*) \frac{\int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) d\boldsymbol{\psi} \pi(\boldsymbol{\theta})}{\int \pi(\boldsymbol{\theta}) \int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) d\boldsymbol{\psi} d\boldsymbol{\theta}} \end{aligned} \tag{10.3.9}$$

where:

$$w^* = \frac{w \int p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta} = \boldsymbol{\theta}_0) d\boldsymbol{\psi}}{p(\mathbf{y})}, \tag{10.3.10}$$

which is the probability of the event $\{\boldsymbol{\theta} \in \boldsymbol{\Theta}_\theta : \boldsymbol{\theta} = \boldsymbol{\theta}_0\}$. (10.3.9) is the density of a mixed random variable. Under this posterior, $\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})$ accounts for only two possible values, ± 1, and:

$$P_{\boldsymbol{\theta}|\mathbf{y}}(\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = -1) = w^*, P_{\boldsymbol{\theta}|\mathbf{y}}(\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = 1) = 1 - w^*.$$

Thus, the expected posterior net loss is:

$$\begin{aligned} E_{\boldsymbol{\theta}|\mathbf{y}}(\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})) &= -w^* + 1 - w^* \\ &= -\frac{w \int p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta} = \boldsymbol{\theta}_0) d\boldsymbol{\psi}}{p(\mathbf{y})} \\ &\quad + \frac{(1-w) \int \pi(\boldsymbol{\theta}) \int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) d\boldsymbol{\psi} d\boldsymbol{\theta}}{p(\mathbf{y})}, \end{aligned}$$

where the second equality is due to (10.3.10).

The decision rule is:

$$\text{reject } H_0 \text{ iff } w \int p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta} = \boldsymbol{\theta}_0) d\boldsymbol{\psi} < (1-w) \int \pi(\boldsymbol{\theta}) \int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) d\boldsymbol{\psi} d\boldsymbol{\theta}.$$

To represent the prior ignorance, w is often set to $1/2$. In this case, the decision rule is:

$$\text{reject } H_0 \text{ iff } B_{01} = \frac{\int p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta} = \boldsymbol{\theta}_0) \mathrm{d}\boldsymbol{\psi}}{\int \int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi}} = \frac{m_0}{m_1} < 1, \tag{10.3.11}$$

where $\{m_k, k = 0, 1\}$ are the marginal likelihoods for H_0 and H_1, respectively.

As the ratio of the marginal likelihoods, B_{01} is the well-known BF (Kass and Raftery, 1995). An important condition for BF to be valid in hypothesis testing is $w = P(\boldsymbol{\theta} = \boldsymbol{\theta}_0) > 0$, which means that H_0 must have a positive probability mass although $\boldsymbol{\theta}$ is a point in the parameter space. If this condition is not satisfied such that $p(\boldsymbol{\theta})$ is continuous, then:

$$\begin{aligned} & \int \triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi})}{p(\mathbf{y})} d\boldsymbol{\psi} d\boldsymbol{\theta} \\ = & \int \triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y})} d\boldsymbol{\psi} \mathrm{d}\boldsymbol{\theta} \\ = & \int \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y})} d\boldsymbol{\psi} \mathrm{d}\boldsymbol{\theta}. \end{aligned}$$

This equation is always positive so that H_0 is always rejected.

In the Bayesian literature, BF serves as the gold standard for model comparison after posterior distributions are obtained for candidate models. BF is intuitively appealing and has a strong probabilistic interpretation but is known to suffer from some theoretical and computational difficulties. First, when a subjective prior $\pi(\boldsymbol{\theta})$ is not available, Jeffreys' prior or reference prior (Jeffreys, 1961; Bernardo and Smith, 2009) are often used to reflect the lack of prior information. Jeffreys' prior and reference prior are generally improper. Thus, $\pi(\boldsymbol{\theta}) = Cf(\boldsymbol{\theta})$, where $f(\boldsymbol{\theta})$ is a nonintegrable function, and C is an arbitrary positive constant. In this case, the BF can be expressed as:

$$B_{01} = \frac{1}{C} \frac{\int_\Psi p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}_0) p(\boldsymbol{\psi}|\boldsymbol{\theta}_0) \mathrm{d}\boldsymbol{\psi}}{\int_\Theta \int_\Psi p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi}}.$$

Clearly, the BF is ill-defined because it depends on the arbitrary constant, C.

Second, to address the ill-defined problem of BF under the improper prior, a proper prior $\pi(\boldsymbol{\theta})$ with a large variance (i.e., a vague prior) has been proposed to represent the prior ignorance. While the BF is well-defined in this case, it has a tendency to favor the null hypothesis, even when the null hypothesis is incorrect, leading to the notorious Jeffreys-Lindley-Bartlett's paradox (see Jeffreys (1961), Lindley (1957), Bartlett (1957), Poirier (1995) and Robert (1993, 2007)).

To explain this statistical paradox, we use the example in Liu et al. (2021). Let $\mathbf{y}$ be an independent and identically distributed random sample from a normal distribution $N(\theta, \sigma^2)$ with a known variance σ^2. Suppose the null hypothesis is $H_0 : \theta = 0$ and the alternative is $H_1 : \theta \neq 0$. Suppose the prior distribution of θ is $N(\mu_0, \tau^2)$. The posterior odds are expressed as the ratio of the posterior probabilities of the alternative and null hypotheses:

$$PO_{10} = \frac{p(H_1|\bar{y})}{p(H_0|\bar{y})} = \frac{p(\bar{y}|H_1)}{p(\bar{y}|H_0)} \times \frac{p(H_1)}{p(H_0)} = BF_{10} \times \frac{p(H_1)}{p(H_0)} \tag{10.3.12}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $p(H_0)$ and $p(H_1)$ are the prior probabilities of H_0 and H_1 respectively; and BF_{10} is the ratio of marginal densities of $\bar{y}$ under the null and alternative hypotheses. The posterior distribution under the alternative hypothesis is defined as follows:

$$\theta|\bar{y} \sim N\left(\mu^*, \sigma^{*2}\right)$$

where:

$$\mu^* = \sigma^{*2}\left(\frac{\mu_0}{\tau^2} + \frac{n\bar{y}}{\sigma^2}\right) = \frac{\mu_0\sigma^2 + n\tau^2\bar{y}}{\sigma^2 + n\tau^2},\ \sigma^{*2} = \frac{1}{\frac{\mu_0}{\tau^2} + \frac{n}{\sigma^2}} = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}.$$

The marginal densities under the different hypotheses are:

$$\bar{y}|H_0 \sim N\left(0, \sigma^2/n\right), \tag{10.3.13}$$

$$\bar{y}|H_1 \sim N\left(\mu_0, \sigma^2/n + \tau^2\right). \tag{10.3.14}$$

Then, the logarithm of BF can be written as:

$$2\log BF_{10} = \left(z\left(\bar{y}\right)\right)^2 - \frac{1}{\sigma^2/n + \tau^2}\left(\frac{\sigma}{\sqrt{n}}z\left(\bar{y}\right) - \mu_0\right)^2 + \log\frac{\sigma^2}{\sigma^2 + n\tau^2}, \tag{10.3.15}$$

where $z\left(\bar{y}\right) = \sqrt{n}\left(\bar{y} - 0\right)/\sigma$ is the standard z-statistic. As noted by Bartlett (1957), for any fixed dataset, and hence, the fixed $\bar{y}$ and the fixed sample size n, (10.3.15) suggests that $\log BF_{10} \to -\infty$ as $\tau^2 \to \infty$. Even when the absolute value of $z\left(\bar{y}\right)$ is sufficiently large so that H_0 is rejected in the frequentist inference, the BF favors H_0. This type of disagreement is due to the arbitrary value of τ which reflects some indeterminacy of the prior density.

Jeffreys-Lindley-Bartlett's paradox leads researchers to find variations to the BF. Examples include *partial Bayes factor* (O'Hagan, 1991), the *intrinsic Bayes factor* (Berger and Pericchi, 1996), and the *fractional Bayes factor* (O'Hagan, 1995). These variants basically split the data $\mathbf{y}$ into a training sample and a testing sample. The training sample is used to update an uninformative prior to obtain an informative prior. Unfortunately, these methods suffer from more or less arbitrary choices of training samples, weights for averaging training samples, and fractions, respectively.

Finally, for the latent variable model and many other models, calculation of the marginal likelihood $M_k, k = 0, 1$ often involves intractable high-dimensional integrals, and as a result, BFs are generally very difficult to calculate; see Han and Carlin (2001) for an excellent review of methods for calculating the BFs from the MCMC output.

Bernardo and Rueda (2002) and the K-L loss function

Bernardo and Rueda (2002, BR) pointed out that if $\boldsymbol{\theta}$ is a continuous parameter, hypothesis testing forces the use of a non-regular (not absolutely continuous) 'sharp' prior to concentrating a positive probability mass so that the null hypothesis H_0 must have a strictly positive prior probability. This nonregular prior structure leads to the theoretical difficulties of BFs. To overcome these difficulties, Bernardo and Rueda (2002) suggested using a continuous loss function based on the Kullback–Leibler (KL) divergence to replace the discrete loss function:

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

where $p(x)$ and $q(x)$ are any two regular probability density functions.

Based on the Kullback–Leibler (KL) divergence, BR takes the form of the net loss function as:

$$\begin{aligned}
&\triangle\mathcal{L}(H_0,\boldsymbol{\theta},\boldsymbol{\psi}) = m\left(\boldsymbol{\theta}_0,\boldsymbol{\theta},\boldsymbol{\psi}\right) - c\\
&m\left(\boldsymbol{\theta}_0,\boldsymbol{\theta},\boldsymbol{\psi}\right) = \min\left\{KL\left[p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi}),p(\mathbf{y}|\boldsymbol{\theta}_0,\boldsymbol{\psi})\right],KL\left[p(\mathbf{y}|\boldsymbol{\theta}_0,\boldsymbol{\psi}),p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi})\right]\right\}
\end{aligned}$$

As shown in section 3.1, as to this type of net loss function, a possible loss function for decision making can be given by:

$$\mathcal{L}(d_0,\boldsymbol{\theta},\boldsymbol{\psi}) = \begin{cases} c_0 & \text{if } \boldsymbol{\theta}=\boldsymbol{\theta}_0\\ c_0+\frac{1}{2}\min\left\{KL\left[p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi}),p(\mathbf{y}|\boldsymbol{\theta}_0,\boldsymbol{\psi})\right],KL\left[p(\mathbf{y}|\boldsymbol{\theta}_0,\boldsymbol{\psi}),p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi})\right]\right\} & \text{if } \boldsymbol{\theta}\neq\boldsymbol{\theta}_0 \end{cases}$$

$$\mathcal{L}(d_1,\boldsymbol{\theta},\boldsymbol{\psi}) = \begin{cases} c_1 & \text{if } \boldsymbol{\theta}=\boldsymbol{\theta}_0\\ c_1-\frac{1}{2}\min\left\{KL\left[p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi}),p(\mathbf{y}|\boldsymbol{\theta}_0,\boldsymbol{\psi})\right],KL\left[p(\mathbf{y}|\boldsymbol{\theta}_0,\boldsymbol{\psi}),p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi})\right]\right\} & \text{if } \boldsymbol{\theta}\neq\boldsymbol{\theta}_0 \end{cases}$$

where c_i measures the cost of action d_i and $c = c_1 - c_0 > 0$. Then, the corresponding hypothesis test statistic can be given by:

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

A Bayesian admissible solution to the decision problem should satisfy:

$$\text{Reject } H_0 \text{ iff } \mathbf{T}_{BR}(\mathbf{y},\boldsymbol{\theta}_0) > c$$

While $\mathbf{T}_{BR}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ is well-defined under improper priors, because the KL divergence function often does not have a closed-form expression, $\mathbf{T}_{BR}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ is difficult to compute for the latent variable model. c is also difficult to determine. BR suggested choosing threshold values based on the normal distribution for c to implement the test. The rationale for basing threshold values on the normal distribution conceivably comes from the fact that many test statistics are asymptotically normally distributed. Therefore, BR's approach is not Bayesian as the sampling distribution of the test statistic is used and it is based on the idea of repeated sampling, not conditional on $\mathbf{y}$.

For the normal mean hypothesis testing problem we have discussed before, $m\left(0,\theta\right)$ takes the form:

$$m\left(0,\theta\right) = \int N\left(\bar{y}|\theta,\frac{\sigma^2}{n}\right)\log\frac{N\left(\bar{y}|\theta,\sigma^2/n\right)}{N\left(\bar{y}|0,\sigma^2/n\right)}d\bar{y} = \frac{1}{2}\frac{n\theta^2}{\sigma^2},$$

Then, we have:

$$\mathbf{T}_{BR}(\mathbf{y},\theta_0) = \frac{1}{2}\int\frac{n\theta^2}{\sigma^2}N\left(\theta|\mu^*,\sigma^{2*}\right)d\theta = \frac{1}{2}\frac{n}{\sigma^2}\left(\mu^{*2}+\sigma^{2*}\right).$$

It can be shown that:

$$\mathbf{T}_{BR}(\mathbf{y},\theta_0) = \frac{1}{2}\frac{n\tau^2}{\sigma^2+n\tau^2}\left(\frac{n\tau^2}{\sigma^2+n\tau^2}z\left(\bar{y}\right)^2 + 2\frac{\sqrt{n}\mu_0\sigma}{\sigma^2+n\tau^2}z\left(\bar{y}\right) + \frac{\sigma^2\mu_0^2}{\tau^2\left(\sigma^2+n\tau^2\right)} + 1\right)$$

which is immune to Jeffreys-Lindly-Bartlett's paradox. Under the condition that $\mu_0 = 0$, $\sigma = 1$, we have:

$$\mathbf{T}_{BR}(\mathbf{y},\theta_0) = \frac{1}{2}\frac{n\tau^2}{1+n\tau^2}\left(\frac{n\tau^2}{1+n\tau^2}z\left(\bar{y}\right)^2+1\right).$$

As $n\to\infty$, we can further obtain:

$$\mathbf{T}_{BR}(\mathbf{y},\theta_0) - \frac{1}{2}\left(z\left(\bar{y}\right)^2+1\right)\xrightarrow{p}0$$

under the null hypothesis.

Li and Yu (2012) and the $\mathcal{Q}$ loss function

To address the computational problem in $\mathbf{T}_{BR}(\mathbf{y},\boldsymbol{\theta}_0)$, Li and Yu (2012, LY) proposed a loss function based on the $\mathcal{Q}$ function used in the EM algorithm (Dempster et al., 1977) to replace the KL divergence function. For any two points such as $\boldsymbol{\vartheta}_1$ and $\boldsymbol{\vartheta}_2$ defined in the parameter space, the $\mathcal{Q}$ function can be expressed as:

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

Compared with the observed data likelihood function $p(\mathbf{y}|\boldsymbol{\vartheta})$, the $\mathcal{Q}$ function is easier to evaluate for the latent variable model. Let $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\psi})$, Li and Yu (2012) defined a new continuous net loss function as:

$$\begin{aligned}\triangle\mathcal{L}(\boldsymbol{\vartheta},\boldsymbol{\vartheta}_0) &= m(\boldsymbol{\theta}_0,\boldsymbol{\theta},\boldsymbol{\psi}) - c\\ m(\boldsymbol{\theta}_0,\boldsymbol{\theta},\boldsymbol{\psi}) &= \{\mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}) - \mathcal{Q}(\boldsymbol{\vartheta}_0|\boldsymbol{\vartheta})\} + \{\mathcal{Q}(\boldsymbol{\vartheta}_0|\boldsymbol{\vartheta}_0) - \mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}_0)\}\end{aligned}$$

Correspondingly, for this type of net loss function, a possible loss function for decision making can be given by:

$$\mathcal{L}(d_0,\boldsymbol{\theta},\boldsymbol{\psi}) = \begin{cases} c_0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0\\ c_0 + \frac{1}{2}\{\mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}) - \mathcal{Q}(\boldsymbol{\vartheta}_0|\boldsymbol{\vartheta})\} + \frac{1}{2}\{\mathcal{Q}(\boldsymbol{\vartheta}_0|\boldsymbol{\vartheta}_0) - \mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}_0)\} & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0\end{cases}$$

$$\mathcal{L}(d_1,\boldsymbol{\theta},\boldsymbol{\psi}) = \begin{cases} c_1 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0\\ c_1 - \frac{1}{2}\{\mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}) - \mathcal{Q}(\boldsymbol{\vartheta}_0|\boldsymbol{\vartheta})\} - \frac{1}{2}\{\mathcal{Q}(\boldsymbol{\vartheta}_0|\boldsymbol{\vartheta}_0) - \mathcal{Q}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}_0)\} & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0\end{cases}$$

where c_i measures the cost of action d_i and $c = c_1 - c_0 > 0$. Thus, LY proposed a posterior-based test statistic as follows:

$$\mathbf{T}_{LY}(\mathbf{y},\boldsymbol{\theta}_0) = E_{\boldsymbol{\vartheta}|\mathbf{y}}\left[m(\boldsymbol{\theta}_0,\boldsymbol{\theta},\boldsymbol{\psi})\right]$$

Although $T_{LY}(\mathbf{y},\boldsymbol{\theta}_0)$ is well-defined with improper priors and is easy to compute for the latent variable model, one still must specify some threshold values for c to make a decision. Again, these threshold values given by LY lack rigorous statistical justifications.

10.4 Bayesian version of LR, LM and Wald test statistics

To address the problem in choosing threshold values, Li et al. (2015, 2022) and Liu et al. (2021) introduce another three loss functions. The corresponding test statistics can be explained as the posterior version of LR, LM and Wald test statistics, which are popular in the frequentist paradigm.

Assuming that $\mathbf{y}$ comes from a probability measure P_0 on the probability space (Ω, F, P_0), let $P_{\boldsymbol{\vartheta}}$ be a collection of candidate models indexed by parameters $\boldsymbol{\vartheta}$. Following White (1987), if there exists $\boldsymbol{\vartheta}$ such that $P_0 \in P_{\boldsymbol{\vartheta}}$, we call that the model $P_{\boldsymbol{\vartheta}}$ is correctly specified. If for any $\boldsymbol{\vartheta}$, $P_0 \notin P_{\boldsymbol{\vartheta}}$, we say the model $P_{\boldsymbol{\vartheta}}$ is misspecified. In this subsection, we assume that the model is correctly specified.

In this subsection, we establish large sample properties for $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)$ under repeated sampling. Let $\mathbf{y}^t := (y_0, y_1, \ldots, y_t)$ for any $0 \le t \le n$ and $l_t\left(\mathbf{y}^t,\boldsymbol{\vartheta}\right) = \log p(\mathbf{y}^t|\boldsymbol{\vartheta}) - \log p(\mathbf{y}^{t-1}|\boldsymbol{\vartheta})$ be the conditional log-likelihood for the t^{th} observation for any $1 \le t \le n$. When there is no confusion, we just write $l_t\left(\mathbf{y}^t,\boldsymbol{\vartheta}\right)$ as $l_t(\boldsymbol{\vartheta})$ so that the log-likelihood function $\mathcal{L}_n(\boldsymbol{\vartheta})$ ($= \log p(\mathbf{y}|\boldsymbol{\vartheta})$ conditional on the initial observation) can be written as $\sum_{t=1}^n l_t(\boldsymbol{\vartheta})$. Let $l_t^{(j)}(\boldsymbol{\vartheta})$ be the j^{th} derivative of $l_t(\boldsymbol{\vartheta})$ and $l_t^{(0)}(\boldsymbol{\vartheta}) = l_t(\boldsymbol{\vartheta})$. Moreover, let:

$$\mathbf{s}(\mathbf{y}^t,\boldsymbol{\vartheta}) = \frac{\partial \log p(\mathbf{y}^t|\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}} = \sum_{i=1}^t l_i^{(1)}(\boldsymbol{\vartheta}),\ \mathbf{h}(\mathbf{y}^t,\boldsymbol{\vartheta}) = \frac{\partial^2 \log p(\mathbf{y}^t|\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} = \sum_{i=1}^t l_i^{(2)}(\boldsymbol{\vartheta}),$$

$$\mathbf{s}_t(\boldsymbol{\vartheta}) = l_t^{(1)}(\boldsymbol{\vartheta}) = \mathbf{s}(\mathbf{y}^t,\boldsymbol{\vartheta}) - \mathbf{s}(\mathbf{y}^{t-1},\boldsymbol{\vartheta}),\ \mathbf{h}_t(\boldsymbol{\vartheta}) = l_t^{(2)}(\boldsymbol{\vartheta}) = \mathbf{h}(\mathbf{y}^t,\boldsymbol{\vartheta}) - \mathbf{h}(\mathbf{y}^{t-1},\boldsymbol{\vartheta}),$$

$$\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{t=1}^{n}\mathbf{h}_t(\boldsymbol{\vartheta}), \bar{\mathbf{J}}_n(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{t=1}^{n}\left[\mathbf{s}_t(\boldsymbol{\vartheta}) - \bar{\mathbf{s}}_t(\boldsymbol{\vartheta})\right]\left[\mathbf{s}_t(\boldsymbol{\vartheta}) - \bar{\mathbf{s}}_t(\boldsymbol{\vartheta})\right]', \bar{\mathbf{s}}_t(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{t=1}^{n}\mathbf{s}_t(\boldsymbol{\vartheta}),$$

$$\mathbf{B}_n(\boldsymbol{\vartheta}) = Var\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} l_t^{(1)}(\boldsymbol{\vartheta})\right], \mathbf{H}_n(\boldsymbol{\vartheta}) = \int \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) g(\mathbf{y})\, d\mathbf{y},\ \mathbf{J}_n(\boldsymbol{\vartheta}) = \int \bar{\mathbf{J}}_n(\boldsymbol{\vartheta}) g(\mathbf{y})\, d\mathbf{y},$$

where $g(\mathbf{y})$ is the data generating process (DGP). In the literature, $\mathbf{H}_n(\boldsymbol{\vartheta})$ and $\mathbf{J}_n(\boldsymbol{\vartheta})$ are generally known as the Hessian matrix and the Fisher information matrix; $\bar{\mathbf{H}}_n(\boldsymbol{\vartheta})$ and $\bar{\mathbf{J}}_n(\boldsymbol{\vartheta})$ are the empirical Hessian matrix and empirical Fisher information matrix.

In this subsection, to show the equivalence between the posterior test statistics and their frequentist versions, we introduce the following regularity conditions:

Assumption 1: For $q = q_\theta + q_\psi$, $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^q$.

Assumption 2: For any $\varepsilon > 0$ and $r > 2$, the α-mixing condition with the coefficient $\alpha(m) = O\left(m^{\frac{-2r}{r-2}-\varepsilon}\right)$ is satisfied for $\{y_t\}_{t=1}^{\infty}$.

Assumption 3: For all t, $l_t(\boldsymbol{\vartheta})$ is three-times differentiable on $\boldsymbol{\Theta}$ almost surely.

Assumption 4: For any $\boldsymbol{\vartheta}, \boldsymbol{\vartheta}' \in \boldsymbol{\Theta}$, $\left\|l_t^{(j)}(\boldsymbol{\vartheta}) - l_t^{(j)}(\boldsymbol{\vartheta}')\right\| \leq c_t^j(\mathbf{y}^t)\left\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\right\|$ in probability, where $c_t^j(\mathbf{y}^t) > 0$, $\sup_t E\left\|c_t^j(\mathbf{y}^t)\right\| < \infty$, $\frac{1}{n}\sum_{t=1}^{n}\left(c_t^j(\mathbf{y}^t) - E\left(c_t^j(\mathbf{y}^t)\right)\right) \xrightarrow{p} 0$, and $j = 0, 1, 2$.

Assumption 5: For all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, there exists $M_t(\mathbf{y}^t) > 0$ such that $l_t^{(j)}(\boldsymbol{\vartheta})$ exists, $\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|l_t^{(j)}(\boldsymbol{\vartheta})\right\| \leqslant M_t(\mathbf{y}^t)$, and $\sup_t E\left\|M_t(\mathbf{y}^t)\right\|^{r+\delta} \leq M$ for some $\delta > 0$ and $M < \infty$, where r is the same as that in Assumption 2, and $j = 0, 1, 2$.

Assumption 6: For $0 \leqslant j \leqslant 1$, $\left\{l_t^{(j)}(\boldsymbol{\vartheta})\right\}$ is L_2-near epoch dependent of size -1 and for $j = 2$, of size $-\frac{1}{2}$ uniformly on $\boldsymbol{\Theta}$.

Assumption 7: For any $\varepsilon > 0$:

$$\lim_{n\to\infty}\sup \sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}\setminus N\left(\boldsymbol{\vartheta}^0,\varepsilon\right)} \frac{1}{n}\sum_{t=1}^{n}\left\{E\left[l_t(\boldsymbol{\vartheta})\right] - E\left[l_t\left(\boldsymbol{\vartheta}^0\right)\right]\right\} < 0, \tag{10.4.1}$$

where $N\left(\boldsymbol{\vartheta}^0, \varepsilon\right)$ is the open ball of radius ε around the true value $\boldsymbol{\vartheta}^0$.

Assumption 8: The prior distribution $p(\boldsymbol{\vartheta})$ is assumed to be thrice continuously differentiable and $0 < p\left(\boldsymbol{\vartheta}_n^0\right) < \infty$ uniformly in n. For some n^*, when $n > n^*$, it is assumed that the posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$ is proper and $\int \|\boldsymbol{\vartheta}\|^2 p(\boldsymbol{\vartheta}|\mathbf{y})\, d\boldsymbol{\vartheta} < +\infty$.

Remark 10.4.1 *For dependent and heterogeneous data, Assumptions 1-7 are standard primitive conditions to develop the maximum likelihood theory (e.g., consistency and asymptotic normality). Assumption 1 is the compactness condition. Assumption 2 shows the weak dependence in y_t. Assumption 3 is the continuity condition. Assumption 4 is the Lipschitz condition for l_t. Assumption 5 is the domination condition for l_t. Assumption 6 shows weak dependence in l_t. Assumption 7 is the identification condition. Assumption 8 is the primitive condition on the prior distribution. More details about these regularity conditions can be found in Gallant and White (1988), Wooldridge (1994), Li et al. (2017, 2022) and Liu et al. (2021).*

10.4.1 Li et al. (2022) and LR-type loss function

Spiegelhalter et al. (2002) defined the following Bayesian deviance function to measure the Bayesian model fit of a candidate model:

$$D(\mathbf{y}) = \int \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}. \tag{10.4.2}$$

Based on the Bayesian deviance function, we can develop a loss function for hypothesis testing as follows:

$$\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_0 + \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - \ln p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases},$$

$$\mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_1 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_1 + \ln p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi}) - \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases},$$

where c_i measures the cost of action d_i. Let $c = c_1 - c_0 > 0$ and we have:

$$\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = 2\left(\ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - \ln p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})\right) - c.$$

According to the original definition given in (10.3.3), the corresponding test statistic can be established as:

$$\begin{aligned} \mathbf{T}_0(\mathbf{y}, \boldsymbol{\theta_0}) &= \int\int \triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi} \\ &= \int\int \left[2\left(\ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - \ln p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})\right) - c\right] p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}. \quad (10.4.3) \end{aligned}$$

Let:

$$\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta_0}) = \int\int \left[2\left(\ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) - \ln p(\mathbf{y}|\boldsymbol{\theta}_0, \boldsymbol{\psi})\right)\right] p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}.$$

Then, $\mathbf{T}_0\left(\mathbf{y}, \boldsymbol{\theta}_0\right) > 0$ is equivalent to $\mathbf{T}_1\left(\mathbf{y}, \boldsymbol{\theta}_0\right) > c$. In this case, H_0 is rejected.

The posterior test statistic $\mathbf{T}_1(\mathbf{y}, \boldsymbol{\theta_0})$ was first introduced by Li et al. (2014, LZY). To determine c for hypothesis testing, given some mild regularity conditions, Li et al. (2014) derived the asymptotic distribution of the test statistic as:

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

where $\boldsymbol{\varepsilon}$ is a standard multivariate normal variate, $\boldsymbol{\vartheta_0} = (\boldsymbol{\theta_0}, \boldsymbol{\psi}_0)'$ is the true value of $\boldsymbol{\vartheta}$, $\mathbf{J}(\boldsymbol{\vartheta_0})$ is the Fisher information matrix given by:

$$\mathbf{J}(\boldsymbol{\vartheta_0}) = \frac{1}{n} \int -L_n^{(2)}(\boldsymbol{\vartheta}_0) p(\mathbf{y}|\boldsymbol{\vartheta}_0) \mathrm{d}\mathbf{y},$$

with $\mathbf{IJ}(\boldsymbol{\vartheta_0})$ being the inverse of $\mathbf{J}(\boldsymbol{\vartheta_0})$, $\mathbf{J}_{11}(\boldsymbol{\vartheta_0})$ and $\mathbf{IJ}_{11}(\boldsymbol{\vartheta_0})$ being the submatrices of $\mathbf{J}(\boldsymbol{\vartheta_0})$ and $\mathbf{IJ}(\boldsymbol{\vartheta_0})$, respectively, both corresponding to $\boldsymbol{\theta}$. Based on the asymptotic distribution, quantiles at certain probability levels can be chosen to make statistical inferences.

The asymptotic distribution is not pivotal. To overcome this disadvantage, Li et al. (2022, LWYZ) revised the loss function as:

$$\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_0 + \left[2\ln p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) - \ln p(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi}) - D_c(\mathbf{y}, \boldsymbol{\theta}_0)\right] & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \quad (10.4.4)$$

$$\mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_1 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_1 - \left[2\ln p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) - \ln p(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi}) - D_c(\mathbf{y}, \boldsymbol{\theta}_0)\right] & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \quad (10.4.5)$$

where $D_c(\mathbf{y}, \boldsymbol{\theta}_0) = \int \ln p(\mathbf{y}, \boldsymbol{\theta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}, \boldsymbol{\theta}_0|\mathbf{y}) \mathrm{d}\boldsymbol{\psi}$ is the Bayesian complete deviance function under the null hypothesis. It can be shown that:

$$\Delta\mathcal{L}\left(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = m\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) - c,$$

where:

$$m\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = 4\ln p(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) - 2\ln p(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\psi}) - 2D_c(\mathbf{y}, \boldsymbol{\theta}_0).$$

The corresponding test statistic is:

$$\mathbf{T}_{LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) = E_{\vartheta|y}\left[m\left(\boldsymbol{\theta}_0,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right]$$
$$= \int \left[4\ln p(\mathbf{y},\hat{\boldsymbol{\vartheta}}) - 2\ln p(\mathbf{y},\boldsymbol{\theta},\boldsymbol{\psi}) - 2D_c(\mathbf{y},\boldsymbol{\theta}_0)\right] p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi}.$$

We can show that we:

$$\text{reject } H_0 \text{ iff } \mathbf{T}_{LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) > c. \tag{10.4.6}$$

Again, one can derive the asymptotic distribution of $\mathbf{T}_{LWYZ}(\mathbf{y},\boldsymbol{\theta}_0)$ so that quantiles at certain probability levels can be chosen to make statistical inferences.

In many models, the MLE $\hat{\boldsymbol{\vartheta}}$ is not easy to find. Based on $\mathbf{T}_{LWYZ}(\mathbf{y},\boldsymbol{\theta}_0)$, to avoid the use of $\hat{\boldsymbol{\vartheta}}$, Li et al. (2022) defined two test statistics:

$$\mathbf{T}_{1,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) = 2\left[D_c(\mathbf{y}) - D_c(\mathbf{y},\boldsymbol{\theta}_0)\right], \mathbf{T}_{2,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) = 2\left[\ln p(\mathbf{y},\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\psi}}) - \ln p(\mathbf{y},\boldsymbol{\theta}_0,\bar{\boldsymbol{\psi}}_0)\right], \tag{10.4.7}$$

where $D_c(\mathbf{y})$, the Bayesian complete deviance function with the prior information, is given by:

$$D_c(\mathbf{y}) = \int\int \left[\ln p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi}) + \ln p(\boldsymbol{\theta},\boldsymbol{\psi})\right] p(\boldsymbol{\theta},\boldsymbol{\psi}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\psi},$$

and:

$$\ln p(\mathbf{y},\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\psi}}) = \ln p(\mathbf{y}|\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\psi}}) + \ln p(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\psi}}), \ln p(\mathbf{y},\boldsymbol{\theta}_0,\bar{\boldsymbol{\psi}}_0) = \ln p(\mathbf{y}|\boldsymbol{\theta}_0,\bar{\boldsymbol{\psi}}_0) + \ln p(\boldsymbol{\theta}_0,\bar{\boldsymbol{\psi}}_0).$$

Under some regularity conditions, Li et al. (2022) showed that under the null and alternative hypotheses:

$$\mathbf{T}_{LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) = \mathbf{T}_{1,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) + q_\theta + O_p(n^{-1/2}) = \mathbf{T}_{2,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) + O_p(n^{-1/2}), \tag{10.4.8}$$

and under the null hypothesis:

$$\mathbf{T}_{1,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) + q_\theta = \mathbf{T}_{2,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) + O_p(n^{-1/2}) = \mathbf{LR} + O_p(n^{-1/2}) \xrightarrow{d} \chi^2(q_\theta).$$

Thus, we can obtain that:

$$\mathbf{T}_{1,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) + q_\theta \xrightarrow{d} \chi^2(q_\theta), \mathbf{T}_{2,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0) \xrightarrow{d} \chi^2(q_\theta).$$

Li et al. (2022) explained $\mathbf{T}_{1,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0)$ as the Bayesian version of the LR statistic, and $\mathbf{T}_{2,LWYZ}(\mathbf{y},\boldsymbol{\theta}_0)$ as the Bayesian plug-in version of LR statistic. In practice, one can use these posterior test statistics for hypothesis testing.

Using the normal mean hypothesis testing example shown above, we can show that:

$$\mathbf{T}_{1,LWYZ}(\mathbf{y},\theta_0) = \frac{n\tau^2}{\sigma^2+n\tau^2}z\left(\bar{y}\right)^2 + \frac{2\sqrt{n}\sigma\mu_0}{\sigma^2+n\tau^2}z\left(\bar{y}\right) - 1 + \frac{\mu_0^2\sigma^2}{\tau^2(\sigma^2+n\tau^2)},$$
$$\mathbf{T}_{2,LWYZ}(\mathbf{y},\theta_0) = \frac{n\tau^2}{\sigma^2+n\tau^2}z\left(\bar{y}\right)^2 + \frac{2\sqrt{n}\sigma\mu_0}{\sigma^2+n\tau^2}z\left(\bar{y}\right) + \frac{\mu_0^2\sigma^2}{\tau^2(\sigma^2+n\tau^2)},$$

which is immune to Jeffreys-Lindly-Bartlett's paradox. If we set $\mu_0 = 0$ and $\sigma = 1$, the expression can be simplified as:

$$\mathbf{T}_{1,LWYZ}(\mathbf{y},\theta_0) = \frac{n\tau^2}{1+n\tau^2}z\left(\bar{y}\right)^2 - 1, \mathbf{T}_{2,LWYZ}(\mathbf{y},\theta_0) = \frac{n\tau^2}{1+n\tau^2}z\left(\bar{y}\right)^2.$$

When $n \to \infty$, under the null hypothesis, we have:

$$\mathbf{T}_{1,LWYZ}(\mathbf{y},\theta_0) + 1 - z\left(\bar{y}\right)^2 \xrightarrow{p} 0, \mathbf{T}_{2,LWYZ}(\mathbf{y},\theta_0) - z\left(\bar{y}\right)^2 \xrightarrow{p} 0.$$

10.4.2 Li et al. (2015) and LM-type loss function

Let:

$$m(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}), \tag{10.4.9}$$

where:

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

$s(\boldsymbol{\vartheta})$ is the score function of $\boldsymbol{\vartheta}$, $C_{\theta\theta}(\boldsymbol{\vartheta})$ 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 posterior mean of $\boldsymbol{\vartheta}$ under H_0, and $\bar{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$ under H_1.

Based on this quadratic loss, a possible loss function for hypothesis testing can be specified as:

$$\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_0 + \frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \tag{10.4.10}$$

$$\mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_1 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_1 - \frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \tag{10.4.11}$$

Thus, based on this loss function, a posterior test statistic proposed by Li et al. (2015, LLY)can be given by:

$$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0) = E_{\vartheta|y}\left[m(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi})\right] = \int (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0)(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}, \tag{10.4.12}$$

where $p(\boldsymbol{\vartheta}|\mathbf{y})$ is the posterior distribution of $\boldsymbol{\vartheta}$ under H_1. Thus, we can further show that:

$$\text{Reject } H_0 \text{ iff } \mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0) > c, \tag{10.4.13}$$

The proposed test can be viewed as the Bayesian version of the LM test. To show this link, let the LM statistic (Breusch and Pagan, 1980) be:

$$\mathbf{LM} = \frac{1}{n} s_\theta(\widehat{\boldsymbol{\vartheta}}_0) \left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}_0\right)\right] s_\theta(\widehat{\boldsymbol{\vartheta}}_0),$$

where $\widehat{\boldsymbol{\vartheta}}_0 = (\boldsymbol{\theta}_0, \widehat{\boldsymbol{\psi}}_0)$ is the MLE of $\boldsymbol{\vartheta}$ under the null hypothesis, $s_\theta(\boldsymbol{\vartheta})$ is the subvector of $s(\boldsymbol{\vartheta})$ corresponding to $\boldsymbol{\theta}$, and $\bar{\mathbf{H}}_{n,\theta\theta}(\boldsymbol{\vartheta})$ is the submatrix of $\bar{\mathbf{H}}_n(\boldsymbol{\vartheta})$ corresponding to $\boldsymbol{\theta}$. Under some regularity assumptions, when the null hypothesis is true and the likelihood dominates the prior, Li et al. (2015) showed that:

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

The test statistic $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ has a few important properties. For example, the test statistic is well-defined under an improper prior and immune to Jeffreys-Lindley-Bartlett's paradox. In addition, for the latent variable model, it is not difficult to compute with the EM algorithm. Finally, it follows a pivotal $\chi^2(\cdot)$ asymptotically, and thus, it is easy to obtain threshold values.

For the normal mean hypothesis testing, we have:

$$\mathbf{T}_{LLY}(\mathbf{y}, \theta_0) = \frac{n\tau^2}{\sigma^2 + n\tau^2} z(\bar{y})^2.$$

which is immune to Jeffreys-Lindly-Bartlett's paradox. Under the condition that $\mu_0 = 0, \sigma = 1$, we have:

$$\mathbf{T}_{LLY}(\mathbf{y}, \theta_0) = \frac{n\tau^2}{1 + n\tau^2} z(\bar{y})^2.$$

As $n \to \infty$, we can further obtain:

$$\mathbf{T}_{LLY}(\mathbf{y}, \theta_0) - z(\bar{y})^2 \xrightarrow{p} 0$$

under the null hypothesis.

10.4.3 Liu et al. (2021) and Wald-type loss function

Although the test statistic proposed by Li et al. (2015) is convenient to calculate and has some good properties, it requires evaluating the first-order derivative of the likelihood function. In many models, particularly in latent variable models, this first-order derivative is not easy to evaluate because the observed-data likelihood function may not have an analytical expression.[1] Another feature of $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ is that it requires estimating both the null model and the alternative model, although under H_0, it is shown to be asymptotically equivalent to the Lagrange Multiplier (LM) test that requires estimating the null model only. Based on another quadratic loss function, Liu et al. (2021, LLYZ) proposed a test statistic that is only a by-product of the MCMC output under H_1 and thus is easier to compute.

Let the posterior covariance matrix under the alterative hypothesis be:

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

where $\bar{\boldsymbol{\vartheta}}$ is the posterior mean of $\boldsymbol{\vartheta}$ under the alternative hypothesis H_1. Liu et al. (2021) proposed the following net loss function for hypothesis testing:

$$m(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0),$$

where $\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})$ is the submatrix of $\mathbf{V}(\bar{\boldsymbol{\vartheta}})$ corresponding to $\boldsymbol{\theta}$, and $\left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1}$ is the inverse matrix of $\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})$.

Based on this quadratic loss function, one can specify two possible loss functions for hypothesis testing given by:

$$\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_0 + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \tag{10.4.14}$$

$$\mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} c_1 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ c_1 - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \tag{10.4.15}$$

Thus, Liu et al. (2021) proposed the following posterior-based test statistic given by:

$$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) = E_{\vartheta|y}\left[m(\boldsymbol{\theta}_0, \boldsymbol{\theta}, \boldsymbol{\psi})\right] = \int (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}, \tag{10.4.16}$$

Consequently, we can show that:

$$\text{Reject } H_0 \text{ iff } \mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) > c, \tag{10.4.17}$$

To show the link between $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0)$ and the Wald statistic, define the Wald statistic by (Engle, 1984):

$$\textbf{Wald} = n\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]^{-1} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right), \tag{10.4.18}$$

where $\widehat{\boldsymbol{\vartheta}} := \left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}\right)$ is the ML estimate of $\boldsymbol{\vartheta}$. Under some regularity assumptions when the null hypothesis is true and the likelihood dominates the prior, Liu et al. (2021) showed that:

$$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta = \textbf{Wald} + o_p(1) \xrightarrow{d} \chi^2(q_\theta).$$

This result is why $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0)$ may be viewed as a Bayesian version of the Wald test.

[1] Advanced techniques, such as automatic differentiation, can help evaluate derivatives. Skaug and Yu (2014) use the automatic differentiation technique, together with the Laplace approximation, to approximate the likelihood function of stochastic volatility models.

$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0)$ shared some nice properties with the test statistic proposed by Li et al. (2015). First, this test statistic is well-defined under improper prior distributions and is immune to Jeffreys-Lindley-Bartlett's paradox. Second, the asymptotic distribution allows threshold values for c to be easily obtained from the $\chi^2(\cdot)$ distribution to make the decision.

An advantage of $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0)$ compared to $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ is that $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0)$ does not require evaluating the first-order derivative of the likelihood function. Another advantage of $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0)$ over $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ is that $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0)$ only must estimate the alternative model but $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ must estimate both the null model and alternative model.

For the normal mean hypothesis testing, we have:

$$\mathbf{T}_{LLYZ}(\mathbf{y}, \theta_0) = \frac{n\tau^2}{\sigma^2 + n\tau^2} z(\bar{y})^2 + \frac{2\sqrt{n}\mu_0\sigma}{\sigma^2 + n\tau^2} z(\bar{y}) + \frac{\mu_0^2\sigma^2}{(\sigma^2 + n\tau^2)\tau^2} + 1.$$

which is immune Jeffreys-Lindly-Bartlett's paradox. Under the condition that $\mu_0 = 0$, $\sigma = 1$, we have:

$$\mathbf{T}_{LLYZ}(\mathbf{y}, \theta_0) = \frac{n\tau^2}{\sigma^2 + n\tau^2} z(\bar{y})^2 + 1.$$

As $n \to \infty$, we can further obtain:

$$\mathbf{T}_{LLYZ}(\mathbf{y}, \theta_0) - 1 - z(\bar{y})^2 \xrightarrow{p} 0$$

under the null hypothesis.

The following table summarizes the posterior-based trinity of the tests and their key properties. Although constructed from the posterior output, which contains random draws from the Bayesian posterior distribution, the statistical inference made by the three tests is not conditional to the data. Instead, the justification of the three tests is performed in a frequentist framework, requiring repeated sampling from the DGP and an asymptotic argument.

10.5 Simulation Studies

We designed two experiments to examine the finite-sample performance of the proposed test with simulated data. In the first experiment, we examine the finite performance of the proposed posterior statistics in terms of the empirical size and empirical power. In the second experiment, we consider a copula model in a similar manner.

10.5.1 Hypothesis testing in a linear regression model

In this subsection, we consider a simple linear regression model as follows:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim iidN(0, \sigma^2), \quad i = 1, 2, \cdots, n,$$

where x_i is randomly drawn from the standard normal distribution and then fixed under repeated sampling. For point-null hypothesis testing problem, we test whether the slope coefficient is zero:

$$H_0 : \beta = 0 \quad \text{vs} \quad H_1 : \beta \neq 0,$$

For the Bayesian analysis, we choose the prior distributions to be the natural conjugate Normal-Gamma priors:

$$(\alpha, \beta)' \sim N(\mu_0, \sigma^2 V_0) \text{ and } h = \frac{1}{\sigma^2} \sim G(a, b),$$

where $\mu_0 = (\mu_\alpha, \mu_\beta)'$, $V_0 = $diag$(V_\alpha, V_\beta)$, $G(a, b)$ denotes the gamma distribution with the shape parameter a and the rate parameter b. Let $\hat{\mu} = (X'X)^{-1} X'Y$ be the OLS estimator of $\mu = (\alpha, \beta)'$, Summary of posterior-based Trinity of Tests

	$\mathbf{T}_{LYWZ}$	$\mathbf{T}_{LLY}$	$\mathbf{T}_{LLYZ}$
Expression	$2\left[D(\mathbf{y}) - D(\mathbf{y}, \boldsymbol{\theta}_0)\right] + q$	$\mathbf{tr}\left[C_{\theta\theta}(\bar{\boldsymbol{\vartheta}}_0) V_{\theta\theta}(\bar{\boldsymbol{\theta}})\right]$	$\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1} \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$
Prior	Improper or proper	Improper or propers	Improper or proper
Jeffreys-Lindley's Paradox	No	No	No
Asymptotic Theory	$\chi^2(q)$	$\chi^2(q)$	$\chi^2(q)$
Asymptotic Pivotal	Yes	Yes	Yes

where $Y = \begin{pmatrix} y_1 & \dots & y_n \end{pmatrix}'$, $X = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix}'$. Then, the posterior distributions of μ and h under H_1 are:

$$\mu|\mathbf{y}, h; H_1 \sim N(\mu_1, \sigma^2 V_1),$$
$$h|\mathbf{y}; H_1 \sim G\left(a + \frac{n}{2}, b + \frac{1}{2}\left(\mathbf{y}'\mathbf{y} + \mu_0' V_0^{-1}\mu_0 - \mu_1' V_1^{-1}\mu_1\right)\right),$$

where $V_1 = \left(X'X + V_0^{-1}\right)^{-1}$ and $\mu_1 = V_1(X'X\hat{\mu} + V_0^{-1}\mu_0) = V_1(X'\mathbf{y} + V_0^{-1}\mu_0)$.

Under H_0, $y_i = \alpha + \varepsilon_i$ and:

$$\alpha|\mathbf{y}; H_0 \sim N(\mu_{\alpha 1}, \sigma^2 V_{\alpha 1}),$$
$$h|\mathbf{y}; H_0 \sim G\left(a + \frac{n}{2}, b + \frac{1}{2}\left(\mathbf{y}'\mathbf{y} + \frac{\mu_\alpha^2}{V_\alpha} - \frac{\mu_{\alpha 1}^2}{V_{\alpha 1}}\right)\right),$$

where $V_{\alpha 1} = \frac{V_\alpha}{nV_\alpha + 1}$ and $\mu_{\alpha 1} = V_{\alpha 1}\left(\sum_{i=1}^n y_i + \frac{\mu_\alpha}{V_\alpha}\right)$. We randomly draw 10,000 iid samples for each parameter. Based on these random samples, we calculate the posterior statistics.

Particularly, we set the true values for α^* and σ^{2*} to equal 1 and 1. We consider three different values for β^*, $\beta^* = 0.0, 0.1, 0.2$, and three different sample sizes, $n = 50$, 500, 2000. For each case, we simulate data from the true DGP and perform hypothesis testing at the 5% significance level 1,000 times. We report the rejection rate of H_0 across the 1,000 replications.

We consider two types of priors, noninformative prior distributions (NP) and informative prior distributions (IP) of parameters of interest as follows:

$$NP : (\mu_\alpha, \mu_\beta, V_\alpha, V_\beta, a, b) = (\alpha^*, \beta^*, 10000, 10000, 1, 1).$$
$$IP : (\mu_\alpha, \mu_\beta, V_\alpha, V_\beta, a, b) = (\alpha^*, \beta^*, 10000, 0.001, 1, 1).$$

The empirical size and power of various statistics under different sample sizes are reported in Table 1 (under uninformative priors) and Table 2 (under informative priors).

The results in Table 1 show good finite sample performance of the proposed posterior statistics in terms of empirical size and empirical power under uninformative prior distributions (NP). For the empirical size, when the sample size is small (e.g., $n = 50$), the empirical sizes of different statistics reported in this study exhibit some variation, ranging from 4.5% to 6.5%. As the sample size increases, the empirical sizes of different statistics increasingly agree with each other and are all estimated to be around the nominal size of 5.0%. The results regarding size are consistent with our theoretical predictions. Regarding the empirical power, as the sample size increases, and β^* deviates from the hypothesized value under the null hypothesis, the empirical powers of all statistics reported in this study increase to 100.0%.

Table 2 shows the empirical size and power of these statistics in a linear regression model under informative prior distributions (IP). The performance of the frequentist $LR, Wald, LM$ statistics remain the same as those under uninformative prior distributions (NP) reported in Table 1. For the posterior statistics, when informative prior information exists, the empirical size and power are strongly affected, particularly when the sample size is small. These results are easy to understand because the prior information matters for the posterior distribution of parameters when data information is insufficient (e.g., when the sample size is small). More specifically, when H_0 is true, the empirical size deviates markedly downwards from the nominal size of 5.0% (e.g., 0.0% when $n = 50, 500$; approximately 1.4% when $n = 2,000$). However, as the sample size increases, and the data information begins to dominate the prior information, the empirical sizes of all the posterior statistics increase to be approximately 5% ($n = 20,000$). When H_1 is true, the empirical powers of $\mathbf{T}_{LYWZ}^1(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$, $\mathbf{T}_{LYWZ}^2(\mathbf{y}, \boldsymbol{\theta}_0)$ and $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$ increase markedly, equaling 100.0% even under a small sample size ($n = 50$). However, for $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$, the empirical power decreases sharply under small sample size, equaling 0% with $n = 50$, and gradually recovers to 100.0% as the sample size increases.

Table 10.1: Empirical size and power of posterior statistics in a linear regression model (NP)

	Empirical Size ($\beta^* = 0.0$)		
	$n = 50$	$n = 500$	$n = 2000$
LR	5.2%	4.1%	4.6%
$Wald$	5.4%	4.1%	4.6%
LM	6.5%	4.3%	4.6%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	4.9%	4.0%	4.6%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	4.9%	4.1%	4.6%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	5.3%	4.1%	4.6%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	4.5%	4.0%	4.7%
	Empirical Power ($\beta^* = 0.1$)		
	$n = 50$	$n = 500$	$n = 2000$
LR	9.2%	59.8%	99.4%
$Wald$	9.2%	59.8%	99.4%
LM	11.5%	60.3%	99.4%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	8.8%	59.7%	99.4%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	8.8%	59.6%	99.4%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	9.2%	60.0%	99.4%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	7.9%	59.4%	99.4%
	Empirical Power ($\beta^* = 0.2$)		
	$n = 50$	$n = 500$	$n = 2000$
LR	25.5%	99.8%	100.0%
$Wald$	25.8%	99.8%	100.0%
LM	28.3%	99.8%	100.0%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	23.9%	99.8%	100.0%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	23.8%	99.8%	100.0%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	25.7%	99.8%	100.0%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	23.0%	99.8%	100.0%

Table 10.2: Empirical size and power of posterior statistics in a linear regression model (IP)

	Empirical Size ($\beta^* = 0.0$)			
	$n = 50$	$n = 500$	$n = 2000$	$n = 20000$
LR	5.2%	4.1%	4.6%	4.8%
$Wald$	5.4%	4.1%	4.6%	4.8%
LM	6.5%	4.3%	4.6%	4.8%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	0.0%	0.0%	1.4%	4.6%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	0.0%	0.0%	1.5%	4.6%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	0.0%	0.0%	1.4%	4.8%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	0.0%	0.0%	1.4%	4.4%
	Empirical Power ($\beta^* = 0.1$)			
	$n = 50$	$n = 500$	$n = 2000$	$n = 20000$
LR	9.2%	59.8%	99.4%	100.0%
$Wald$	9.2%	59.8%	99.4%	100.0%
LM	11.5%	60.3%	99.4%	100.0%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	0.0%	8.4%	97.8%	100.0%
	Empirical Power ($\beta^* = 0.2$)			
	$n = 50$	$n = 500$	$n = 2000$	$n = 20000$
LR	25.5%	99.8%	100.0%	100.0%
$Wald$	25.8%	99.8%	100.0%	100.0%
LM	28.3%	99.8%	100.0%	100.0%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	0.0%	81.5%	100.0%	100.0%

10.5.2 Hypothesis testing in copula models

In this subsection, we use the Gaussian copula model with a Gaussian margin to show the finite sample performance of the proposed posterior statistics. Let r_{1t}, r_{2t} be two time series random variables specified as:

$$\begin{aligned} r_{1t} &= \mu_1 + \sigma_1 z_{1t}, \\ r_{2t} &= \mu_2 + \sigma_2 z_{2t}, \\ C(F(r_{1t}), F(r_{2t}); \delta) &= \Phi(r_{1t}, r_{2t}; \rho), \end{aligned}$$

where μ_i and σ_i are the mean and the standard deviation of r_{it} and $C(\cdot)$ is the copula function $C(\cdot)$ with the dependence parameter ρ. This model is known to be the same as a bivariate normal distribution with the correlation coefficient being ρ. To simulate the data reasonably, we set $\mu^* = (0.02, 0.03)'$ and the true variance covariance matrix as:

$$\Phi^* = \begin{pmatrix} 2 & 2\sqrt{2}\rho^* \\ 2\sqrt{2}\rho^* & 4 \end{pmatrix} \tag{10.5.1}$$

where ρ^* is the true value of ρ. We thus test the following hypothesis:

$$H_0 : \rho = 0 \quad \text{vs} \quad H_1 : \rho \neq 0.$$

In this model, we have parameters $\boldsymbol{\vartheta} = (\mu_1, \mu_2, h_1, h_2, \rho)'$, where $(\mu_1, \mu_2)'$ and $(h_1, h_2)'$ are means and precision for the two random normally distributed variables, respectively, and ρ is of primary interest. The prior distributions for the parameters are set as:

$$\begin{aligned} &\mu_1 \sim N(\mu_{10}, \sigma_{10}^2), \quad \mu_2 \sim N(\mu_{20}, \sigma_{20}^2), \\ &h_1 \sim Gamma(s_{10}, r_{10}), \quad h_2 \sim Gamma(s_{20}, r_{20}), \\ &\rho \sim TN(\mu_\rho, \sigma_\rho^2, lb, ub), \end{aligned}$$

where $TN(\cdot)$ is the truncated normal distribution.

The uninformative prior distributions (NP) and informative prior distributions (IP) of the parameters of interest are set as:

$$\begin{aligned} &NP : (\mu_{10}, \mu_{20}, \sigma_{10}^2, \sigma_{20}^2, s_{10}, s_{20}, r_{10}, r_{20}, \mu_\rho, \sigma_\rho^2, lb, ub) \\ &= (\mu_1^*, \mu_2^*, 10000, 10000, 0.01, 0.01, 0.02, 0.04, \rho^*, 10000, -1, 1) \end{aligned}$$

$$\begin{aligned} &IP : (\mu_{10}, \mu_{20}, \sigma_{10}^2, \sigma_{20}^2, s_{10}, s_{20}, r_{10}, r_{20}, \mu_\rho, \sigma_\rho^2, lb, ub) \\ &= (\mu_1^*, \mu_2^*, 10000, 10000, 0.01, 0.01, 0.02, 0.04, \rho^*, 0.0009, -1, 1) \end{aligned}$$

To perform a simulation study, we consider three different ρ^* ($\rho^* = 0.0, 0.1, 0.2$) and three different sample sizes $n = 100, 500, 5000$. For each case, we repeat the data generation process 1,000 times independently and report the rejection rate of H_0 based on different statistics at the 5% significance level across the 1,000 replications. The empirical size and power of the proposed statistics are reported in Table 3 (under uninformative priors) and Table 4 (under informative priors).

The results in Table 3 exhibit a similar pattern as those in Table 1. For the empirical size when the sample size is small (e.g., $n = 100$), marked disagreement among the empirical size of different statistics appears, and some are far away from the nominal size 5% (e.g., the empirical size of $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$ is as large as 13.6%, while the empirical size of $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ is as small as 2.7%). However, as the sample size increases, the empirical size of different statistics reported in this study converge towards each other, being approximately the nominal size of 5%. Regarding

Table 10.3: Empirical size and power of posterior statistics in a Gaussian copula model (NP)

	Empirical Size ($\rho^* = 0.0$)		
	$n = 100$	$n = 500$	$n = 5000$
LR	5.5%	6.0%	5.7%
$Wald$	6.3%	6.6%	5.7%
LM	5.5%	6.4%	5.7%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	5.9%	6.4%	5.7%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	5.8%	6.1%	5.5%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	13.6%	9.4%	6.1%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	2.7%	5.4%	5.1%
	Empirical Power ($\rho^* = 0.1$)		
	$n = 100$	$n = 500$	$n = 5000$
LR	15.8%	58.0%	100.0%
$Wald$	17.0%	58.6%	100.0%
LM	16.1%	58.1%	100.0%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	19.5%	58.9%	100.0%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	15.9%	57.4%	100.0%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	27.2%	59.2%	100.0%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	7.0%	53.0%	100.0%
	Empirical Power ($\rho^* = 0.2$)		
	$n = 100$	$n = 500$	$n = 5000$
LR	55.1%	99.4%	100.0%
$Wald$	57.4%	99.4%	100.0%
LM	56.3%	99.4%	100.0%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	55.5%	99.4%	100.0%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	53.5%	99.3%	100.0%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	59.9%	99.3%	100.0%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	34.8%	99.0%	100.0%

Table 10.4: Empirical size and power of posterior statistics in a Gaussian copula model (IP)

	Empirical Size ($\rho^* = 0.0$)			
	$n = 100$	$n = 500$	$n = 5000$	$n = 50000$
LR	5.5%	6.0%	5.7%	5.0%
$Wald$	6.3%	6.6%	5.7%	5.0%
LM	5.5%	6.4%	5.7%	5.0%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	0.1%	0.1%	3.0%	4.7%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	0.1%	0.1%	3.3%	4.8%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	0.0%	0.1%	4.0%	5.8%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	0.0%	0.2%	2.8%	4.5%
	Empirical Power ($\rho^* = 0.1$)			
	$n = 100$	$n = 500$	$n = 5000$	$n = 50000$
LR	15.8%	58.0%	100.0%	100.0%
$Wald$	17.0%	58.6%	100.0%	100.0%
LM	16.1%	58.1%	100.0%	100.0%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	100.0%	99.9%	100.0%	100.0%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	0.0%	9.1%	100.0%	100.0%
	Empirical Power ($\rho^* = 0.2$)			
	$n = 100$	$n = 500$	$n = 5000$	$n = 50000$
LR	55.1%	99.4%	100.0%	100.0%
$Wald$	57.4%	99.4%	100.0%	100.0%
LM	56.3%	99.4%	100.0%	100.0%
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$	100.0%	100.0%	100.0%	100.0%
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	0.0%	76.5%	100.0%	100.0%

the empirical power, as the sample size increases, and as ρ^* deviates from the hypothesised value under the null hypothesis, the empirical powers of all statistics in this study approach 100%.

Table 4 shows the empirical size and power of these statistics in a Gaussian copula model under informative prior distributions (IP), which exhibit similar patterns to that in Table 2. The performance of the frequentist $LR, Wald, LM$ statistics remain the same as those under uninformative prior distributions (NP) reported in Table 3. For the posterior statistics, the empirical size and power are strongly affected by strong prior information, particularly under a relatively small sample size. When H_0 is true, the empirical size is distorted and nears 0.0% when $n = 100, 500$. As the sample size increases up to 50,000, the empirical sizes of all the posterior statistics increase to near the nominal size of 5.0%. When H_1 is true, the empirical powers of $\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + q_\theta$, $\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$ and $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$ again increase to 100.0%, even with $n = 100$. However, for $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$, the empirical power first decreases to 0.0% ($n = 100$), and then gradually increases to 100.0% as the sample size increases.

10.6 Empirical Illustrations

This section applies the proposed test statistics to two popular examples in economics and finance. The first example contains asset pricing models with a t error distributions. The likelihood functions of these models have an analytical form and can also be rewritten in a latent variable form. The second example is a Gaussian copula model with Gaussian margins.

10.6.1 Hypothesis testing for asset pricing models with multivariate t distribution

Asset pricing models are important models in modern finance and generally assume that the return distribution is normal. Unfortunately, there has been overwhelming empirical evidence against normality for asset returns, which have led researchers to investigate asset pricing models with heavy-tailed distributions. Zhou (1993) and Kan and Zhou (2017) suggested using the multivariate t distribution to replace the multivariate normal distribution. Based on the efficient market theory, the asset excess premium should not be statistically different from zero. Finally, the multivariate t distribution can be rewritten in scale-mixture form to become a latent variable model. Thus, based on Zhang et al. (2019), we consider the following asset pricing model, which has two equivalent representations:

$$R_t = \boldsymbol{\alpha} + \boldsymbol{\beta}' \boldsymbol{F}_t + \varepsilon_t, \varepsilon_t \sim t[\mathbf{0}, \boldsymbol{\Sigma}, \nu];$$
$$R_t = \boldsymbol{\alpha} + \boldsymbol{\beta}' \boldsymbol{F}_t + \varepsilon_t, \varepsilon_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}/\omega_t), \omega_t \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right),$$

where R_t is the excess return of portfolio at period t with $N \times 1$ dimension, $\boldsymbol{F}_t$ a $K \times 1$ vector of factor portfolio excess returns, $\boldsymbol{\alpha}$ a $N \times 1$ vector of intercepts, $\boldsymbol{\beta}$ a $N \times K$ vector of scaled covariances, ε_t the random error, and $t = 1, 2, \cdots, n$. For convenience, we restrict $\boldsymbol{\Sigma}$ to be a diagonal matrix with diagonal elements $\sigma^2_{ii}, i = 1, 2, \cdots, N$, and ν to be a known constant as $\nu = 3$. The scale-mixture representation of t distribution is used to write the "model.txt" file that is passed to WinBUGS to generate MCMC outputs from the posterior distribution of parameters.

The data used in this study are monthly returns of 25 portfolios that are constructed at the end of each month as the intersections of 5 portfolios formed on size (market equity, ME) and 5 portfolios formed on the ratio of book equity to market equity (BE/ME). The Fama/French's five factors, market excess return, SMB (Small Minus Big), HML (High Minus Low), RMW(Robust Minus Weak), CMA(Conservative Minus Aggressive) are used as the explanatory factors (Fama and French, 2015). The sample period is from July 1963 to July 2021; thus, $N = 25$, $K = 5$, $n = 697$. These data are freely available from the data library of Kenneth French.[2]

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

Table 10.5: Asset pricing testing under a multivariate t distribution

hypothesis	$\boldsymbol{\alpha} = 0 \times \mathbf{1}_N$
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + N$	72.23
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	72.45
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - N$	71.59
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	71.61

Making inferences for asset pricing models has attracted considerable attention in the empirical asset pricing literature. Avramov and Zhou (2010) provided an excellent review of the literature on Bayesian portfolio analysis. For Bayesian inference, we must specify the prior distributions for parameters. In this study, to represent the prior ignorance, we assign some vague conjugate prior distributions:

$$\alpha_i \sim N[0, 100], \beta_{ij} \sim N[0, 100], \sigma^2_{ii} \sim IG[0.01, 0.0001].$$

In the R language, we use R2WinBUGS to obtain the MCMC outputs and draw 100,000 random observations from the posterior distributions in each model where the first 40,000 is used as the burn-in sample, and the next 60,000 iterations is collected with every 3rd observation as an effective observation. Thus, 20,000 effective observations are considered.

In asset pricing theory, the efficient market theory suggests that the excess premium $\boldsymbol{\alpha}$ should be zero. Thus, we can write this problem as a hypothesis to be tested as:

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

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

In section 3, we have shown that the threshold values from Bernardo and Rueda (2002) and Li and Yu (2012) are difficult to calibrate. Thus, in this study, we only consider the statistics developed by Li et al. (2015), Liu et al. (2021) and Li et al. (2022). Based on 20,000 MCMC samples, we calculate the four posterior test statistics, $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$, $\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0)$, $\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$, and $\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$. For more details about computing these test statistics, see Li et al. (2015, 2022) and Liu et al. (2021). We report the results in Table 5.

From these results, according to the critical values (37.65) from $\chi^2(25)$ under the 5% significance level, all the test statistics reject the null hypothesis. Thus, we can conclude that the mean-variance efficiency does not hold in practice.

10.6.2 Hypothesis testing for copula models

In this subsection, we use the Gaussian copula model with a Gaussian margin to fit real data and perform hypothesis testing based on the proposed posterior statistics. Recalling the simulation study in section 5.2, the model is specified as:

$$\begin{aligned} r_{1t} &= \mu_1 + \sigma_1 z_{1t}, \\ r_{2t} &= \mu_2 + \sigma_2 z_{2t}, \\ C(F(r_{1t}), F(r_{2t}); \delta) &= \Phi(r_{1t}, r_{2t}; \rho), \end{aligned}$$

where r_{1t} and r_{2t} are now daily log returns on the S&P 100 and S&P 600 indices at time t, μ_i, σ_i are mean and standard deviation of r_{it}, respectively, $i = 1, 2$. $\Phi(\cdot)$ is the cumulative density function of bivariate normal distribution. The log likelihood function at time t is:

$$\ln L_t = -\ln 2\pi - \frac{1}{2} \ln \left(\frac{1-\rho^2}{h_1 h_2} \right) - \frac{z^2_{1t} + z^2_{2t} - 2\rho z_{1t} z_{2t}}{2(1-\rho^2)},$$

Table 10.6: Posterior mean and standard error of parameters in a Gaussian copula model

Parameters	μ_1	h_1	μ_2	h_2	ρ
Posterior Means	0.0361	0.6765	0.0475	0.4938	0.8354
SE	0.0158	0.0116	0.0184	0.0087	0.0036

Table 10.7: Dependence testing based on a Gaussian copula model

hypothesis	$\rho = 0$
$\mathbf{T}^1_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0) + 1$	7333.372
$\mathbf{T}^2_{LYWZ}(\mathbf{y}, \boldsymbol{\theta}_0)$	7333.244
$\mathbf{T}_{LLYZ}(\mathbf{y}, \boldsymbol{\theta}_0) - 1$	53074.090
$\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	344.023

where $h_i = \frac{1}{\sigma_i^2}$ is the precision parameter, and $z_{it} = (r_{it} - \mu_i)h_i^{1/2}$. The required parameters are $\boldsymbol{\vartheta} = (\mu_1, h_1, \mu_2, h_2, \rho)'$; the first four are parameters related to the normal marginal distributions, and the last one is for the copula function.

The data used in this study are daily percentage returns on the S&P 100 and S&P 600 Indices from 17 August 1995 to 22 September 2021; thus, $T = 6122$. We are interested in the dependence between the two series of index returns. Based on the Gaussian copula model with a Gaussian margin, the hypothesis testing problem can be formulated as:

$$H_0 : \rho = 0, H_1 : \rho \neq 0.$$

We assign the following prior distributions on parameters:

$$\begin{aligned} \mu_i &\sim N(0, 25), \quad i = 1, 2, \\ h_i &\sim Gamma(0.1, 1), \quad i = 1, 2, \\ \rho &\sim TN[0, 100, -1, 1]. \end{aligned}$$

We iterate 100,000 times starting at the initial value $\theta_0 = (0, 1, 0, 1, 0)'$, and burn in the first 40,000 of the chain. For the remaining 60,000 observations, we take one of every 3 observations and finally obtain an effective MCMC output of sample size 20,000 for each parameter. Based on this MCMC output, we report the parameter estimation results based on the MCMC output in Table 6 and the values of the proposed posterior statistics in Table 7.

According to the posterior statistics shown in Table 7, all four posterior statistics strongly reject the null hypothesis that the two series of index returns are uncorrelated at the 5% significance level. This result is consistent with the expectations that the two indices (SP100, SP600) returns are naturally correlated with each other by construction. The estimation result for ρ in Table 6, which is 0.8354 with a relatively small standard error (0.0036), also provides strong evidence against the null hypothesis that assumes $\rho = 0$.

10.7 Discussion and future research

In this chapter, we review several statistics for hypothesis testing, which can be regarded as the Bayesian version of the "trinity" of test statistics widely used in the frequentist domain, the LR test, the LM test and the Wald test. Their asymptotic distributions are discussed based on a set of regular conditions. We show that these approaches have good theoretical properties and do not require tedious additional computations. We also demonstrate the methods using econometric models with simulation studies and real examples.

Bibliography

Avramov, D. and G. Zhou (2010). Bayesian portfolio analysis. *Annu. Rev. Financ. Econ. 2*(1), 25–47.

Bartlett, M. (1957). A comment on D. V. lindley's statistical paradox. *Biometrika 44*, 533–534.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2 ed.). Springer Science & Business Media.

Berger, J. O. and L. R. Pericchi (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association 91*(433), 109–122.

Bernardo, J. M. and R. Rueda (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review 70*(3), 351–372.

Bernardo, J. M. and A. F. Smith (2009). *Bayesian theory* (2 ed.), Volume 405. Chichester: John Wiley & Sons Canada, Limited.

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

Chen, X., T. M. Christensen, and E. Tamer (2018). Monte Carlo confidence sets for identified sets. *Econometrica 86*(6), 1965–2018.

Chernozhukov, V. and H. Hong (2003). An MCMC approach to classical estimation. *Journal of Econometrics 115*(2), 293–346.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association 90*(432), 1313–1321.

Chopin, N. and O. Papaspiliopoulos (2020). *An Introduction to Sequantial Monte Carlo*. Springer Series in Statistics.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological) 39*(1), 1–22.

Engle, R. F. (1984). Wald, likelihood ratio, and lagrange multiplier tests in econometrics. *Handbook of Econometrics 2*, 775–826.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Friel, N. and A. N. Pettitt (2008). Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(3), 589–607.

Gallant, A. R. and H. White (1988). *A unified theory of estimation and inference for nonlinear dynamic models*. Blackwell.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2 ed.). Taylor & Francis, UK.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statistical Science 7*, 473–511.

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

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

Kan, R. and G. Zhou (2017). Modelling non-normality using multivariate t: implications for asset pricing. *China Finance Review International 7*(1), 2–32.

Kass, R. E. and A. E. Raftery (1995). Bayes factor. *Journal of the Americana Statistical Association 90*(430), 773–795.

Kline, B. and E. Tamer (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics 7*(2), 329–366.

Li, Y., X. B. Liu, and J. Yu (2015). A bayesian chi-squared test for hypothesis testing. *Journal of Econometrics 189*(1), 54–69.

Li, Y., N. L. Wang, and J. Yu (2021). Improved marginal likelihood estimation via power posteriors and importance sampling. *Journal of Econometrics,forthcoming*.

Li, Y., N. L. Wang, J. Yu, and Y. Zhang (2022). Hypothesis Testing Using Posterior-test-based Bayes Factors. *Working paper*.

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

Li, Y., J. Yu, and T. Zeng (2017). Deviation information criterion for bayesian model comparison: justification and variation. *Working paper*.

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

Liao, Y. and A. Simoni (2019). Bayesian inference for partially identified smooth convex models. *Journal of Econometrics 211*(2), 338–360.

Lindley, D. V. (1957). A statistical paradox. *Biometrika 44*(1/2), 187–192.

Liu, X., Y. Li, J. Yu, and T. Zeng (2021). Posterior-based Wald-type statistics for hypothesis testing. *Journal of Econometrics 230*(1), 83 – 113.

Moon, H. R. and F. Schorfheide (2012). Bayesian and frequentist inference in partially identified models. *Econometrica 80*(2), 755–782.

O'Hagan, A. (1991). Discussion on posterior bayes factors (by m. aitkin). *Journal of Royal Statistic Society B 53*, 136.

O'Hagan, A. (1995). Fractional bayes factors for model comparison. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(1), 99–118.

Pishro-Nik, H. (2016). *Introduction to probability, statistics, and random processes*. Kappa Research, LLC.

Poirier, D. J. (1995). *Intermediate statistics and econometrics: a comparative approach*. The MIT Press.

Robert, C. P. (1993). A note on jeffreys-lindley paradox. *Statistica Sinica 3*(2), 601–608.

Robert, C. P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*, Volume 2. Springer.

Skaug, H. J. and J. Yu (2014). A flexible and automated likelihood based framework for inference in stochastic volatility models. *Computational Statistics and Data Analysis 76*, 642–654.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology) 64*(4), 583–639.

Wakefield, J. (2013). *Bayesian and frequentist regression Methods*, Volume 23. Springer Science & Business Media.

White, H. (1987). *Specification testing in dynamic models*, Volume 1. Cambridge University Press New York, NY, USA.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. *Handbook of econometrics 4*, 2639–2738.

Zhang, H., N. Wang, Y. Li, and Y. Zhan (2019). Bayesian asset pricing testing under multivariate t-distribution. *Applied Economics Letters 26*(11), 898–901.

Zhou, G. (1993). Asset-pricing tests under alternative distributions. *The Journal of Finance 48*(5), 1927–1942.