Hypothesis Testing via Posterior-Test-Based Bayes Factors*

Yong Li
Renmin University of China

Nianling Wang
Capital University of Economics and Business

Jun Yu
Singapore Management University

Yonghui Zhang
Renmin University of China

March 13, 2023

Abstract

Hypothesis testing via p-value has been criticized in recent years. Bayes factors (BFs) have been tipped as a possible replacement of p-value for hypothesis testing. However, the standard BFs suffer from some theoretical and practical difficulties. For example, they are not well defined under improper priors and are subject to Jeffreys-Lindley-Bartlett's paradox under vague priors. Moreover, they are difficult to compute for many models. In this paper, we propose to compare sampling distributions of the posterior-test-based statistics for hypothesis testing. Two posterior-test-based BFs are constructed from the posterior version of the likelihood ratio test and the Wald test, respectively. Under regularity conditions, we show that the new methods can avoid the p-hacking problem and the problems in the standard BFs. The advantages of the proposed methods are investigated using several simulation and empirical studies.

JEL classification: C11, C12
Keywords: Bayes factor, Consistency, p-value, p-hacking, Posterior likelihood ratio test, Posterior Wald test

*Yong Li & Yonghui Zhang, School of Economics, Renmin University of China, Beijing, 200872, China. Nianling Wang, School of Finance, Capital University of Economics and Business, Beijing, 100070, China. Jun Yu, School of Economics, Singapore Management University, 90 Stamford Road, Singapore.

1 Introduction

Hypothesis testing is ubiquitous in empirical research in many fields in sciences and social sciences. In the frequentist paradigm, the p-value of a test statistic (e.g. the likelihood ratio test, the Wald test, and the Lagrange multiplier test) under the null hypothesis (defined as H_0) is the most popular indicator of the statistical significance of H_0. It represents the probability of observing an outcome or a more extreme outcome when H_0 is true. When the p-value is small enough (smaller than a pre-determined level α, say 5%), we have $(1-\alpha)\times 100\%$ confidence to reject H_0. Typically, H_0 corresponds to no effect or simplification of a larger model specified in the alternative hypothesis (defined as H_1). Not surprisingly, in a typical case, empirical researchers look for evidence against H_0 so that they can claim a statistically significant effect.

In recent years, hypothesis testing based on the p-value has been criticized in many fields. Many researchers, including prominent scientists and statisticians, claim that we should abandon the p-value for hypothesis testing; see, for example, Amrhein et al. (2019) and Wasserstein et al. (2019). More than 800 researchers have added their names as signatories to support the movement against the use of the p-value (Amrhein et al., 2019).

There are several complaints about the p-value. First, the p-value does not represent the probability of H_0 being true. Second, the p-value does not work under a large sample size unless the null hypothesis is exactly true (Berkson, 1938). This is because, as the sample size increases, the sampling distribution (and the asymptotic distribution) of any reasonable test statistic becomes more concentrated. Thus, in general, the p-value can be made arbitrarily small by increasing the sample size. This concern becomes more relevant in the big-data era. Third, p-value-based testing is asymmetric, that is, H_0 and H_1 are not treated equally by the p-value. Fourth, the p-value does not measure the size of an effect. Different observed effects can have the same p-value (Goodman, 2008).

Last but not least, the use of the p-value in academic research causes the so-called "p-hacking" problem, which is associated with *publication bias* in the scientific literature and usually occurs when researchers select and manipulate data and statistical analyses until some statistically significant evidence is found against H_0; see Andrews and Kasy (2019) and Abadie (2020). The researchers are often motivated to do "p-hacking" since most journals prefer publishing papers with significant results. This problem of p-hacking has been investigated extensively by many researchers such as Head et al. (2015) in science, Baker (2015) in psychology, Kim et al. (2018) in accounting, Kim and Ji (2015) and Harvey (2017) in finance, and Brodeur et al. (2018) in economics. Recently, Elliott et al. (2022) propose several tests for p-hacking based on the distribution of p-values across multiple studies. The recommendation from these studies is unanimous: one should be careful with using the p-value and avoid doing p-hacking.

In the literature, it is well known that the Bayes factor (BF) of Kass and Raftery (1995) can be an effective alternative to the p-value for hypothesis testing (Marden, 2000). Unlike the p-value whose interpretation is rooted in the sampling distribution or the asymptotic distribution of the proposed test statistic, the BF compares the posterior probabilities of alternative model specifications. It is well documented that the BF enjoys the consistency property. That is, $\text{Prob}(H_0|\mathbf{y}) \to 1$ under H_0 and $\text{Prob}(H_1|\mathbf{y}) \to 1$ under H_1 as the sample size $n \to \infty$, where $\mathbf{y} = (y_1, \ldots, y_n)'$ denote the observed data. This is the reason why the BF does not suffer from the p-hacking problem.

Unfortunately, the BF is not trouble-free as it suffers from some theoretical and computational difficulties. First, the BF is not well defined under improper priors. Second, the BF is subject to Jeffreys-Lindley-Bartlett's (JLB) paradox when proper but vague priors are used. That is, the BF tends to favor H_0 when a vague prior is used for parameters in H_0; see Kass and Raftery (1995). Third, the calculation of the BF requires evaluation of two marginal likelihoods, $p(\mathbf{y}|H_0)$ and $p(\mathbf{y}|H_1)$. In many cases, marginal likelihoods involve high-dimensional integrations that may be numerically challenging. Although some interesting approaches have been proposed to calculate the BF from posterior outputs, such as those in Chib (1995), Friel and Pettitt (2008), and Li et al. (2023), the BF remains challenging to calculate, especially in the big data environment.

In this paper, we propose to combine the strengths of frequentist-based test statistics and those of the BF to design new statistics for hypothesis testing. Our idea is related to that of Johnson (2005, 2008) who propose to compare the sampling distributions of a frequentist-based test statistic under H_0 and H_1. In particular, instead of comparing the marginal likelihoods of H_0 and H_1, Johnson (2005, 2008) suggest comparing the pivotal asymptotic distribution under H_0 with that under a local alternative hypothesis evaluated at a frequentist test statistic. For example, when the frequentist test is the likelihood ratio (LR) statistic, we just compare a central χ^2 distribution with a non-central χ^2 distribution, both evaluated at the LR statistics. This BF-like approach is based on modelling frequentist test statistics.

As documented in Johnson (2005, 2008), the frequentist-test-based BFs share the strengths in the frequentist test statistic as well as those in the standard BF. First, they are not subject to the p-hacking problem because they have the consistency property. Second, they are free from the JLB paradox as it is independent of priors.

However, when the prior information is available and important, the frequentist-test-based BF cannot use it to improve statistical inferences. Some prominent researchers believe that, in some cases, it is important to use priors to reflect their belief about the validity of underlying theory (An and Schorfheide, 2007). Moreover, the calculation of the frequentist-test-based BF requires one to obtain optimization-based frequentist estimators of parameters (such as the maximum likelihood (ML) estimator (MLE) in the case of the LR test). In many models such as latent variable models, optimization-based frequentist estimators are difficult to obtain. On the other hand, Bayesian estimation based on posterior sampling (such as MCMC) has become a powerful alternative to frequentist estimation. Hence, it is useful to extend the idea of the frequentist-test-based BFs to construct posterior-test-based BFs.

This paper proposes new posterior-test-based BFs. The first is constructed from the posterior LR (PLR) statistic that modifies the PLR test of Li et al. (2014). The second is constructed from the posterior Wald statistic (PWald) of Liu et al. (2022). The proposed posterior-test-based BFs enjoy several good statistical and numerical properties. First, they are well defined under improper priors. Second, they can avoid the JLB paradox. Third, they can incorporate the prior information when it is available. Fourth, they are based on posterior outputs, and hence, can be easier to compute, compared with Johnson (2005, 2008) when optimization-based estimation is difficult but posterior sampling is easy. Last but not least, they enjoy the consistency property, and hence, can avoid the p-hacking problem.

The rest of the paper is organized as follows. Section 2 reviews the p-value, the standard BF, and the frequentist-test-based BFs for hypothesis testing. Section 3 proposes the posterior-test-based BFs based on two posterior statistics and obtains the consistency property. Section 4 checks the finite sample performance and the advantages of the posterior-test-based BFs. Section 5 explains the advantages of the posterior-test-based BFs via three empirical studies. Section 6 concludes. The appendix collects proofs of the propositions and the theorems in the paper while the Online Supplement collects proofs of the two lemmas.

2 A Literature Review

Assume that the data $\mathbf{y}$ is fitted by a correctly specified probability model $M \equiv \{p(\mathbf{y}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q\}$, where Θ is the parameter space. Write $\boldsymbol{\theta} = (\boldsymbol{\vartheta}', \boldsymbol{\psi}')'$, where $\boldsymbol{\vartheta} \in \Theta_\vartheta \subset \mathbb{R}^{q_\vartheta}$ is a vector of parameters of interest and $\boldsymbol{\psi} \in \Theta_\psi \subset \mathbb{R}^{q_\psi}$ $(q_\vartheta + q_\psi = q)$ collects nuisance parameters. Consider the following hypothesis testing problem:

$$H_0 : \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 \text{ vs } H_1 : \boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0. \tag{2.1}$$

In the literature there are mainly two classes of approaches to hypothesis testing, frequentist approaches and Bayesian approaches. In the frequentist paradigm, a test statistic is normally introduced. Based on the asymptotic distribution of the test statistic under H_0, the p-value is computed so that hypothesis testing can be done by comparing the p-value with a subjective significance level such as 5%. The p-value is generally denoted as either $p = \text{Prob}\left(T \geqslant t_n(\mathbf{y})|H_0\right)$ or $p = \text{Prob}\left(T \leqslant t_n(\mathbf{y})|H_0\right)$ for one-side testing or $p = \text{Prob}\left(|T| \geqslant t_n(\mathbf{y})|H_0\right)$ for two-side testing, where $t_n(\mathbf{y})$ is a test statistic that depends on a frequentist estimator of $\boldsymbol{\theta}$ and T is generally the asymptotic distribution (or the finite sample distribution in rare cases) of $t_n(\mathbf{y})$ under H_0. The probabilities of Type I error and Type II error can be expressed as Prob(Reject $H_0|H_0$) and Prob(Not reject $H_0|H_1$), respectively.

A typical justification of p-value is that it is the probability of observing an outcome or a more extreme outcome when H_0 is true, and hence, can be viewed as a measure of the "strength of evidence" against H_0. The smaller the p-value, the more significant the statistical evidence against H_0. However, the p-value is not the probability of H_0 being true. For more discussions on the p-value, one can refer to Marden (2000).

A serious criticism about its usage is the common practice of "p-hacking" in the scientific literature to search for statistical significance against H_0. Simmons et al. (2014) and Simonsohn et al. (2011) show that "p-hacking" can increase the probability that a study examining a non-existent effect "works" from the nominal 5% to well above 50%. In practice, conventional significance levels (such as 1%, 5%, 10%) are exclusively and arbitrarily used with little consideration of contexts, including the sample size, the power of the test, and the expected loss. Furthermore, there is strong evidence of publication bias in favor of statistically significant results. This is not surprising, the researchers are often motivated to report only statistically significant results without disclosing flexibility in data collection and/or multiple testing because top journals want to publish papers with positive results.

In response to these rising criticisms and concerns, to avoid further misinterpretation, misuse and large scale confusion, the American Statistical Association provides a formal statement to clarify several widely-agreed principles underlying the proper use and interpretation of the p-values; see Wasserstein and Lazar (2016).

Many researchers have suggested that the p-value should only be abandoned before an effective alternative is found; see, for example, Benjamin et al. (2018). Recently, researchers such as Harvey (2017) recommend a simple alternative — BF, which has long been a statistic used in the Bayesian paradigm, even before the p-value approach has been criticized.

BFs can overcome some difficulties of the p-value and enjoy many desirable properties. In the context of hypothesis testing specified in (2.1), BF is defined as the ratio of two marginal likelihoods:

$$\mathrm{BF}_{01} = \frac{p(\mathbf{y}|H_0)}{p(\mathbf{y}|H_1)} = \frac{\int p(\mathbf{y}|\boldsymbol{\vartheta}_0, \boldsymbol{\psi}, H_0) p(\boldsymbol{\psi}|H_0) d\boldsymbol{\psi}}{\int p(\mathbf{y}|\boldsymbol{\vartheta}, \boldsymbol{\psi}, H_1) p(\boldsymbol{\vartheta}, \boldsymbol{\psi}|H_1) d\boldsymbol{\vartheta} d\boldsymbol{\psi}},$$

where $p(\boldsymbol{\psi}|H_0)$ is the prior of $\boldsymbol{\psi}$ under H_0 and $p(\boldsymbol{\vartheta}, \boldsymbol{\psi}|H_1)$ is the prior of $(\boldsymbol{\vartheta}, \boldsymbol{\psi})$ under H_1. When the prior probabilities of two competing hypotheses are the same ($\mathrm{Prob}(H_0) = \mathrm{Prob}(H_1) = 0.5$), BF_{01} is the same as the posterior odds, $\mathrm{BF}_{01} = \mathrm{Prob}(H_0|\mathbf{y})/\mathrm{Prob}(H_1|\mathbf{y})$. In general, BF_{01} requires calculating the two marginal likelihoods, $\int p(\mathbf{y}|\boldsymbol{\vartheta}_0, \boldsymbol{\psi}, H_0) p(\boldsymbol{\psi}|H_0) d\boldsymbol{\psi}$ and $\int p(\mathbf{y}|\boldsymbol{\vartheta}, \boldsymbol{\psi}, H_1) p(\boldsymbol{\vartheta}, \boldsymbol{\psi}|H_1) d\boldsymbol{\vartheta} d\boldsymbol{\psi}$.

As explained in the Introduction, while a major advantage of BFs over the p-value is that it can avoid the p-hacking problem, it suffers from several problems. These problems motivate Johnson (2005, 2008) to introduce frequentist-test-based BFs for hypothesis testing. Instead of comparing the marginal likelihoods of H_0 and H_1 which are directly data dependent, the frequentist-test-based BFs of Johnson compare the asymptotic distribution of a frequentist test statistic under H_0 with that under a local alternative, both evaluated at the test statistic. Since the method is based on frequentist statistics, no prior information on model parameters is used. Moreover, frequentist estimation (such as ML) is required.

Take the well known LR statistic as an example. The LR test is defined as

$$\mathrm{LR} = 2\left[\ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}\right) - \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0\right)\right],$$

where $\widehat{\boldsymbol{\theta}}_0 = \left(\boldsymbol{\vartheta}_0', \widehat{\boldsymbol{\psi}}_0'\right)'$ and $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\vartheta}}', \widehat{\boldsymbol{\psi}}'\right)'$ are the constrained and unconstrained MLE of $\boldsymbol{\theta} = (\boldsymbol{\vartheta}', \boldsymbol{\psi}')'$ under H_0 and under H_1, respectively. Let $\boldsymbol{\theta}_n^*$ be the true value that minimizes the Kullback–Leibler (KL) loss between the data generating process (DGP) and the candidate model

$$\boldsymbol{\theta}_n^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \frac{1}{n} \int \ln \frac{p(\mathbf{y})}{p(\mathbf{y}|\boldsymbol{\theta})} p(\mathbf{y}) d\mathbf{y}, \tag{2.2}$$

where $p(\cdot)$ is the pdf of $\mathbf{y}$.[1] Assume regularity conditions hold. Under H_0, as $n \to \infty$, $\text{LR} \xrightarrow{d} \chi^2(q_\vartheta)\,(:= Z|H_0)$. Consider the following sequence of local alternatives for $\boldsymbol{\vartheta}$

$$H_{L1} : \boldsymbol{\vartheta}_n^* = \boldsymbol{\vartheta}_0 + \boldsymbol{\delta}/\sqrt{n}, \tag{2.3}$$

where $\boldsymbol{\delta} \neq 0$. Assume that the nuisance parameters $\boldsymbol{\psi}$ is fixed at $\boldsymbol{\psi}^*$, which is the true value. Thus, under H_{L1}, $\boldsymbol{\theta}_n^* = (\boldsymbol{\vartheta}_n^{*\prime}, \boldsymbol{\psi}^{*\prime})'$. Clearly, as $n \to \infty$, $\boldsymbol{\theta}_n^* \to \boldsymbol{\theta}_0^* := (\boldsymbol{\vartheta}_0', \boldsymbol{\psi}^{*\prime})'$. Davidson and MacKinnon (1987) show that, under H_{L1}, as $n \to \infty$,

$$\text{LR} \xrightarrow{d} \chi^2(q_\vartheta, \tau)\,(:= Z|H_{L1}) \text{ with } \tau = \boldsymbol{\delta}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\delta},$$

where $\chi^2(q_\vartheta, \tau)$ is a non-central χ^2 variate with the non-centrality parameter τ, $\boldsymbol{\Sigma}_{11}$ is the submatrix of $\boldsymbol{\Sigma}$ corresponding to $\boldsymbol{\vartheta}$, $\boldsymbol{\Sigma} := \mathbf{H}^{-1}$ is the inverse matrix of $\mathbf{H}$, the negative expected Hessian of the log likelihood defined by

$$\mathbf{H} = \lim_{n\to\infty} E\left[-\frac{1}{n}\frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=(\boldsymbol{\vartheta}_0', \boldsymbol{\psi}^{*\prime})'}\right].$$

It can be shown that

$$\text{LR} = \begin{cases} Z|H_0 + O_p(n^{-1/2}) & \text{under } H_0 \\ Z|H_{L1} + O_p(n^{-1/2}) & \text{under } H_{L1} \end{cases}. \tag{2.4}$$

Based on the LR statistic and its asymptotic distributions under H_0 and H_{L1}, Johnson (2005) proposes the following LR-test-based BF,

$$\text{BF}_{01}^J(\text{LR}) = \frac{p_{Z|H_0}(\text{LR})}{p_{Z|H_{L1}}(\text{LR})} = \frac{p_{Z|H_0}(\text{LR})}{\int p_{Z|H_{L1}}(\text{LR})p(\boldsymbol{\delta})d\boldsymbol{\delta}}.$$

Given the prior distribution of $\boldsymbol{\delta} \sim N[0, c\boldsymbol{\Sigma}_{11}]$ with $\boldsymbol{\Sigma}_{11}$ being defined above, it has

$$\text{BF}_{01}^J(\text{LR}) = (cn+1)^{\frac{q_\vartheta}{2}} \exp\left\{-\frac{\text{LR}}{2(cn+1)/(cn)}\right\}, \tag{2.5}$$

or

$$\ln \text{BF}_{01}^J(\text{LR}) = \frac{q_\vartheta}{2}\ln(cn+1) - \frac{\text{LR}}{2(cn+1)/(cn)}. \tag{2.6}$$

---

[1] Here we use the subscript n in $\boldsymbol{\theta}_n^*$ to indicate that the true parameters may change with the sample size such as under the Pitman local alternatives or when data are heterogeneous over time. Note that when data is stationary and under fixed alternatives, $\boldsymbol{\theta}_n^*$ should be $\boldsymbol{\theta}^*$.

Remark 2.1 *Under H_0, as $n \to \infty$, $LR = O_p(1)$, $\frac{cn}{cn+1} \to 1$, $\frac{q_\vartheta}{2}\ln(cn+1) \to \infty$. Hence, $\ln BF_{01}^J(LR) \to \infty$ and $BF_{01}^J(LR) \to \infty$, selecting H_0. Under H_1, Johnson (2008) shows that when $LR \sim O_p(n)$, $\ln BF_{01}^J(LR) \to -\infty$ and $BF_{01}^J(LR) \to 0$, selecting H_1. In this case, his LR-test-based BF has the consistency property. Moreover, for large n,*

$$\ln BF_{01}^J(LR) \approx -\frac{LR}{2} + \frac{q_\vartheta}{2}\ln n + \frac{q_\vartheta}{2}\ln c. \tag{2.7}$$

When $c = 1$, $\ln BF_{01}^J(LR)$ reduces to the well known BIC of Schwarz (1978). We will set $c = 1$ in the simulation and empirical studies.

Remark 2.2 *Since $BF_{01}^J(LR)$ is based on the LR statistic that requires MLE under H_0 and H_1, there is no need to specify prior distributions. Consequently, $BF_{01}^J(LR)$ is always well defined and can avoid the JLB paradox. However, this convenience comes with a cost. The first type of cost is that it cannot incorporate the prior information when it exists. The second type of cost is that for many complicated models such as latent variable models, MLE is generally difficult to obtain. In this case, Bayesian methods such as MCMC may be appealing to practitioners. These problems are the motivations for us to introduce posterior-test-based BFs, which can incorporate prior information and are based on posterior outputs.*

3 Posterior-test-based BFs

In this section, we propose two posterior-test-based BFs constructed from two posterior test statistics. Before we introduce them, we first briefly review the statistical decision framework for hypothesis testing in Section 3.1.[2] In Section 3.2, we give regularity conditions under which two posterior-test-based BFs are justified asymptotically. In Section 3.3, we propose the posterior LR (PLR)-test-based BF. In Section 3.4, we propose the posterior Wald (PWald)-test-based BF.

3.1 Hypothesis testing under the statistical decision framework

It is well known that hypothesis testing can be regarded as a statistical decision problem. For the hypothesis testing problem considered in (2.1), we can define two statistical decisions in the decision space, that is, not rejecting H_0 (name it d_0) or rejecting H_0 (name it d_1). We

[2]See Li (2023) for a more detailed review of the posterior hypothesis testing literature.

can assign a loss function to each decision denoted, respectively, by $\{\mathcal{L}(d_i, \boldsymbol{\vartheta}, \boldsymbol{\psi}), i = 0, 1\}$. Therefore, the net loss function is

$$\Delta\mathcal{L}(H_0, \boldsymbol{\vartheta}, \boldsymbol{\psi}) = \mathcal{L}(d_0, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - \mathcal{L}(d_1, \boldsymbol{\vartheta}, \boldsymbol{\psi}).$$

If the expected posterior loss of d_0 is larger than that of d_1, we then reject H_0. In other words, H_0 is rejected if and only if (iff)

$$E_{\boldsymbol{\theta}|\mathbf{y}}(\Delta\mathcal{L}(H_0, \boldsymbol{\vartheta}, \boldsymbol{\psi})) = \int_{\Theta} [\mathcal{L}(d_0, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - \mathcal{L}(d_1, \boldsymbol{\vartheta}, \boldsymbol{\psi})] \, p(\boldsymbol{\vartheta}, \boldsymbol{\psi}|\mathbf{y}) \, d\boldsymbol{\vartheta} d\boldsymbol{\psi} > 0, \tag{3.1}$$

where $p(\boldsymbol{\vartheta}, \boldsymbol{\psi}|\mathbf{y})$ is the posterior distribution under H_1. Naturally a posterior test statistic can be defined as

$$\mathrm{T}(\mathbf{y}, \boldsymbol{\vartheta}_0) = E_{\boldsymbol{\theta}|\mathbf{y}}(\Delta\mathcal{L}(H_0, \boldsymbol{\vartheta}, \boldsymbol{\psi})). \tag{3.2}$$

Remark 3.1 *BFs can be cast into this framework. If the loss functions are*

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

then the posterior test statistic is

$$T(\mathbf{y}, \boldsymbol{\vartheta}_0) = E_{\boldsymbol{\theta}|\mathbf{y}}(\triangle\mathcal{L}(H_0, \boldsymbol{\vartheta}, \ \boldsymbol{\psi})) > 0.$$

In this case, Bernardo and Rueda (2002) show that the decision is equivalent to

$$reject \ H_0 \ iff \ BF_{01} = \frac{\int p(\mathbf{y}|\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0, \boldsymbol{\psi}) \, p(\boldsymbol{\psi}|\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0) \, d\boldsymbol{\psi}}{\int\int p(\mathbf{y}|\boldsymbol{\vartheta}, \boldsymbol{\psi}) \, p(\boldsymbol{\psi}|\boldsymbol{\vartheta}) \, \pi(\boldsymbol{\vartheta}) \, d\boldsymbol{\vartheta} d\boldsymbol{\psi}} < 1, \tag{3.3}$$

where BF_{01} is the standard BF. That is, BFs can be regarded as a decision problem with a simple zero-one loss function when it is used for hypothesis testing. Bernardo and Rueda (2002) show that it is this zero-one loss that leads to the JLB paradox. Based on alternative but continuous loss functions, Bernardo and Rueda (2002), Li and Yu (2012), Li et al. (2014), Li et al. (2015), and Liu et al. (2022) have introduced different posterior statistics for hypothesis testing.

3.2 Regularity conditions

Let $\mathbf{y}^t$ denote $(y_0, y_1, \ldots, y_t)$ for any $0 \leq t \leq n$. Let $l_t(\mathbf{y}^t, \boldsymbol{\theta})$ or simply $l_t(\boldsymbol{\theta})$ denote $\ln p(\mathbf{y}^t|\boldsymbol{\theta}) - \ln p(\mathbf{y}^{t-1}|\boldsymbol{\theta})$ which is the log-likelihood for the t^{th} observation for any $1 \leq t \leq n$.

Thus, the log-likelihood function $\mathcal{L}_n(\boldsymbol{\theta})(:= \ln p(\mathbf{y}|\boldsymbol{\theta}, y_0))$ can be written as $\sum_{t=1}^{n} l_t(\boldsymbol{\theta})$. Let $l_t^{(j)}(\boldsymbol{\theta})$ denote the j^{th} derivative of $l_t(\boldsymbol{\theta})$ so that

$$l_t^{(0)}(\boldsymbol{\theta}) = l_t(\boldsymbol{\theta}),\ l_t^{(1)}(\boldsymbol{\theta}) = \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \text{ and } l_t^{(2)}(\boldsymbol{\theta}) = \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

Moreover, let

$$\begin{aligned}
\mathbf{s}(\mathbf{y}^t|\boldsymbol{\theta}) &:= \frac{\partial \ln p(\mathbf{y}^t|\boldsymbol{\theta})}{\partial \theta} = \sum_{i=1}^{t} l_i^{(1)}(\boldsymbol{\theta}),\ \mathbf{h}(\mathbf{y}^t|\boldsymbol{\theta}) := \frac{\partial^2 \ln p(\mathbf{y}^t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \sum_{i=1}^{t} l_i^{(2)}(\boldsymbol{\theta}), \\
\mathbf{s}_t(\boldsymbol{\theta}) &:= l_t^{(1)}(\boldsymbol{\theta}) = \mathbf{s}(\mathbf{y}^t|\boldsymbol{\theta}) - \mathbf{s}(\mathbf{y}^{t-1}|\boldsymbol{\theta}),\ \mathbf{h}_t(\boldsymbol{\theta}) := l_t^{(2)}(\boldsymbol{\theta}) = \mathbf{h}(\mathbf{y}^t|\boldsymbol{\theta}) - \mathbf{h}(\mathbf{y}^{t-1}|\boldsymbol{\theta}), \\
\bar{\mathbf{H}}_n(\boldsymbol{\theta}) &:= \frac{1}{n}\sum_{t=1}^{n} \mathbf{h}_t(\boldsymbol{\theta}),\ \bar{\mathbf{J}}_n(\boldsymbol{\theta}) := \frac{1}{n}\sum_{t=1}^{n} \left[\mathbf{s}_t(\boldsymbol{\theta}) - \bar{\mathbf{s}}_n(\boldsymbol{\theta})\right]\left[\mathbf{s}_t(\boldsymbol{\theta}) - \bar{\mathbf{s}}_n(\boldsymbol{\theta})\right]',\ \bar{\mathbf{s}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{s}_t(\boldsymbol{\theta}), \\
\mathbf{H}_n(\boldsymbol{\theta}) &:= \int \bar{\mathbf{H}}_n(\boldsymbol{\theta}) p(\mathbf{y})\, d\mathbf{y},\ \mathbf{J}_n(\boldsymbol{\theta}) := \int \bar{\mathbf{J}}_n(\boldsymbol{\theta}) p(\mathbf{y})\, d\mathbf{y},\ \mathbf{H}(\boldsymbol{\theta}) := \lim_{n\to\infty} \mathbf{H}_n(\boldsymbol{\theta}),\ \mathbf{J}(\boldsymbol{\theta}) := \lim_{n\to\infty} \mathbf{J}_n(\boldsymbol{\theta}),
\end{aligned}$$

where $\mathbf{H}_n(\boldsymbol{\theta})$ is the Hessian matrix, $\bar{\mathbf{H}}_n(\boldsymbol{\theta})$ the empirical Hessian matrix, $\mathbf{J}_n(\boldsymbol{\theta})$ the Fisher information matrix, and $\bar{\mathbf{J}}_n(\boldsymbol{\theta})$ the empirical Fisher information matrix. For the development of our proposed BFs, the following regularity conditions are imposed.

Assumption 1: $\boldsymbol{\theta} = (\boldsymbol{\vartheta}', \boldsymbol{\psi}')' \in \Theta$ that is a compact subset of $\mathbb{R}^q$.

Assumption 2: $\{y_t\}_{t=1}^{\infty}$ is α-mixing with size of $\alpha(m) = O\left(m^{\frac{-2r}{r-2}-\varepsilon}\right)$ for some $\varepsilon > 0$ and $r > 2$.

Assumption 3: For all t, $l_t(\boldsymbol{\theta})$ satisfies the standard measurability and continuity condition. Moreover, it is eight-times differentiable on $\mathcal{F}_{-\infty}^{t} \times \Theta$ where $\mathcal{F}_{-\infty}^{t}(:= \sigma(y_t, y_{t-1}, \cdots))$ is the σ-field generated by $(y_t, y_{t-1}, \ldots)$.

Assumption 4: For $j = 0, 1, 2$, for any $\boldsymbol{\theta}$, $\boldsymbol{\theta}' \in \Theta$,

$$\left\| l_t^{(j)}(\boldsymbol{\theta}) - l_t^{(j)}(\boldsymbol{\theta}') \right\| \le c_t^j\left(\mathbf{y}^t\right) \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|$$

in probability, where $c_t^j(\mathbf{y}^t)$ is a positive random variable with $\sup_t E\left\|c_t^j(\mathbf{y}^t)\right\| < \infty$ and $\frac{1}{n}\sum_{t=1}^{n}\left[c_t^j(\mathbf{y}^t) - E\left(c_t^j(\mathbf{y}^t)\right)\right] \xrightarrow{p} 0$.

Assumption 5: For $j = 0, 1, 2, 3$, $l_t^{(j)}(\boldsymbol{\theta})$ exists and there exists a function $M_t(\mathbf{y}^t)$ such that for any $\boldsymbol{\theta} \in \Theta$,

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\| l_t^{(j)}(\boldsymbol{\theta}) \right\| \leqslant M_t(\mathbf{y}^t) \text{ and} \sup_t E\left\|M_t(\mathbf{y}^t)\right\|^{r+\delta} \le M < \infty$$

for some $\delta > 0$, where r is the same as that in Assumption 2.

Assumption 6: $\left\{l_t^{(j)}(\boldsymbol{\theta})\right\}$ is L_2-near epoch dependent (NED) with respect to $\{\mathbf{y}_t\}$ of size -1 for $0 \leqslant j \leqslant 1$ and $-\frac{1}{2}$ for $j=2$ uniformly in Θ.

Assumption 7: For $\delta > 0$ and $N_0(\delta) \subseteq \Theta$, there exists $K(\delta) > 0$ such that

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}_n^*}\left(\sup_{\boldsymbol{\theta}\in\Theta\backslash N_0(\delta)} \frac{1}{n}\left[\mathcal{L}_n(\boldsymbol{\theta}) - \mathcal{L}_n(\boldsymbol{\theta}_n^*)\right] < -K(\delta)\right) = 1,$$

where $P_{\boldsymbol{\theta}_n^*}$ is the probability under $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*$ and $N_0(\delta)$ is an open ball of radius δ around $\boldsymbol{\theta}_n^*$.

Assumption 8: The sequence $\{\mathbf{H}_n(\boldsymbol{\theta}_n^*)\}$ are negative definite, uniformly in n.

Assumption 9: The prior density $p(\boldsymbol{\theta})$ is eight-times continuously differentiable with $p(\boldsymbol{\theta}_n^*) > 0$ and $\int \|\boldsymbol{\theta}\|^2 p(\boldsymbol{\theta})d\boldsymbol{\theta} < \infty$.

Assumption 10: Let $\boldsymbol{\psi}_n^{0*}$ be the quasi-true value that minimizes the KL loss between the DGP and the candidate model

$$\boldsymbol{\psi}_n^{0*} = \arg\min_{\boldsymbol{\psi}\in\Theta_\psi} \frac{1}{n}\int \ln \frac{p(\mathbf{y})}{p(\mathbf{y}|\boldsymbol{\vartheta}_0, \boldsymbol{\psi})} p(\mathbf{y})d\mathbf{y},$$

where Θ_ψ is the support space of $\boldsymbol{\psi}$, and $\left\{\boldsymbol{\psi}_n^{0*}\right\}$ is the sequence of minimizers interior to Θ_ψ uniformly in n, and $p(\cdot)$ is DGP of $\mathbf{y}$. Under H_1 (i.e. $\boldsymbol{\vartheta}_n^* \neq \boldsymbol{\vartheta}_0$), Assumptions 1-9 also hold for the misspecified model $p(\mathbf{y}|\boldsymbol{\vartheta}_0, \boldsymbol{\psi})$.

Remark 3.2 *Assumptions 1-8 are popular primitive conditions for establishing the ML theory, namely consistency and asymptotic normality, for dependent and heterogeneous data; see, for example, Gallant and White (1988) and Wooldridge (1994). For more discussions on these conditions, one can refer to Liu et al. (2022) and Li et al. (2020).*

Recall that $\boldsymbol{\theta}_n^* = (\boldsymbol{\vartheta}_n^{*\prime}, \boldsymbol{\psi}_n^{*\prime})'$ is the true parameters. Its limit is $\boldsymbol{\theta}^* = (\boldsymbol{\vartheta}^{*\prime}, \boldsymbol{\psi}^{*\prime})'$. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\vartheta}}', \widehat{\boldsymbol{\psi}}')'$ and $\widehat{\boldsymbol{\theta}}_0 = (\boldsymbol{\vartheta}_0', \widehat{\boldsymbol{\psi}}_0')'$ denote the unconstrained MLE and constrained MLE, respectively. The Bayesian estimator of $\boldsymbol{\theta}$ under H_1 and H_0 are given by $\overline{\boldsymbol{\theta}}(:= (\overline{\boldsymbol{\vartheta}}', \overline{\boldsymbol{\psi}}')' = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}, H_1)d\boldsymbol{\theta})$, and $\overline{\boldsymbol{\theta}}_0 = (\boldsymbol{\vartheta}_0', \overline{\boldsymbol{\psi}}_0')'$, respectively, where $\overline{\boldsymbol{\psi}}_0$ is the Bayesian estimator of $\boldsymbol{\psi}$ under H_0, defined by $\overline{\boldsymbol{\psi}}_0 = \int \boldsymbol{\psi} p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0, H_0)d\boldsymbol{\psi}$.

Remark 3.3 *Under Assumptions 1-10, Li et al. (2022) show that*

$$\begin{aligned}
\overline{\boldsymbol{\theta}} &= E\left[\boldsymbol{\theta}|\mathbf{y}, H_1\right] = \widehat{\boldsymbol{\theta}} + O_p(n^{-1}), \\
\mathbf{V}\left(\widehat{\boldsymbol{\theta}}\right) &= E\left[\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)' |\mathbf{y}, H_1\right] = -\frac{1}{n}\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\theta}}\right) + O_p(n^{-2}),
\end{aligned}$$

$$\overline{\boldsymbol{\psi}}_0 = \widehat{\boldsymbol{\psi}}_0 + O_p(n^{-1}),$$
$$\mathbf{V}_{\psi\psi}\left(\widehat{\boldsymbol{\theta}}_0\right) = E\left[\left(\boldsymbol{\psi} - \widehat{\boldsymbol{\psi}}_0\right)\left(\boldsymbol{\psi} - \widehat{\boldsymbol{\psi}}_0\right)' |\boldsymbol{\vartheta}_0, \mathbf{y}, H_0\right] = -\frac{1}{n}[\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\theta}}_0\right)]_{\psi\psi} + O_p(n^{-2}),$$

where $\left[\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\theta}}_0\right)\right]_{\psi\psi}$ *is the submatrix of* $\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\theta}}_0\right)$ *corresponding to* $\boldsymbol{\psi}$.

3.3 PLR-based BFs

Under the decision theoretical framework, based on the following net loss function that is defined as the difference of two log-likelihood functions under H_0 and H_1,

$$\triangle\mathcal{L}[H_0, (\boldsymbol{\vartheta}, \boldsymbol{\psi})] = 2\ln p(\mathbf{y}|\boldsymbol{\vartheta}, \boldsymbol{\psi}) - 2\ln p(\mathbf{y}|\boldsymbol{\vartheta}_0, \boldsymbol{\psi}), \tag{3.4}$$

Li et al. (2014) propose an LR-like posterior test statistic

$$\mathrm{T}_{LZY}(\mathbf{y}, \boldsymbol{\vartheta}_0) = 2\int \left[\ln p(\mathbf{y}|\boldsymbol{\vartheta}, \boldsymbol{\psi}) - \ln p(\mathbf{y}|\boldsymbol{\vartheta}_0, \boldsymbol{\psi})\right] p(\boldsymbol{\vartheta}, \boldsymbol{\psi}|\mathbf{y}) d\boldsymbol{\vartheta} d\boldsymbol{\psi}. \tag{3.5}$$

Unfortunately, $\mathrm{T}_{LZY}(\mathbf{y}, \boldsymbol{\vartheta}_0)$ is not asymptotically pivotal.

To obtain a pivotal asymptotic distribution, we introduce loss functions as

$$\mathcal{L}(d_0, \boldsymbol{\vartheta}, \boldsymbol{\psi}) = \begin{cases} c_0 & \text{if } \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 \\ c_0 + \left[2\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - \ln p(\mathbf{y}, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - D_c(\mathbf{y}, \boldsymbol{\vartheta}_0))\right] & \text{if } \boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0 \end{cases}, \tag{3.6}$$
$$\mathcal{L}(d_1, \boldsymbol{\vartheta}, \boldsymbol{\psi}) = \begin{cases} c_1 & \text{if } \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 \\ c_1 - \left[2\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - \ln p(\mathbf{y}, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)\right] & \text{if } \boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0 \end{cases}, \tag{3.7}$$

where $D_c(\mathbf{y}, \boldsymbol{\vartheta}_0) = \int \ln p(\mathbf{y}, \boldsymbol{\vartheta}_0, \boldsymbol{\psi}) p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi}$ is the Bayesian complete deviance function under H_0, $\widehat{\boldsymbol{\theta}}_m$ the posterior mode, $c_i (i = 0, 1)$ the cost of action d_i with $c = c_1 - c_0 > 0$. The net loss function is

$$\Delta\mathcal{L}\left(H_0, \boldsymbol{\vartheta}, \boldsymbol{\psi}\right) = 4\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - 2\ln p(\mathbf{y}, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - 2D_c(\mathbf{y}, \boldsymbol{\vartheta}_0) - c := m\left(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}, \boldsymbol{\psi}\right) - c, \tag{3.8}$$

where $m\left(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}, \boldsymbol{\psi}\right) = 4\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - 2\ln p(\mathbf{y}, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - 2D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)$ is a non-negative discrepancy measure between two statistical decisions, d_0 and d_1.

The posterior test statistic can be defined as:

$$\mathrm{T}(\mathbf{y}, \boldsymbol{\vartheta}_0) = \int \left[4\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - 2\ln p(\mathbf{y}, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - 2D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)\right] p(\boldsymbol{\vartheta}|\mathbf{y}) d\boldsymbol{\vartheta} d\boldsymbol{\psi}. \tag{3.9}$$

Note that for $\mathrm{T}(\mathbf{y}, \boldsymbol{\vartheta}_0)$ in (3.9), the posterior mode $\widehat{\boldsymbol{\theta}}_m$ is not easy to obtain in general. Hence, we can consider two alternative versions of PLR test statistics. Let $D_c(\mathbf{y})$ be the Bayesian complete deviance function under H_1 given as

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

Rewrite $\ln p(\mathbf{y}, \overline{\boldsymbol{\vartheta}}, \overline{\boldsymbol{\psi}}) = \ln p(\mathbf{y}|\overline{\boldsymbol{\vartheta}}, \overline{\boldsymbol{\psi}}) + \ln p(\overline{\boldsymbol{\vartheta}}, \overline{\boldsymbol{\psi}})$ and $\ln p(\mathbf{y}, \boldsymbol{\vartheta}_0, \overline{\boldsymbol{\psi}}_0) = \ln p(\mathbf{y}|\boldsymbol{\vartheta}_0, \overline{\boldsymbol{\psi}}_0) + \ln p(\boldsymbol{\vartheta}_0, \overline{\boldsymbol{\psi}}_0)$. Then, we define two PLR test statistics as:

$$\mathrm{PLR1} = 2(D_c(\mathbf{y}) - D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)) \text{ and } \mathrm{PLR2} = 2\left[\ln p(\mathbf{y}, \overline{\boldsymbol{\vartheta}}, \overline{\boldsymbol{\psi}}) - \ln p(\mathbf{y}, \boldsymbol{\vartheta}_0, \overline{\boldsymbol{\psi}}_0)\right]. \quad (3.10)$$

Remark 3.4 *In the literature, $D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)$ and $D_c(\mathbf{y})$ are generally referred to as the Bayesian deviances for measuring the Bayesian model fit. Hence, the posterior test statistics defined in (3.10) represent the difference between the two Bayesian deviances to measure the evidence against H_0. For more details about the Bayesian deviance for measuring model fit, see Spiegelhalter et al. (2002).*

We are now in the position to establish the large sample relationship among PLR1, PLR2, $\mathrm{T}(\mathbf{y}, \boldsymbol{\vartheta}_0)$ and LR. We then establish the large sample properties for PLR1 and PLR2, and introduce their corresponding test-based BFs.

Lemma 3.1 *Under Assumptions 1-10, we have*

$$\overline{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_m + O_p(n^{-1}) \text{ and } \overline{\boldsymbol{\psi}}_0 = \widehat{\boldsymbol{\psi}}_{m0} + O_p(n^{-1}),$$

where $\widehat{\boldsymbol{\theta}}_m$ and $\widehat{\boldsymbol{\psi}}_{m0}$ are the posterior mode of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ under H_0, respectively.

Proposition 3.1 *Suppose Assumptions 1-10 hold. Under both H_0 and H_1, we have*

$$T(\mathbf{y}, \boldsymbol{\vartheta}_0) = PLR1 + 2q + O_p(n^{-1}) \text{ and } PLR1 = PLR2 - q_{\vartheta} + O_p(n^{-1}). \quad (3.11)$$

Remark 3.5 *Proposition 3.1 establishes the large sample relationship among PLR1, PLR2 and $T(\mathbf{y}, \boldsymbol{\vartheta}_0)$. One can observe that PLR1 and PLR2 share the same size and power properties asymptotically. Compared with $T(\mathbf{y}, \boldsymbol{\vartheta}_0)$, they do not involve the posterior mode, and hence, are relatively easy to compute. In addition, compared with PLR1, PLR2 only involves the plug-in parameter estimator, and hence, is even easier to obtain.*

Lemma 3.2 *Suppose Assumptions 1-10 hold. Under H_0 and H_{L1}, we have*

$$PLR1 + q_{\vartheta} = LR + O_p(n^{-\frac{1}{2}}) \text{ and } PLR2 = LR + O_p(n^{-\frac{1}{2}}). \tag{3.12}$$

Remark 3.6 *Due to the results in the above lemma, PLR1 and PLR2 can be explained as the posterior version of the LR statistic.*

Remark 3.7 *The proposed PLR is related to some statistics proposed in the literature on Bayesian statistics. For example, Aitkin et al. (2005) introduce the following likelihood ratio*

$$LR_{ABC} = \frac{p(\mathbf{y}|\boldsymbol{\vartheta}_0, \boldsymbol{\psi})}{p(\mathbf{y}|\boldsymbol{\vartheta}, \boldsymbol{\psi})},$$

and then suggest evaluating the posterior probability

$$p\left(LR_{ABC} < k|\mathbf{y}\right)$$

for any pre-specified k, such as 1, 0.1 or 0.01. For $k = 1$, if $p\left(LR_{ABC} < 1|\mathbf{y}\right) > 1 - p$, where p is some small probability, then H_0 is rejected. However, in practice the choice of k and p is arbitrary.

Remark 3.8 *The proposed PLR is also related to some statistics proposed in the literature on Bayesian econometrics. For example, Chen et al. (2018) examine the asymptotic behavior of LR$= -2\left[\ln p\left(\mathbf{y}|\boldsymbol{\vartheta}^*, \boldsymbol{\psi}^*\right) - \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}, \widehat{\boldsymbol{\psi}}\right)\right]$ under the frequentist framework and CLR$= -2\left[\ln p\left(\mathbf{y}|\boldsymbol{\vartheta}, \boldsymbol{\psi}\right) - \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}, \widehat{\boldsymbol{\psi}}\right)\right]$ conditional on $\mathbf{y}$ under the Bayesian framework. They show that the two statistics have the same asymptotic distribution. Hence, based on posterior outputs, Chen et al. (2018) use CLR to calibrate the confidence interval for parameters in partially identified models.*

Proposition 3.2 *Suppose Assumptions 1-10 hold. Under H_0, we have*

$$PLR1 + q_{\vartheta} \xrightarrow{d} \chi^2(q_{\vartheta}) \text{ and } PLR2 \xrightarrow{d} \chi^2(q_{\vartheta}).$$

Under H_{L1}, we have

$$PLR1 + q_{\vartheta} \xrightarrow{d} \chi^2(q_{\vartheta}, \tau) \text{ and } PLR2 \xrightarrow{d} \chi^2(q_{\vartheta}, \tau),$$

where $\tau := \boldsymbol{\delta}'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\delta}$, $\boldsymbol{\Sigma}_{11}$ is the submatrix of $\mathbf{H}^{-1}(\boldsymbol{\theta}^*)$ corresponding to $\boldsymbol{\vartheta}$ with $\boldsymbol{\theta}^* = (\boldsymbol{\vartheta}^{*\prime}, \boldsymbol{\psi}^{*\prime})'$.

We are now in the position to construct posterior-test-based BFs via PLR. Let

$$\mathrm{BF}_{01}(\mathrm{PLR1}) = (cn+1)^{\frac{q_\vartheta}{2}} \exp\left\{-\frac{\mathrm{PLR1}+q_\vartheta}{2(cn+1)/(cn)}\right\}, \tag{3.13}$$

$$\mathrm{BF}_{01}(\mathrm{PLR2}) = (cn+1)^{\frac{q_\vartheta}{2}} \exp\left\{-\frac{\mathrm{PLR2}}{2(cn+1)/(cn)}\right\}. \tag{3.14}$$

When $\mathrm{BF}_{01}(\mathrm{PLR}) > 1$ or equivalently $\ln\mathrm{BF}_{01}(\mathrm{PLR}) > 0$, we find evidence to support H_0; otherwise, we find evidence against H_0.

Theorem 3.1 *Suppose Assumptions 1-10 hold. Under H_0, we have*

$$\ln BF_{01}^{J}(LR) = O_p(\ln n),\ \ \ln BF_{01}(PLR1) = O_p(\ln n),\ \ \textit{and}\ \ln BF_{01}(PLR2) = O_p(\ln n).$$

Under the alternative hypothesis with $\boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}_0 = O(n^{-1/2+a})$ for some $a>0$, we have

$$\begin{aligned}\ln BF_{01}^{J}(LR) &= O_p(n^{\max\{-\frac{1}{2}+\alpha,2\alpha\}}),\\ \ln BF_{01}(PLR1) &= O_p(n^{\max\{-\frac{1}{2}+\alpha,2\alpha\}}),\ \ln BF_{01}(PLR2) = O_p(n^{\max\{-\frac{1}{2}+\alpha,2\alpha\}}).\end{aligned}$$

Remark 3.9 *This theorem shows that, like the LR-test-based BFs of Johnson, our proposed PLR-test-based BFs have the consistency property. However, unlike the LR-test-based BFs of Johnson, our proposed PLR-test-based BFs can incorporate the prior information. Moreover, $BF_{01}^{J}(LR)$ is developed for i.i.d. data by Johnson. This assumption is relaxed in our proposed PLR-test-based BFs.*

Remark 3.10 *The proposed PLR-test-based BFs can avoid the JLB paradox. To see this, consider the example in Li et al. (2014). Let $y \sim N(\theta, 1)$ and consider $H_0 : \theta = 0$. Set the prior distribution of θ to $N(0, \tau_0^2)$. Then*

$$\theta|y \sim N(\mu(y), \omega^2)\ \textit{where}\ \mu(y) = \frac{\tau_0^2 y}{1+\tau_0^2}\ \textit{and}\ \omega^2 = \frac{\tau_0^2}{1+\tau_0^2},$$

$$\frac{1}{BF_{01}} = \sqrt{\frac{1}{1+\tau_0^2}} \exp\left\{\frac{\tau_0^2 y^2}{2(1+\tau_0^2)}\right\}.$$

When $\tau^2 \to +\infty$, $BF_{01} \to \infty$, that is, the test always supports H_0, giving rise to the JLB paradox. On the contrary, it is easy to show that

$$PLR1 = \frac{n\tau^2}{1+n\tau^2} z(\bar{y})^2 - 1\ \textit{and}\ PLR2 = \frac{n\tau^2}{1+n\tau^2} z(\bar{y})^2,$$

where $z(\bar{y}) = \sqrt{n}(\bar{y}-\theta)$ is the standard z-statistic. When H_0 is true, $z(\bar{y})$ converges to $N(0,1)$ and both $PLR1+1$ and $PLR2$ are asymptotically distributed as $\chi^2(1)$, avoiding the paradox. Consequently, $BF_{01}(PLR1)$ and $BF_{01}(PLR2)$ can also avoid the paradox.

Remark 3.11 *BF_{01}(PLR1) and BF_{01}(PLR2) share the coherence property of $BF_{01}^{J}(LR)$. When M_1 is nested within M_2 and M_2 is nested within M_3, we have $BF_{31}(PLR) = BF_{32}(PLR) \times BF_{21}(PLR)$. Hence, they can be used to do multiple hypothesis testing; see Hu and Johnson (2009).*

3.4 PWald-based BFs

Liu et al. (2022) propose the following net loss function for hypothesis testing

$$\triangle\mathcal{L}[H_0, (\boldsymbol{\vartheta}, \boldsymbol{\psi})] = (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \, \mathbf{V}_{\vartheta\vartheta}^{-1}(\overline{\boldsymbol{\theta}}) \, (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0),$$

where $\mathbf{V}_{\vartheta\vartheta}(\overline{\boldsymbol{\theta}})$ is the submatrix of $\mathbf{V}(\overline{\boldsymbol{\theta}})$ corresponding to $\boldsymbol{\vartheta}$, and $\mathbf{V}(\overline{\boldsymbol{\theta}})$ is the posterior covariance matrix under H_1 given by

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

with $\overline{\boldsymbol{\theta}}$ being the posterior mean of $\boldsymbol{\theta}$ under H_1. They then construct a Wald-like test statistic based on posterior outputs as:

$$\text{PWald} = \int (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \, \mathbf{V}_{\vartheta\vartheta}^{-1}(\overline{\boldsymbol{\theta}}) \, (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \, p(\boldsymbol{\vartheta}|\mathbf{y}) d\boldsymbol{\vartheta}. \tag{3.15}$$

Note that the Wald statistic is

$$\text{Wald} = (\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)' \left[-\frac{1}{n} [\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\theta}})]_{11} \right]^{-1} (\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0),$$

where $[\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\theta}})]_{11}$ is the corresponding submatrix of $\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\vartheta}$. Under Assumptions 1-10, when H_0 holds and the likelihood information dominates the prior information, Liu et al. (2022) show that

$$\text{PWald} - q_\vartheta = \text{Wald} + o_p(1) \xrightarrow{d} \chi^2(q_\vartheta). \tag{3.16}$$

Hence, PWald can be understood as a posterior version of Wald.

Proposition 3.3 *Suppose Assumptions 1-10 hold. Under H_{L1}, as $n \to \infty$, we have*

$$PWald - q_\vartheta = Wald + o_p(1) \xrightarrow{d} \chi^2(q_\vartheta, \tau) \text{ with } \tau = \boldsymbol{\delta}' \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\delta}, \tag{3.17}$$

where $\boldsymbol{\Sigma}_{11} = [\mathbf{H}^{-1}(\boldsymbol{\theta}^*)]_{11}$.

Let the posterior-test-based BF from PWald be

$$\mathrm{BF}_{01}(\mathrm{PWald}) = (cn+1)^{\frac{q_\vartheta}{2}} \exp\left\{-\frac{\mathrm{PWald} - q_\vartheta}{2(cn+1)/(cn)}\right\}. \tag{3.18}$$

Remark 3.12 *Equations (3.16) and (3.17) give the asymptotic distributions of PWald under H_0 and H_{L1}, respectively. When $BF_{01}(PWald) > 1$ or $\ln BF_{01}(PWald) > 0$, we find evidence to support H_0; otherwise, we find evidence to support H_1.*

Theorem 3.2 *Suppose Assumptions 1-10 hold. Under H_0, we have*

$$\ln BF_{01}(PWald) = O_p(\ln n). \tag{3.19}$$

Under the alternative hypothesis with $\boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}_0 = O(n^{-1/2+a})$ for some $a > 0$, we have

$$\ln BF_{01}(PWald) = -O_p(n^{2a}). \tag{3.20}$$

Remark 3.13 *Theorem 3.2 shows that the proposed PWald-test-based BFs are consistent.*

Similar to the idea in Equation (2.5), one can define a frequentist-test-based BF via the Wald statistic as:

$$\mathrm{BF}_{01}^{J}(\mathrm{Wald}) = (cn+1)^{\frac{q_\vartheta}{2}} \exp\left\{-\frac{\mathrm{Wald}}{2(cn+1)/(cn)}\right\}. \tag{3.21}$$

This is closely related to Equation (1) in Johnson (2008) although his definition is developed for a specific model.

Remark 3.14 *Unlike $BF_{01}^{J}(Wald)$, $BF_{01}(PWald)$ incorporates the prior information when it is available. Unlike $BF_{01}(PLR)$ that requires posterior outputs under H_0 and H_1, $BF_{01}(PWald)$ only requires posterior outputs under H_1. Moreover, unlike $BF_{01}(PLR)$ that is based on the likelihood function, $BF_{01}(PWald)$ does not need the likelihood function. Hence, for models where the likelihood function is difficult to calculate, $BF_{01}(PWald)$ is easier to obtain than $BF_{01}(PLR)$.*

4 Simulation Studies

In this section, we design two simulation experiments to investigate the finite sample performance of the proposed posterior-test-based BFs and to compare the performance with that of the frequentist test-based BFs of Johnson and the conventional BF. In the first experiment, we consider a simple linear regression model to illustrate how the proposed posterior-test-based BFs can avoid the p-hacking problem and the JLB paradox as well as examine the usefulness of informative priors. In the second experiment, we consider a nonlinear regression model where the likelihood function is multi-modal, making MLE sensitive to initial values and distorting the performance of the frequentist test statistics and the corresponding test-based BFs. However, the posterior distributions are not subject to the multi-modality problem when the sample size is large or when the informative prior is used. Therefore, our posterior-test-based BFs have good finite-sample performances.

4.1 Model 1: Simple linear regression model

In this subsection, we consider the following simple linear regression model where all BFs have closed-form expressions:

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

where $x_i \sim i.i.d.N(0,1)$ and fixed under repeated sampling. We test $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$. For the Bayesian analysis, the conjugate Normal-Gamma priors are used, i.e.,

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

where $\mu_0 = (\mu_\alpha, \mu_\beta)'$, V_0 =diag(V_α, V_β), $\Gamma(a, b)$ denotes the gamma distribution with the shape parameter a and the rate parameter b. Let $\mu = (\alpha, \beta)'$, $\mathbf{y} = \begin{pmatrix} y_1 & \ldots & y_n \end{pmatrix}'$, $\mathbf{X} = \begin{pmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_n \end{pmatrix}'$. The posterior distributions under H_1 are:

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

where $V_1 = \left(\mathbf{X}'\mathbf{X} + V_0^{-1}\right)^{-1}$ and $\mu_1 = V_1(\mathbf{X}'\mathbf{X}\widehat{\mu} + V_0^{-1}\mu_0) = V_1(\mathbf{X}'\mathbf{y} + V_0^{-1}\mu_0)$ with $\widehat{\mu}$ being the usual OLS estimator of μ. The posterior distributions under H_0 are

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

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 draw 10,000 random samples from each posterior distribution. Based on these random samples, we then calculate the posterior-test-based BFs.[3]

[3]The closed-form expressions for BF_{01}, $\text{BF}_{01}^J(\text{LR})$ and $\text{BF}_{01}^J(\text{Wald})$ can be easily obtained for this model.

Table 1: Rejection rate, mean, standard deviation of BFs when $\beta^* = 0.0$

	Non-informative prior (NP)			
n	$n=50$	$n=500$	$n=5000$	$n=15000$
BF_{01}	0.00% (539.12, 218.33)	0.00% (1536.88, 632.25)	0.00% (5000.73, 1871.56)	0.00% (8669.58, 3302.44)
BF_{01}^J(LR)	4.40% (5.11, 2.05)	1.00% (15.58, 6.40)	0.20% (50.59, 18.93)	0.00% (87.02, 33.15)
BF_{01}^J(Wald)	4.50% (5.07, 2.01)	1.00% (15.57, 6.39)	0.20% (50.59, 18.93)	0.00% (87.02, 33.14)
BF_{01}(PLR1)	4.10% (5.09, 1.99)	1.00% (15.58, 6.38)	0.20% (50.60, 18.95)	0.00% (87.03, 33.17)
BF_{01}(PLR2)	4.10% (5.09, 1.99)	1.00% (15.58, 6.39)	0.20% (50.59, 18.93)	0.00% (87.02, 33.14)
BF_{01}(PWald)	4.60% (5.07, 2.01)	1.10% (15.57, 6.39)	0.20% (50.60, 18.92)	0.00% (87.05, 33.12)
	Informative Prior (IP)			
	$n=50$	$n=500$	$n=5000$	$n=15000$
BF_{01}	31.10% (1.00, 0.04)	29.00% (1.05, 0.19)	13.50% (1.81, 0.60)	7.90% (2.88, 1.05)
BF_{01}^J(LR)	4.40% (5.11, 2.05)	1.00% (15.58, 6.40)	0.20% (50.59, 18.93)	0.00% (87.02, 33.15)
BF_{01}^J(Wald)	4.50% (5.07, 2.01)	1.00% (15.57, 6.39)	0.20% (50.59, 18.93)	0.00% (87.02, 33.14)
BF_{01}(PLR1)	0.00% (7.05, 0.73)	0.00% (19.32, 3.57)	0.00% (52.88, 17.46)	0.00% (88.45, 32.32)
BF_{01}(PLR2)	0.00% (6.96, 0.26)	0.00% (19.29, 3.53)	0.00% (52.88, 17.44)	0.00% (88.44, 32.30)
BF_{01}(PWald)	0.00% (6.96, 0.25)	0.00% (19.29, 3.53)	0.00% (52.87, 17.46)	0.00% (88.43, 32.32)

To examine the influence of prior, two prior distributions are considered, a non-information prior and an informative prior. The informativeness level is set such that under the small sample size (say $n = 50$), the prior information dominates the data information, while under the large sample size ($n = 5,000$), the data information exceeds the prior information. To be more specific,

- for the non-informative prior (NP), we set $(\mu_\alpha, \mu_\beta, V_\alpha, V_\beta, a, b) = (\alpha^*, \beta^*, 10000, 10000, 1, 1)$;
- for the informative prior (IP), we set $(\mu_\alpha, \mu_\beta, V_\alpha, V_\beta, a, b) = (\alpha^*, \beta^*, 10000, 0.001, 1, 1)$.

We set the true values for α^* and σ^{2*} to 1. Two β values $(0.0, 0.1)$ are considered to obtain the Type-I and Type-II errors. When $\beta = 0$, four sample sizes (n=50, 500, 5,000, 15,000) are considered; When $\beta = 0.1$, three sample sizes (n=50, 500, 5,000) are considered. For each case, we simulate data from the true DGP and perform hypothesis testing based on various BFs, replicating the experiment 1,000 times. We report the rejection rate of H_0, the mean, and the standard deviation of BFs across 1,000 replications. The results are reported in Table 1 when $\beta^* = 0.0$ and in Table 2 when $\beta^* = 0.1$.

Table 2: Rejection rate, mean, standard deviation of BFs when $\beta^* = 0.1$

	Non-informative prior (NP)		
	$n = 50$	$n = 500$	$n = 5000$
BF_{01}	0.30% (461.82, 245.87)	4.40% (484.21, 610.65)	99.80% (0.01, 0.13)
$\mathrm{BF}_{01}^{J}(\mathrm{LR})$	12.40% (4.39, 2.31)	38.30% (4.92, 6.19)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}^{J}(\mathrm{Wald})$	12.40% (4.35, 2.28)	38.40% (4.91, 6.19)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}(\mathrm{PLR1})$	11.60% (4.38, 2.26)	38.20% (4.93, 6.19)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}(\mathrm{PLR2})$	11.50% (4.38, 2.26)	38.10% (4.92, 6.19)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}(\mathrm{PWald})$	12.50% (4.35, 2.28)	38.50% (4.91, 6.19)	100.00% (0.00, 0.00)
	Informative prior (IP)		
	$n = 50$	$n = 500$	$n = 5000$
BF_{01}	64.50% (1.00, 0.90)	84.80% (0.60, 1.41)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}^{J}(\mathrm{LR})$	12.40% (4.39, 2.31)	38.30% (4.92, 6.19)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}^{J}(\mathrm{Wald})$	12.40% (4.35, 2.28)	38.40% (4.91, 6.19)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}(\mathrm{PLR1})$	100.00% (0.05, 0.04)	99.30% (0.07, 0.17)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}(\mathrm{PLR2})$	100.00% (0.05, 0.04)	99.30% (0.07, 0.17)	100.00% (0.00, 0.00)
$\mathrm{BF}_{01}(\mathrm{PWald})$	100.00% (0.07, 0.08)	98.90% (0.08, 0.19)	100.00% (0.00, 0.00)

Some interesting findings emerge from the two tables. First, all BFs can consistently select the true model, as indicated by the last column of both tables. The only exception is for the standard BF under the informative prior. When $n = 15,000$, its Type I error is 7.9%. Since the standard BF is consistent, we expect the rejection rate eventually becomes zero. Our unreported simulations suggest that a sample size of more than 10^8 is required to achieve zero rejection rate. It is surprising that the standard BF requires such a big sample size under such a simple model.

Second, as expected, the standard BF suffers from the JLB paradox. For example, when the true value of β is 0.1 and n is 50 (or 500), the standard BF under the vague prior only rejects H_0 in 0.30% (or 4.4%) of replications.

Third, two frequentist-test-based BFs of Johnson, $\mathrm{BF}_{01}^{J}(\mathrm{LR})$ and $\mathrm{BF}_{01}^{J}(\mathrm{Wald})$, can avoid the JLB paradox but are independent of the prior distribution. That is, the prior information could not improve their performance.

Fourth, three posterior-test-based BFs, $\mathrm{BF}_{01}(\mathrm{PLR1})$, $\mathrm{BF}_{01}(\mathrm{PLR2})$, and $\mathrm{BF}_{01}(\mathrm{PWald})$

can avoid the JLB paradox. When the true value of β is 0.1, three posterior-test-based BFs under the vague prior lead to nearly identical rejection rates to two frequentist-test-based BFs. Under the informative prior, they lead to much better rejection rates when n is 50 or 500. When the true value of β is 0.0, under the informative prior, they lead to much better rejection rates (always zero) than BF_{01} and two frequentist-test-based BFs.

Finally, to illustrate the p-hacking problem in p-value-based methods and how the test-based BFs can avoid it, we conduct a small trial using data simulated from the simple linear regression model with the 5% significance level. There are three common practices to do p-hacking.

1. Random sampling and selective reporting. As pointed out by Rouder et al. (2009), the p-value can randomly walk below the pre-specified significance level even when H_0 is true. Therefore, one can repeatedly collect different observations until a small p-value is found but selectively report the significant results only. In our simulation, if we set β^* at 0, the random seed for generating $\mathbf{y}$ at 88, the random seed for generating x at 12345, and the prior the noninformative one, we would find that all the p-value-based methods, including LR, Wald, PLR1, PLR2, and PWald, result in a p-value around $0.0058 < 0.05$ when $n = 15,000$, rejecting H_0 falsely. On the contrary, all the posterior-test-based BFs are around 2.72, suggesting H_0 cannot be rejected.

2. Increase the sample size. It is possible that with $\beta^* = 0$, the test statistics do not reject H_0 when n is small, but reject it when n is bigger. For instance, if set the random seed for generating $\mathbf{y}$ at 88, the random seed for generating $\{x_t\}$ at 12345, and the prior the noninformative one, we would find that all p-value-based methods result in a p-value around $0.0845 > 0.05$ when $n = 5,000$, suggesting the null hypothesis cannot be rejected at the 5% significance level. However, as shown in 1, when $n = 15,000$, these test statistics reject H_0. On the contrary, the corresponding posterior-test-based BFs take values of around 16.00 and 2.72 for $n = 5,000$ and $n = 15,000$ respectively, suggesting that H_0 cannot be rejected.

3. Use a subsample of data. It is possible that one can obtain either significant results or insignificant results depending on the subsample he uses. For instance, given the data set used in 1, we find that all the p-value-based statistics cannot reject H_0 if the first 1/3 subsample or the last 1/3 subsample is used (with the p-value being 0.0845

and 0.7638 respectively). However, all the p-value-based statistics reject H_0 if one uses the middle 1/3 subsample with the p-value being $0.0067 < 0.05$. The corresponding test-based BFs based on the three subsamples are 16.00, 3.11, and 67.63, respectively, all suggesting that H_0 cannot be rejected.

4.2 Model 2: Nonlinear regression model

In this subsection, we use a nonlinear regression model where the likelihood function is multi-modal. In this case, the ML estimation can be sensitive to the initial values, and hence, distort the performance of frequentist-test-based BFs. On the contrary, the posterior distribution of parameters is immune to multi-modality of likelihood, especially when the sample size is reasonably large or when the prior is informative. Consequently, we expect better performances of the posterior-test-based BFs than the frequentist-test-based BFs.

Consider the nonlinear regression model of Dorsey and Mayer (2000),

$$y_i = \theta_1 + \theta_1^2 x_{1i} + \theta_2 x_{2i} + \theta_2^2 x_{3i} + \varepsilon_i, i = 1, \cdots, n$$

where $x_{ji} \sim i.i.d.U(0,1)$ for $j = 1, 2, 3$ and $\varepsilon_i \sim i.i.d.N(0,1)$. To simulate data, we set $\theta_1^* = 2$ and $\theta_2^* = -2$ or 0 and test $H_0 : \theta_2 = 0$ against $H_1 : \theta_2 \neq 0$. When $\theta_2 = 0$, four sample sizes (n=100, 1000, 10000, 30000) are considered; when $\theta_2 = 0.2$, three sample sizes (n=100, 1000, 10000) are considered. We replicate the experiment for 500 times and report the empirical rejection rate of H_0 based on various test-based BFs across 500 replications.

The prior distributions are specified as

$$\theta_1 \sim N(\theta_1^*, 10000), \text{ and } \theta_2 \sim N(\theta_2^*, 10000) \text{ or } \theta_2 \sim N(\theta_2^*, 0.01).$$

As there is multi-modality in the likelihood function, we use the Sequential Monte Carlo (SMC) technique of Herbst and Schorfheide (2014) for posterior sampling. SMC utilizes a set of particles to approximate the posterior distribution and is robust to multi-modality in the target distribution.[4] With the posterior draws, we can compute PLR1, PLR2, PWald, BF_{01}(PLR1), BF_{01}(PLR2), BF_{01}(PWald). For LR, Wald, and their corresponding test-based BFs, one needs to obtain the MLE numerically and the initial values of θ_1 and θ_2 are randomly drawn from $N(0, 100)$ for numerical optimizations.

[4]For SMC, the number of particles is set at $M = 1,000$, the number of grids between zero and one ($S = 500$), and the grids $b_s = (\frac{s}{S})^\lambda, s = 1, 2, \cdots, S, \lambda = 2$. In each iteration with respect to b_s, the mutation step conducts Metropolis-Hastings sampling for once.

Table 3: Rejection rate of BFs for the nonlinear regression model

	Non-informative prior (NP)						
	$\theta_2^* = 0$				$\theta_2^* = -2$		
	$n = 100$	$n = 1000$	$n = 10000$	$n = 30000$	$n = 100$	$n = 1000$	$n = 10000$
BF_{01}^J(LR)	91.00%	0.20%	0.20	0.00%	75.80%	53.00%	53.60%
BF_{01}^J(Wald)	90.40%	96.40%	99.20%	99.20%	100.00%	99.80%	100.00%
BF_{01}(PLR1)	1.60%	0.60%	0.60%	0.00%	100.00%	100.00%	100.00%
BF_{01}(PLR2)	1.40%	0.60%	0.60%	0.00%	100.00%	100.00%	100.00%
BF_{01}(PWald)	1.80%	1.00%	0.60%	0.00%	100.00%	100.00%	100.00%
	Informative prior (IP)						
	$\theta_2^* = 0$				$\theta_2^* = -2$		
	$n = 100$	$n = 1000$	$n = 10000$	$n = 30000$	$n = 100$	$n = 1000$	$n = 10000$
BF_{01}^J(LR)	90.80%	0.40%	0.40%	0.00%	73.20%	50.20%	54.20%
BF_{01}^J(Wald)	90.60%	96.20%	98.80%	99.40%	100.00%	100.00%	100.00%
BF_{01}(PLR1)	0.00%	0.00%	0.40%	0.00%	100.00%	100.00%	100.00%
BF_{01}(PLR2)	0.00%	0.00%	0.40%	0.00%	100.00%	100.00%	100.00%
BF_{01}(PWald)	0.00%	0.00%	0.60%	0.00%	100.00%	100.00%	100.00%

The rejection rates of H_0 of alternative test-based BFs are reported in Table 3. There are three main findings from the table. First, BF_{01}^J(LR) and BF_{01}^J(Wald) do not perform well. For example, when H_0 is true and $n = 100$, under the non-informative prior, BF_{01}^J(LR) incorrectly rejects H_0 in 91% of replications; BF_{01}^J(Wald) incorrectly rejects H_0 in 90.4% of replications. This very high false rejection rate even goes up for BF_{01}^J(Wald) when n increases. The rejection rates under the informative prior are nearly unchanged as they are independent of prior.[5] When H_0 is false, under the non-informative prior, BF_{01}^J(LR) incorrectly accepts H_0 in 25%, 47%, and 47% of replications if n=100, 1000, 10000, respectively. Second, BF_{01}(PLR1), BF_{01}(PLR2) and BF_{01}(PWald) perform much better regardless of n or prior. When H_0 is true, the rejection rates are either zero or very close to zero. When H_0 is false, the rejection rates are always one. Third, the use of informative prior improves the finite sample performance of BF_{01}(PLR1), BF_{01}(PLR2) and BF_{01}(PWald) when H_0 is true and n is 100 or 1000.

To understand why three posterior-test-based BFs perform much better than two frequentist-test-based BFs, we obtain the rejection rate of H_0 for LR, Wald, PLR1, PLR2, and PWald under the same simulation design. Table 4 reports the rejection rates. It is clear that there are serious size distortions in LR when $n = 100$ and in Wald for all sample sizes. Their empirical size and empirical power do not approach their nominal levels. This suggests that MLE, which depends on the initial values, may correspond to a local maximum of the likelihood function, and hence, the finite sample distributions of LR and Wald are not close to their

[5]They are slightly different because the initial values are randomly picked.

Table 4: Rejection rate of LR, Wald, PLR1, PLR2, and PWald for the nonlinear regression model

Non-informative prior (NP)						
	$\theta_2^* = 0$			$\theta_2^* = -2$		
	$n = 100$	$n = 1000$	$n = 10000$	$n = 100$	$n = 1000$	$n = 10000$
LR	91.20%	3.20%	3.20%	75.80%	53.00%	53.60%
Wald	91.40%	97.00%	99.60%	100.00%	99.80%	100.00%
PLR1	2.60%	4.40%	5.80%	100.00%	100.00%	100.00%
PLR2	2.40%	4.40%	5.80%	100.00%	100.00%	100.00%
PWald	3.20%	4.40%	5.80%	100.00%	100.00%	100.00%
Informative prior (IP)						
	$\theta_2^* = 0$			$\theta_2^* = -2$		
	$n = 100$	$n = 1000$	$n = 10000$	$n = 100$	$n = 1000$	$n = 10000$
LR	91.20%	2.60%	1.80%	73.20%	50.20%	54.20%
Wald	91.40%	96.80%	99.40%	100.00%	100.00%	100.00%
PLR1	0.00%	0.60%	5.40%	100.00%	100.00%	100.00%
PLR2	0.00%	0.60%	5.40%	100.00%	100.00%	100.00%
PWald	0.00%	0.60%	5.20%	100.00%	100.00%	100.00%

asymptotic distributions. Whereas, the posterior distribution has a less serious problem in terms of multi-modality.

To get the support of this argument, we simulate three sample paths with $\theta^* = (2, -2)$ and $n = 10, 10000$. Based on the simulated paths, we plot the posterior densities of θ_1 and θ_2 and the likelihood function of θ in Figures 1, 2 under the non-informative prior when $n = 10, 10000$. We also plot the posterior densities of θ_1 and θ_2 in Figure 3 under the information prior when $n = 10$. It is clear that when $n = 10$ and under the non-informative prior (NP), both the likelihood function and the posterior distributions are multi-modal. The multi-modality problem disappears in the posterior distributions when n increases but stays in the likelihood function even when $n = 10000$. Moreover, the multi-modality problem disappears in the posterior distributions when $n = 10$ but the informative prior is used.

5 Empirical Studies

In this section, we apply the proposed test-based BFs to three empirical examples, a linear regression model, a time-varying parameter model, and a stochastic volatility (SV) model. Specifically, for a time-varying parameter model, the likelihood function must be computed from recursion, making MLE not easy to obtain. Thus, we only report two PLR-based BFs and the PWald-based BF. For the SV model, the likelihood function does not have an analytical expression, making the likelihood function and MLE even more difficult

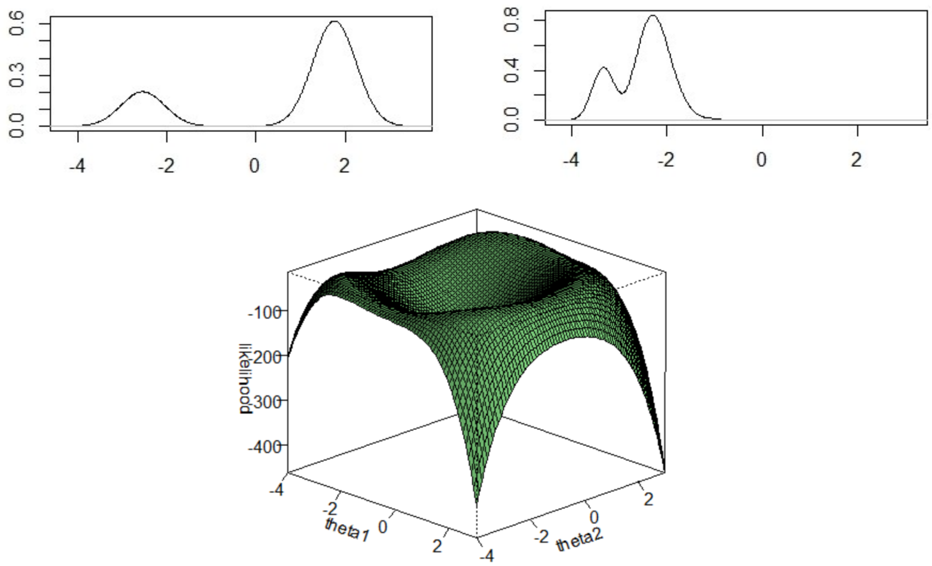

Figure 1: Posterior densities of θ_1 (top left) and θ_2 (top right), and the likelihood function of θ (bottom) under the non-informative prior and when $\theta^* = (2, -2), n = 10$

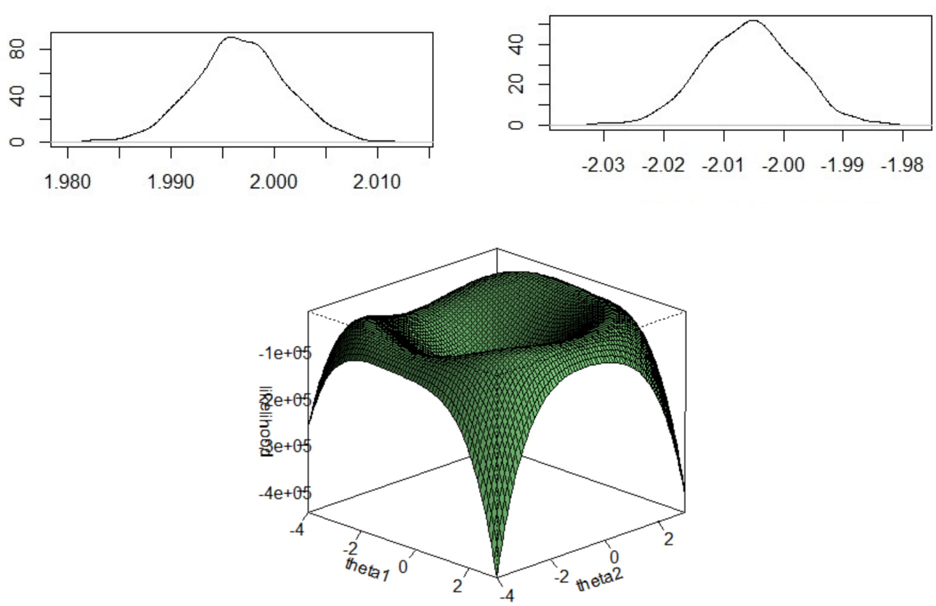

Figure 2: Posterior densities of θ_1 (top left) and θ_2 (top right), and the likelihood function of θ (bottom) under the non-informative prior and when $\theta^* = (2, -2), n = 10000$

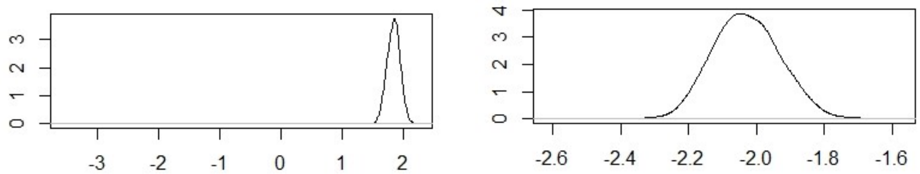

Figure 3: Posterior densities of θ_1 (left) and θ_2 (right) under the informative prior and when $\theta^* = (2, -2), n = 10$

to obtain. Consequently, LR, Wald, PLR1, and PLR2 are all difficult to obtain. In practice, the SV model is often estimated by MCMC. From MCMC outputs, one can easily obtain PWald. This is why we only report PWald-based BF for the SV model.

5.1 A simple linear regression model

The first empirical study is a simple linear regression model where the dependent variable is the daily log returns of S&P 500 (denoted by $\Delta \ln s_t$) and the independent variable is the corresponding log returns of the futures (denoted by $\Delta \ln f_t$). The sample period is from January 22, 2019 to October 14, 2022. The effective sample size is 902. The model is

$$\Delta \ln s_t = \alpha + \beta \Delta \ln f_t + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.t(0, \sigma^2, v),$$

where β captures the optimal hedge ratio. We test $H_0 : \beta = 1$ against $H_1 : \beta \neq 1$.

The prior distribution for the parameters are set as

$$(\alpha, \beta)' \sim N(\mu_0, V_0), \mu_0 = (0, 0)', V_0 = \text{diag}(100, V_\beta), \sigma^{-2} \sim \Gamma(1, 1), v - 2 \sim \text{Exp}(0.05).$$

To compute the posterior statistics, we use WinBUGS to obtain posterior samplers and compute the corresponding statistics. Since we assume the t distribution for the error term, which is empirically more reasonable, the standard BF does not have an analytical expression. We use the algorithm proposed by Li et al. (2023) to calculate the standard BF. To check the sensitivity to the prior, we let V_β vary from 10^0 to 10^{15}.

Various BFs are reported in Table 5. Again, the JLB paradox is clearly seen in the standard BF. When $V_\beta = 10^{15}$ (i.e., a very vague prior), BF supports H_0. Whereas, when the prior becomes more informative, BF supports H_1. By taking nearly identical values, all the other test-based BFs are immune to the JLB paradox and reject H_0.

Table 5: Various BFs for the linear regression model

Hyper-parameters	$V_\beta = 1$	$V_\beta = 10^5$	$V_\beta = 10^{10}$	$V_\beta = 10^{15}$
$\ln \mathrm{BF}_{01}$	-7.22	1.43	13.23	39.63
$\ln \mathrm{BF}_{01}^J(\mathrm{LR})$	-4.06	-4.06	-4.06	-4.06
$\ln \mathrm{BF}_{01}^J(\mathrm{Wald})$	-3.83	-3.83	-3.83	-3.83
$\ln \mathrm{BF}_{01}(\mathrm{PLR1})$	-3.95	-3.91	-3.93	-3.93
$\ln \mathrm{BF}_{01}(\mathrm{PLR2})$	-3.98	-3.95	-3.95	-3.95
$\ln \mathrm{BF}_{01}(\mathrm{PWald})$	-3.74	-3.94	-3.92	-3.75

5.2 An extended CAPM

In this section, we test the extended capital asset pricing model (CAPM) of Sharpe (1964), where beta is allowed to be time-varying. To model the time-varying beta, following Mergner and Bulla (2008), we write the extended CAPM in a state-space form,

$$r_t^s = \alpha + \beta_t r_t^m + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2),$$
$$\beta_{t+1} = \mu + \phi(\beta_t - \mu) + \eta_t, \quad \eta_t \sim i.i.d.N(0, \sigma_\eta^2),$$

where r_t^s, r_t^m are weekly excess returns for the pan-European insurance industry portfolio and the DJ STOXX 600 index from December 2, 1987 to January 14, 2016, respectively. When there is no pricing error, α must be zero. Hence, we test $H_0 : \alpha = 0$ vs $H_1 : \alpha \neq 0$. The sample size is 1,467.

Although the likelihood function of this model can be computed via the Kalman filter, it is known that ML estimation for this model can be unstable. Moreover, the Hessian matrix is not easy to compute. Whereas, the MCMC analysis is relatively easier to conduct. Therefore, we use our posterior-test-based BFs to perform hypothesis testing. To make posterior draws, following Li et al. (2018), we use the following priors,

$$\alpha \sim N(0, 10^3),\ \mu \sim N(0, 100^3),\ \phi \sim Beta(1, 1),\ \sigma_\varepsilon^{-2} \sim \Gamma(0.001, 0.001),\ \sigma_\eta^{-2} \sim \Gamma(0.001, 0.001).$$

For the model under H_0, we draw 500,000 MCMC samples with the first 50,000 as burn-in. We then take 1 observation for every 45 iterations, resulting in 10,000 effective samples. For the model under H_1, we draw 150,000 MCMC samples with the first 20,000 as burn-in. We then take 1 observation for every 13 iterations, also resulting in 10,000 effective samples. The posterior-test-based BFs are reported in Table 6. They all favor H_0, suggesting the extended CAPM can price the returns of the pan-European insurance industry portfolio well.

Table 6: Posterior-test-based BFs for the extended CAPM

$\ln \text{BF}_{01}(\text{PLR1})$	3.0267
$\ln \text{BF}_{01}(\text{PLR2})$	3.0368
$\ln \text{BF}_{01}(\text{PWald})$	3.0384

5.3 A stochastic volatility model

The SV model is a nonlinear non-Gaussian state space model where the likelihood function is intractable as it involves high-dimensional integrals. As a result, the MLE is very difficult to obtain. In practice, MCMC is often used to provide full likelihood-based estimation.

Following Yu (2005), we formulate the SV model with the leverage effect as,

$$y_t - \alpha = \exp(h_t/2)u_t, \quad u_t \sim N(0,1),$$
$$h_{t+1} = \mu + \phi(h_t - \mu) + v_{t+1}, \quad v_{t+1} \sim N(0, \tau^2),$$

and $h_0 = \mu$. Let $Corr(u_t, v_{t+1}) = \rho$ captures the leverage effect if $\rho < 0$. We test $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$.

We use 945 daily mean-corrected returns on Pound/Dollar exchange rates from October 1, 1981 to June 28, 1985. Following Meyer and Yu (2000), we use the following priors:

$$\mu \sim N(0, 100), \phi \sim Beta(1, 1,), 1/\tau^2 \sim \Gamma(0.001, 0.001), \rho \sim U(-1, 1).$$

For posterior sampling, we iterate the Gibbs sampler 110,000 times and throw away the first 10,000 iterations. We collect every 20th iterations to get 5,000 effective draws. We then use the 5,000 posterior samples to obtain $\text{BF}_{01}(\text{PWald})$, which equals 30.67. Hence, we find no evidence of a leverage effect in the exchange rate data.

6 Conclusion

This paper is concerned with hypothesis testing. The p-value-based methods, including all the frequentist tests relying on the asymptotic theory, are subject to the p-hacking problem. One of the manifestations of the p-hacking problem is publication bias. Not surprisingly, hypothesis testing based on p-values has been criticized by many researchers.

As important alternatives, BFs have been suggested as a replacement for hypothesis testing. BFs have a consistent property so that they can select the true model with probability going to one. Unfortunately, the standard BF suffers from both theoretical and computational difficulties, including the JLB paradox. In this paper, based on the posterior test statistics, we mainly propose two new BF-like statistics for hypothesis testing. The first one is constructed from a posterior version of the LR test while the second one is from the posterior Wald test.

The two proposed posterior-test-BFs inherit some of the good properties of the standard BFs and the frequentist test statistics and avoid many problems in the standard BFs and the frequentist test statistics. In particular, they inherit the consistent property of the standard BF and hence, avoid the p-hacking problem. Moreover, compared with the standard BF, they avoid the JLB paradox because they are constructed from the posterior test statistics, which are based on continuous loss functions. Furthermore, compared with frequentist test statistics and the corresponding frequentist-test-based BFs of Johnson (2005, 2008), they can incorporate the prior information to improve the test behavior. This is an important advantage when the informative prior is indeed available. Last but not least, they are based on posterior outputs, and hence, avoid the need to do frequentist estimation. An important frequentist estimation technique is maximum likelihood. For many important models, maximum likelihood is difficult to use because (1) the likelihood function does not have a closed-form expression; (2) the likelihood function is multi-modal.

We have designed two experiments to study the finite-sample performances of the proposed posterior-test-based BFs and compare them with those of the standard BF and the frequentist-test-based BFs of Johnson (2005, 2008). The simulation studies confirm the consistency property and hence, avoid the p-hacking problem. The simulation studies also show that the posterior test-based BFs outperform the standard and frequentist-test-based BFs. When applying the proposed posterior-test-based BFs to real data, we continue to find that the proposed posterior-test-based BFs lead to good performances and are easier to implement, even for models where MLE is difficult to obtain.

Appendix

Appendix 1: Proof of Proposition 3.1

Note that

$$\int \ln\left[p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\right]p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} = \int [\ln p(\mathbf{y}|\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} + \int [\ln p(\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}. \tag{6.1}$$

It is sufficient to derive large sample approximations for $\int [\ln p(\boldsymbol{\theta})]\, p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$ and $\int [\ln p(\mathbf{y}|\boldsymbol{\theta})]\, p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$, respectively.

First, note that

$$\begin{aligned}\int [\ln p(\boldsymbol{\theta})] p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} &= \int \ln p(\boldsymbol{\theta}) \frac{p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y})} d\boldsymbol{\theta} \\ &= \frac{\int \ln p(\boldsymbol{\theta}) \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{p(\mathbf{y})} = \frac{\int \ln p(\boldsymbol{\theta}) \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{\int \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}.\end{aligned}$$

From this formula, we let

$$g_0(\boldsymbol{\theta}) = \ln p(\boldsymbol{\theta}),\ b_D(\boldsymbol{\theta}) = 1,\ \text{and}\ h_n(\boldsymbol{\theta}) = -\frac{1}{n} \ln\left[p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right].$$

According to Lemma 6.1, and noting that $b_D^{(1)}(\boldsymbol{\theta}) = 0$, we can show that

$$\int [\ln p(\boldsymbol{\theta})] p(\boldsymbol{\theta}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} = \widehat{g}_0 + \frac{1}{n} B_0 + O_p\left(\frac{1}{n^2}\right), \tag{6.2}$$

where

$$B_0 = \frac{1}{2} \mathrm{tr}\left[\left(\widehat{h}^{(2)}\right)^{-1} \widehat{g}_0^{(2)}\right] - \frac{1}{2} \mathrm{vec}\left(\left(\widehat{h}^{(2)}\right)^{-1}\right) \widehat{h}^{(3)} \left(\widehat{h}^{(2)}\right)^{-1} \widehat{g}_0^{(1)}.$$

Second, let $g_t(\boldsymbol{\theta}) = \ln p(y_t|I_{t-1}, \boldsymbol{\theta})$, where I_{t-1} is the information set generated by $\{y_0, y_1, y_2, \cdots, y_{t-1}\}$. We can write

$$\ln p(\mathbf{y}|\boldsymbol{\theta}) = \sum_{t=1}^{n} \ln p(y_t|I_{t-1}, \boldsymbol{\theta}) = \sum_{t=1}^{n} g_t(\boldsymbol{\theta}).$$

Hence, we can show that

$$\begin{aligned}&\int [\ln p(\mathbf{y}|\boldsymbol{\theta})] p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = \int \ln p(\mathbf{y}|\boldsymbol{\theta}) \frac{p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y})} d\boldsymbol{\theta} \\ &= \frac{\int \ln p(\mathbf{y}|\boldsymbol{\theta}) \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{p(\mathbf{y})} = \frac{\int \ln p(\mathbf{y}|\boldsymbol{\theta}) \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{\int p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \\ &= \frac{\int \ln p(\mathbf{y}|\boldsymbol{\theta}) \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{\int \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}} = \frac{\int \sum_{t=1}^{n} \ln p(\mathbf{y}_t|I_{t-1}, \boldsymbol{\theta}) \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{\int \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}} \\ &= \sum_{t=1}^{n} \left[\frac{\int g_t(\boldsymbol{\theta}) \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{\int \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}\right].\end{aligned}$$

From this formula, we also can observe that

$$b_D(\boldsymbol{\theta}) = 1 \text{ and } h_n(\boldsymbol{\theta}) = -\frac{1}{n} \ln\left[p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right]. \tag{6.3}$$

According to Lemma 6.1, we can show that

$$\frac{\int g_t(\boldsymbol{\theta})\exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\right)\right]d\boldsymbol{\theta}}{\int \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\right)\right]d\boldsymbol{\theta}} = \widehat{g}_t + \frac{1}{n}B_t + O_p\left(\frac{1}{n^2}\right),$$

where

$$B_t = \frac{1}{2}\mathrm{tr}\left[\left(\widehat{h}^{(2)}\right)^{-1}\widehat{g}_t^{(2)}\right] - \frac{1}{2}\mathrm{vec}\left(\left(\widehat{h}^{(2)}\right)^{-1}\right)\widehat{h}^{(3)}\left(\widehat{h}^{(2)}\right)^{-1}\widehat{g}_t^{(1)}.$$

From (6.2) and (6.3), we can further show that

$$\begin{aligned}
\int\left[\ln p(\boldsymbol{\theta}) + \ln p(\mathbf{y}|\boldsymbol{\theta})\right]p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} &= \int[\ln p(\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} + \int[\ln p(\mathbf{y}|\boldsymbol{\theta})]p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} \\
&= \sum_{t=0}^{n}\left[\frac{\int g_t(\boldsymbol{\theta})\exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\right)\right]d\boldsymbol{\theta}}{\int \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\right)\right]d\boldsymbol{\theta}}\right] \\
&= \sum_{t=0}^{n}\widehat{g}_t + \frac{1}{n}\sum_{t=0}^{n}B_t + O_p\left(\frac{1}{n}\right). \qquad (6.4)
\end{aligned}$$

Note that

$$\sum_{t=0}^{n}g_t(\boldsymbol{\theta}) = \sum_{t=1}^{n}\ln p(\mathbf{y}_t|I_{t-1},\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta}) = \ln p(\mathbf{y}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta}).$$

Based on the definition of the posterior mode (i.e. $h_n(\widehat{\boldsymbol{\theta}}_m) = 0$), we know that

$$\sum_{t=0}^{n}\widehat{g} = \sum_{t=0}^{n}\widehat{g}(\widehat{\boldsymbol{\theta}}_m), \sum_{t=0}^{n}\widehat{g}^{(1)} = \sum_{t=0}^{n}\widehat{g}^{(1)}(\widehat{\boldsymbol{\theta}}_m) = \frac{\partial[\ln p(\mathbf{y}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_m} = -nh_n(\widehat{\boldsymbol{\theta}}_m) = 0,$$

$$\sum_{t=0}^{n}\widehat{g}^{(2)} = \sum_{t=0}^{n}\widehat{g}^{(2)}(\widehat{\boldsymbol{\theta}}_m) = \frac{\partial^2[\ln p(\mathbf{y}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_m} = -n\widehat{h}^{(2)}(\widehat{\boldsymbol{\theta}}_m) = -n\widehat{h}^{(2)}.$$

Hence, we can show that

$$\begin{aligned}
\sum_{t=0}^{n}B_t &= \sum_{t=0}^{n}\left\{\frac{1}{2}\mathrm{tr}\left[\left(\widehat{h}^{(2)}\right)^{-1}\widehat{g}_t^{(2)}\right] - \frac{1}{2}\mathrm{vec}\left(\left(\widehat{h}^{(2)}\right)^{-1}\right)\widehat{h}^{(3)}\left(\widehat{h}^{(2)}\right)^{-1}\widehat{g}_t^{(1)}\right\} \\
&= \frac{1}{2}\mathrm{tr}\left[\left(\widehat{h}^{(2)}\right)^{-1}\left(\ sum_{t=0}^{n}\widehat{g}_t^{(2)}\right)\right] - \frac{1}{2}\mathrm{vec}\left(\left(\widehat{h}^{(2)}\right)^{-1}\right)\widehat{h}^{(3)}\left(\widehat{h}^{(2)}\right)^{-1}\left(\sum_{t=0}^{n}\widehat{g}_t^{(1)}\right) \\
&= -\frac{n}{2}q + 0 = -\frac{n}{2}q.
\end{aligned}$$

Consequently, we can derive that

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

$$
\begin{aligned}
&= \sum_{t=0}^{n} \widehat{g}_t + \frac{1}{n}\sum_{t=0}^{n} B_t + O_p\left(\frac{1}{n}\right) \\
&= \ln p(\widehat{\boldsymbol{\theta}}_m) + \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_m) - \frac{q}{2} + O_p(n^{-1}). \qquad (6.5)
\end{aligned}
$$

Thus, we have,

$$
2\left[\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - \int [\ln p(\mathbf{y}, \boldsymbol{\theta})] p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}\right] = q + O_p(n^{-1}). \qquad (6.6)
$$

Naturally, the posterior test statistic can be derived as follows:

$$
\begin{aligned}
\mathrm{T}(\mathbf{y}, \boldsymbol{\vartheta}_0) &= \int \Delta\mathcal{L}\left(H_0, \boldsymbol{\vartheta}, \boldsymbol{\psi}\right) p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\
&= \int \left[4\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - 2\ln p(\mathbf{y}, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - 2D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)\right] p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\
&= \int \left[4\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - 4\ln p(\mathbf{y}, \boldsymbol{\vartheta}, \boldsymbol{\psi}) + 2\ln p(\mathbf{y}, \boldsymbol{\vartheta}, \boldsymbol{\psi}) - 2D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)\right] p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\
&= 2q + 2\left[D_c(\mathbf{y}) - D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)\right] + O_p(n^{-1}) \\
&= PLR1 + 2q + O_p(n^{-1}). \qquad (6.7)
\end{aligned}
$$

Similarly to the result in (6.6), when $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$, we can also derive that

$$
2\left[\ln p(\mathbf{y}, \widehat{\boldsymbol{\psi}}_{m0}|\boldsymbol{\vartheta}_0) - \int [\ln p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\vartheta}_0)] p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi}\right] = q_\psi + O_p(n^{-1}), \qquad (6.8)
$$

that is,

$$
2\int [\ln p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\vartheta}_0)] p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi} = 2\ln p(\mathbf{y}, \widehat{\boldsymbol{\psi}}_{m0}|\boldsymbol{\vartheta}_0) - q_\psi + O_p(n^{-1}).
$$

By (6.6) and (6.8), we can show that

$$
D_c(\mathbf{y}) = \int [\ln p(\mathbf{y}, \boldsymbol{\theta})] p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = \ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - \frac{q}{2} + O_p(n^{-1}),
$$

and

$$
\begin{aligned}
D_c(\mathbf{y}, \boldsymbol{\vartheta}_0) &= \int [\ln p(\mathbf{y}, \boldsymbol{\vartheta}_0)] p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi} \\
&= \int \left[\ln p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\vartheta}_0) + \ln p(\boldsymbol{\vartheta}_0)\right] p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi} \\
&= \int [\ln p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\vartheta}_0)] p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi} + \ln p(\boldsymbol{\vartheta}_0) \\
&= \ln p(\mathbf{y}, \widehat{\boldsymbol{\psi}}_{m0}|\boldsymbol{\vartheta}_0) + \ln p(\boldsymbol{\vartheta}_0) - \frac{q_\psi}{2} + O_p(n^{-1}) \\
&= \ln p(\mathbf{y}, \boldsymbol{\vartheta}_0, \widehat{\boldsymbol{\psi}}_{m0}) - \frac{q_\psi}{2} + O_p(n^{-1}). \qquad (6.9)
\end{aligned}
$$

Based on Lemma 3.1, using the Talyor expansion and the definition of the posterior mode, we can further show that

$$\begin{aligned}\ln p(\mathbf{y},\overline{\boldsymbol{\theta}}) &= \ln p(\mathbf{y},\widehat{\boldsymbol{\theta}}_m) + \frac{\partial \ln p(\mathbf{y},\widehat{\boldsymbol{\theta}}_m)}{\partial \boldsymbol{\theta}'}(\overline{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_m) + \frac{1}{2}(\overline{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_m)'\frac{\partial^2 \ln p(\mathbf{y},\widetilde{\boldsymbol{\theta}}_m)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}(\overline{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}_m) \\ &= \ln p(\mathbf{y},\widehat{\boldsymbol{\theta}}_m) + 0 + O_p(n^{-1})O_p(n)O_p(n^{-1}) \\ &= \ln p(\mathbf{y},\widehat{\boldsymbol{\theta}}_m) + O_p(n^{-1}),\end{aligned} \tag{6.10}$$

where $\widetilde{\boldsymbol{\theta}}_m$ is some intermediate value between $\overline{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_m$. Similarly, we can further show that

$$\ln p(\mathbf{y},\overline{\boldsymbol{\psi}}_0|\boldsymbol{\vartheta}_0) = \ln p(\mathbf{y},\widehat{\boldsymbol{\psi}}_{m0}|\boldsymbol{\vartheta}_0) + O_p(n^{-1}).$$

Hence, from (6.9), it is easy to show that

$$\begin{aligned}PLR1 &= 2\left[D_c(\mathbf{y}) - D_c(\mathbf{y},\boldsymbol{\vartheta}_0)\right] \\ &= 2\left[\ln p(\mathbf{y},\widehat{\boldsymbol{\theta}}_m) - \frac{q}{2} + O_p(n^{-1})\right] - 2\left[\ln p(\mathbf{y},\boldsymbol{\vartheta}_0,\widehat{\boldsymbol{\psi}}_{m0}) - \frac{q_\psi}{2} + O_p(n^{-1})\right] \\ &= 2\left[\ln p(\mathbf{y},\widehat{\boldsymbol{\theta}}_m) - \ln p(\mathbf{y},\boldsymbol{\vartheta}_0,\widehat{\boldsymbol{\psi}}_{m0})\right] - q_\vartheta + O_p(n^{-1}) \\ &= 2\left[\ln p(\mathbf{y},\overline{\boldsymbol{\theta}}) - \ln p(\mathbf{y},\boldsymbol{\vartheta}_0,\overline{\boldsymbol{\psi}}_0)\right] - q_\vartheta + O_p(n^{-1}) \\ &= PLR2 - q_\vartheta + O_p(n^{-1}).\end{aligned}$$

Appendix 2: Proof of Proposition 3.2

According to Lemma 3.2, we have

$$\mathrm{PLR1} + q_\vartheta = \mathrm{LR} + o_p(1) \text{ and } \mathrm{PLR2} = \mathrm{LR} + o_p(1)$$

under either H_0 or H_{L1}. Under H_0, it can be easily shown that

$$\mathrm{PLR1} + q_\vartheta = \mathrm{LR} + o_p(1) \xrightarrow{d} \chi^2(q_\vartheta) \text{ and } \mathrm{PLR2} = \mathrm{LR} + o_p(1) \xrightarrow{d} \chi^2(q_\vartheta).$$

Also the proof under H_0 can be seen as a special case under H_{L1} with $\boldsymbol{\delta} = 0$. Hence, we only establish the limiting distribution of the LR test under the local alternatives $H_{L1}: \boldsymbol{\vartheta}_n^* = \boldsymbol{\vartheta}_0 + \boldsymbol{\delta}/\sqrt{n}$ with the regularity conditions imposed.

Recall that $\boldsymbol{\theta}_n^* = (\boldsymbol{\vartheta}_n^{*\prime}, \boldsymbol{\psi}_n^{*\prime})'$ is the true value and $\boldsymbol{\theta}^* = (\boldsymbol{\vartheta}^{*\prime}, \boldsymbol{\psi}^{*\prime})'$ is the limit. Clearly, under H_{L1}, $\boldsymbol{\theta}^* = (\boldsymbol{\vartheta}_0', \boldsymbol{\psi}^{*\prime})'$. Also note that $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_0$ are the unconstrained MLE and the constrained MLE, respectively. When the model is correctly specified, we have $E\left[l_t^{(1)}\left(\boldsymbol{\theta}_n^*\right)|\mathcal{F}_\infty^{t-1}\right] =$

0 where $\mathcal{F}_{\infty}^{t-1} = \sigma(y_{t-1}, \ldots)$. Under Assumptions 1-9, it is straightforward to verify that Assumptions DG, OP', MX', SM, DM", NE", ID', PD', CN and DR in Theorem 7.4 (i) in Gallant and White (1988, page 125) hold. Then we can establish the following results:

(a) $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_n^* + o_p(1)$;

(b) $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_n^* + o_p(1)$;

(c) $\bar{\mathbf{H}}_n(\boldsymbol{\theta}) - \mathbf{H}_n(\boldsymbol{\theta}) = o_p(1)$ a.s. uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$;

(d) $\mathbf{J}_n^{-1/2}(\boldsymbol{\theta}_n^*)\sqrt{n}\bar{\mathbf{s}}_n(\boldsymbol{\theta}_n^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_q)$ and $\mathbf{H}_n(\boldsymbol{\theta}_n^*) + \mathbf{J}_n(\boldsymbol{\theta}_n^*) = 0$.

To be specific, following the proof of Lemma 7.3 in Gallant and White (1988), we can show (a)-(c). When the model is correctly specified, we can obtain the first part of (d) by verifying the CLT for a martingale difference sequence (MDS) using Assumption 3; the second part of (d) is the information identity, which can be easily justified.

Using the Taylor expansion and the first-order condition (FOC) $\bar{\mathbf{s}}_n\left(\widehat{\boldsymbol{\theta}}\right) = 0$, we can rewrite the LR testing statistic as follows

$$\begin{aligned}
\mathrm{LR} &= 2n\left[\mathcal{L}_n\left(\widehat{\boldsymbol{\theta}}_0\right) - \mathcal{L}_n\left(\widehat{\boldsymbol{\theta}}\right)\right] \\
&= 2n\bar{\mathbf{s}}_n\left(\widehat{\boldsymbol{\theta}}\right) + n\left(\widehat{\boldsymbol{\theta}}_0 - \widehat{\boldsymbol{\theta}}\right)'\bar{\mathbf{H}}_n\left(\widetilde{\boldsymbol{\theta}}\right)\left(\widehat{\boldsymbol{\theta}}_0 - \widehat{\boldsymbol{\theta}}\right) \\
&= \sqrt{n}\left(\widehat{\boldsymbol{\theta}}_0 - \widehat{\boldsymbol{\theta}}\right)'\bar{\mathbf{H}}_n\left(\widetilde{\boldsymbol{\theta}}\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_0 - \widehat{\boldsymbol{\theta}}\right),
\end{aligned}$$

where $\widetilde{\boldsymbol{\theta}}$ is a value between $\widehat{\boldsymbol{\theta}}_0$ and $\widehat{\boldsymbol{\theta}}$. We complete the proof by showing that

(i) $\bar{\mathbf{H}}_n\left(\widetilde{\boldsymbol{\theta}}\right) - \mathbf{H}_n(\boldsymbol{\theta}_n^*) = o_p(1)$;

(ii) $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_0 - \widehat{\boldsymbol{\theta}}\right) = \mathbf{H}_n^{-1}(\boldsymbol{\theta}_n^*)\mathbf{B}'[\mathbf{B}\mathbf{H}_n^{-1}(\boldsymbol{\theta}_n^*)\mathbf{B}']^{-1}\mathbf{B}\mathbf{H}_n^{-1/2}(\boldsymbol{\theta}_n^*)\mathbf{Z} + o_p(1)$, where $\mathbf{B} = (\mathbf{I}_{q_\vartheta}, \mathbf{0}_{q_\vartheta \times q_\psi})$ is a $q_\vartheta \times q$ matrix and $\mathbf{Z} := N\left(-\mathbf{H}^{1/2}(\boldsymbol{\theta}^*)\mathbf{C}_\delta, \mathbf{I}_q\right)$ with $\mathbf{C}_\delta = (\boldsymbol{\delta}', \mathbf{0}_{1\times q_\psi})'$.

By (a) and (b), we have $\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_n^* = o_p(1)$. Together with (c) it leads to (i). For (ii), we first show that $\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_n^*\right) = -\mathbf{H}_n(\boldsymbol{\theta}_n^*)\sqrt{n}\bar{\mathbf{s}}_n(\boldsymbol{\theta}_n^*) + o_p(1)$. For the unconstrained MLE $\widehat{\boldsymbol{\theta}}$, we have the FOC, $\bar{\mathbf{s}}_n\left(\widehat{\boldsymbol{\theta}}\right) = 0$. Taking the first-order Taylor expansion at the true value $\boldsymbol{\theta}_n^*$ leads to

$$0 = \bar{\mathbf{s}}_n\left(\widehat{\boldsymbol{\theta}}\right) = \bar{\mathbf{s}}_n(\boldsymbol{\theta}_n^*) + \bar{\mathbf{H}}_n\left(\boldsymbol{\theta}^\dagger\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_n^*\right),$$

where $\boldsymbol{\theta}^\dagger$ lies between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_n^*$. By (a) and Assumptions 4-5, we can show that $\bar{\mathbf{H}}_n\left(\boldsymbol{\theta}^\dagger\right) - \bar{\mathbf{H}}_n(\boldsymbol{\theta}_n^*) = o_p(1)$. Together with (d) it gives

$$\bar{\mathbf{H}}_n\left(\boldsymbol{\theta}^\dagger\right) - \mathbf{H}_n(\boldsymbol{\theta}_n^*) = \left[\bar{\mathbf{H}}_n\left(\boldsymbol{\theta}^\dagger\right) - \bar{\mathbf{H}}_n(\boldsymbol{\theta}_n^*)\right] + \left[\bar{\mathbf{H}}_n(\boldsymbol{\theta}_n^*) - \mathbf{H}_n(\boldsymbol{\theta}_n^*)\right] = o_p(1).$$

Then we obtain

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_n^*\right)=-\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\sqrt{n}\bar{\mathbf{s}}_n\left(\boldsymbol{\theta}_n^*\right)+o_p\left(1\right). \tag{6.11}$$

For the constrained MLE $\widehat{\boldsymbol{\theta}}_0$, following Davidson (2000, page 300), we obtain that

$$\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_n^*\right)=&-\left[\bar{\mathbf{H}}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)-\bar{\mathbf{H}}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\bar{\mathbf{H}}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\bar{\mathbf{H}}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\right]\sqrt{n}\bar{\mathbf{s}}_n\left(\boldsymbol{\theta}_n^*\right)\\
&-\bar{\mathbf{H}}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\bar{\mathbf{H}}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{C}_\delta+o_p\left(1\right)\\
=&-\left[\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)-\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\right]\sqrt{n}\bar{\mathbf{s}}_n\left(\boldsymbol{\theta}_n^*\right)\\
&-\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{C}_\delta+o_p\left(1\right).
\end{aligned} \tag{6.12}$$

where we use the result in (c) in the last step. Using (6.11)-(6.12), we have

$$\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_0-\widehat{\boldsymbol{\theta}}\right)=&\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}^*\right)-\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_n^*\right)\\
=&\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\sqrt{n}\bar{\mathbf{s}}_n\left(\boldsymbol{\theta}_n^*\right)\\
&-\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}_n^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{C}_\delta+o_p\left(1\right)\\
=&\mathbf{H}^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}_n^*\right)\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}_n^*\right)\mathbf{J}^{1/2}\left(\boldsymbol{\theta}_n^*\right)N\left(0,\mathbf{I}_q\right)\\
&-\mathbf{H}^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}_n^*\right)\mathbf{H}^{1/2}\left(\boldsymbol{\theta}_n^*\right)\mathbf{C}_\delta+o_p\left(1\right)\\
=&\mathbf{H}^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}_n^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}_n^*\right)\mathbf{Z}+o_p\left(1\right),
\end{aligned}$$

with $\mathbf{Z}:=N\left(-\mathbf{H}^{1/2}\left(\boldsymbol{\theta}^*\right)\mathbf{C}_\delta,\mathbf{I}_q\right)$, where we have used the definitions of $\mathbf{H}\left(\cdot\right)$ and $\mathbf{J}\left(\cdot\right)$ and the result in (d) in the third step, and the fact $\boldsymbol{\theta}_n^*\to\boldsymbol{\theta}^*$ as $n\to\infty$ in defining the normal random variable $\mathbf{Z}$.s Lastly, we have

$$\begin{aligned}
\mathrm{LR}=&2n\left[\mathcal{L}_n\left(\widehat{\boldsymbol{\theta}}_0\right)-\mathcal{L}_n\left(\widehat{\boldsymbol{\theta}}\right)\right]\\
=&\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_0-\widehat{\boldsymbol{\theta}}\right)'\mathbf{H}\left(\boldsymbol{\theta}^*\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_0-\widehat{\boldsymbol{\theta}}\right)+o_p\left(1\right)\\
\xrightarrow{d}&\mathbf{Z}'\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}^*\right)\mathbf{Z}\\
=&\mathbf{Z}'\mathbf{P}\left(\boldsymbol{\theta}^*\right)\mathbf{Z},
\end{aligned}$$

where $\mathbf{P}\left(\boldsymbol{\theta}^*\right)=\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}^*\right)\mathbf{B}'[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{B}']^{-1}\mathbf{B}\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}^*\right)$. Clearly, $\mathbf{P}\left(\boldsymbol{\theta}^*\right)\mathbf{P}\left(\boldsymbol{\theta}^*\right)=\mathbf{P}\left(\boldsymbol{\theta}^*\right)$ and $\mathrm{tr}\left(\mathbf{P}\left(\boldsymbol{\theta}^*\right)\right)=q_\vartheta$. Then LR $\to_d\chi^2\left(q_\vartheta,\tau\right)$ with $\tau=\mathbf{C}_\delta'[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{B}']^{-1}\mathbf{C}_\delta=\boldsymbol{\delta}'[\mathbf{H}^{-1}]_{11}^{-1}\boldsymbol{\delta}$, $\mathbf{H}=\mathbf{H}\left(\boldsymbol{\theta}^*\right)$, and $[\mathbf{H}^{-1}]_{11}$ being the $q_\vartheta\times q_\vartheta$ submatrix of $\mathbf{H}^{-1}$ corresponding to $\boldsymbol{\vartheta}$.

Appendix 3: Proof of Theorem 3.1

The proof is a direct result of Lemmas 3.1 and 3.2. Note that

$$\mathrm{BF}_{01}(\mathrm{PLR1}) = (cn+1)^{\frac{q_\vartheta}{2}} \exp\left\{-\frac{\mathrm{PLR1}+q_\vartheta}{2(cn+1)/(cn)}\right\}.$$

Hence, we can show that

$$\ln \mathrm{BF}_{01}(\mathrm{PLR1}) = \frac{q_\vartheta}{2}\ln(cn+1) - \left\{\frac{\mathrm{PLR1}+q_\vartheta}{2(cn+1)/(cn)}\right\}. \tag{6.13}$$

Under H_0, $\mathrm{PLR1} = O_p(1)$, $\frac{cn}{cn+1} = O(1)$, and $\frac{q_\vartheta}{2}\ln(cn+1) = O(\ln n)$. Hence, $\ln\mathrm{BF}^J_{01}(\mathrm{LR}) = O_p(\ln n)$. Under H_{L1}, we have $\boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}_0 = O(n^{-1/2+a})$ for some $a > 0$. Based on Lemma 3.1, it can be shown that $\mathrm{PLR1} = \mathrm{LR} + o_p(1)$.

Furthermore, we can write

$$\mathrm{LR} = 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) - \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)\right] = 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) - \ln p(\mathbf{y}|\boldsymbol{\theta}_n^*)\right] + \left[\ln p(\mathbf{y}|\boldsymbol{\theta}_n^*) - \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)\right]. \tag{6.14}$$

By the standard ML theory, for the first term in (6.14), we have

$$2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) - \ln p(\mathbf{y}|\boldsymbol{\theta}_n^*)\right] = O_p(1).$$

In the following, we only need to derive the probability order of $2\left[\ln p(\mathbf{y}|\boldsymbol{\theta}_n^*) - \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)\right]$.

Based on the first-order Taylor expansion, we can show that

$$\frac{\partial \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)}{\partial \boldsymbol{\theta}} = \frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \ln p(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_1)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_n^*),$$

where $\widetilde{\boldsymbol{\theta}}_1$ is some intermediate value between $\boldsymbol{\theta}_n^*$ and $\widehat{\boldsymbol{\theta}}_0$. Hence, we can further show that

$$0 = \frac{\partial \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)}{\partial \boldsymbol{\psi}} = \frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\psi}} + \frac{\partial^2 \ln p(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_1)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}_0 - \boldsymbol{\vartheta}_n^*) + \frac{\partial^2 \ln p(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_1)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}(\widehat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_n^*).$$

From the above formula, we can get that

$$\begin{aligned}\widehat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_n^* &= \left[\frac{\partial \ln p(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_1)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right]^{-1}\left[\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\psi}} + \frac{\partial^2 \ln p(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_1)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\vartheta}'}(\boldsymbol{\vartheta}_0 - \boldsymbol{\vartheta}_n^*)\right] \\ &= O_p(n^{-1})\left[O_p(n^{\frac{1}{2}}) + O_p(n)O(n^{-\frac{1}{2}+\alpha})\right] = O_p(n^{\frac{1}{2}+\alpha}).\end{aligned}$$

Based on the second-order Taylor expansion, it can be shown as

$$2[\ln p(\mathbf{y}|\boldsymbol{\theta}_n^*) - \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)]$$

$$
\begin{aligned}
&= 2\frac{\partial \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)}{\partial \boldsymbol{\theta}'}\left(\boldsymbol{\theta}_n^* - \widehat{\boldsymbol{\theta}}_0\right) + \left(\boldsymbol{\theta}_n^* - \widehat{\boldsymbol{\theta}}_0\right)'\left[\frac{\partial^2 \ln p(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_2)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right]\left(\boldsymbol{\theta}_n^* - \widehat{\boldsymbol{\theta}}_0\right) \\
&= 2\frac{\partial \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)}{\partial \boldsymbol{\vartheta}'}\left(\boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}_0\right) + \left(\boldsymbol{\theta}_n^* - \widehat{\boldsymbol{\theta}}_0\right)'\left[\frac{\partial^2 \ln p(\mathbf{y}|\widetilde{\boldsymbol{\theta}}_2)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right]\left(\boldsymbol{\theta}_n^* - \widehat{\boldsymbol{\theta}}_0\right) \\
&= O_p(n)O_p(n^{-\frac{1}{2}+\alpha}) + O_p(n^{-\frac{1}{2}+\alpha})O_p(n)O_p(n^{-\frac{1}{2}+\alpha}) \\
&= O_p(n^{-\frac{1}{2}+\alpha}) + O_p(n^{2\alpha}) = O_p(n^{\max\{-\frac{1}{2}+\alpha, 2\alpha\}}),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\theta}}_2$ is some intermediate value between $\boldsymbol{\theta}_n^*$ and $\widehat{\boldsymbol{\theta}}_0$.

Hence, from Equation (6.13), we can get the order of LR$= O_p(\ln n) - O_p(n^{\max\{-\frac{1}{2}+\alpha,2\alpha\}})$ such that LR $= -O_p(n^{\max\{-\frac{1}{2}+\alpha,2\alpha\}})$, $\ln\mathrm{BF}_{01}^J(\mathrm{LR}) \to -\infty$, and $\mathrm{BF}_{01}^J(\mathrm{LR}) \to 0$ with probability approaching one. In this case, the proposed LR-test-based BFs have the consistency property.

From Equation (6.26) in Lemma 3.2, we can show that

$$
\begin{aligned}
\mathrm{PLR1} &= 2\left[D_c(\mathbf{y}) - D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)\right] \\
&= \mathrm{PLR2} - q_{\vartheta} + O_p(n^{-1}) \\
&= \mathrm{LR} - \left[\ln p(\widehat{\boldsymbol{\theta}}) - \ln p(\widehat{\boldsymbol{\theta}}_0)\right] - q_{\vartheta} + O_p(n^{-1}) \\
&= -O_p(n^{\max\{-\frac{1}{2}+\alpha,2\alpha\}}). \qquad (6.15)
\end{aligned}
$$

Hence, from Equation (6.15), we can get the order of PLR1$= -O_p(n^{\max\{-\frac{1}{2}+\alpha,2\alpha\}})$, $\ln\mathrm{BF}_{01}(\mathrm{PLR1}) \to -\infty$, and $\mathrm{BF}_{01}(\mathrm{PLR1}) \to 0$ with probability approaching one. Using the same approach, we can show that $\ln\mathrm{BF}_{01}(\mathrm{PLR2}) \to -\infty$ and $\mathrm{BF}_{01}(\mathrm{PLR2}) \to 0$ with probability approaching one. In this case, both the proposed LR-test-based BFs and PLR-test-based BFs have the consistency property.

Appendix 4: Proof of Proposition 3.3

First, from Remark 3.6, we note that

$$
\begin{aligned}
\mathbf{V}\left(\overline{\boldsymbol{\theta}}\right) &= E\left[(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})'|\mathbf{y}\right] \\
&= E\left[\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\right)'|\mathbf{y}\right] \\
&= E\left[\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)'|\mathbf{y}\right] + E\left[\left(\widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\right)\left(\widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\right)'|\mathbf{y}\right] \\
&\quad + E\left[\left(\widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)'|\mathbf{y}\right] + E\left[\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)\left(\widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\right)'|\mathbf{y}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbf{V}\left(\widehat{\boldsymbol{\theta}}\right) + E\left[\left(\widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\right)\left(\widehat{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\right)' |\mathbf{y}\right] \\
&= \mathbf{V}\left(\widehat{\boldsymbol{\theta}}\right) + o_p(n^{-1/2})o_p(n^{-1/2}) \\
&= \mathbf{V}\left(\widehat{\boldsymbol{\theta}}\right) + o_p(n^{-1}) \\
&= -\frac{1}{n}\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\theta}}) + o_p\left(n^{-1}\right) = O_p\left(\frac{1}{n}\right),
\end{aligned} \tag{6.16}
$$

and that

$$
\begin{aligned}
\mathbf{V}_{\vartheta\vartheta}^0 &= E\left[\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right)\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right)' |\mathbf{y}\right] \\
&= E\left[\left(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}} + \overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}} + \overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' |\mathbf{y}\right] \\
&= \mathbf{V}_{\vartheta\vartheta}\left(\overline{\boldsymbol{\vartheta}}\right) + \left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \overline{\boldsymbol{\vartheta}}\right)' + \left(\overline{\boldsymbol{\vartheta}} - \overline{\boldsymbol{\vartheta}}\right)\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' + \left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' \\
&= \mathbf{V}_{\vartheta\vartheta}\left(\overline{\boldsymbol{\vartheta}}\right) + \left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)'.
\end{aligned} \tag{6.17}
$$

Furthermore, under the local hypothesis and the standard ML theory, we have

$$
\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = \widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_n^* + \boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}_0 = O_p(n^{-1/2}) + O_p(n^{-1/2}) = O_p(n^{-1/2}).
$$

Hence, we can further derive that

$$
\begin{aligned}
\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' &= \left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}} + \widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}} + \widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' \\
&= \left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}}\right)\left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}}\right)' + \left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' \\
&\quad + \left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}}\right)\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' + \left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}}\right)' \\
&= o_p(n^{-1/2})o_p(n^{-1/2}) + \left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' + 2o_p(n^{-1/2})O_p(n^{-1/2}) \\
&= \left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' + o_p(n^{-1}).
\end{aligned} \tag{6.18}
$$

Based on Equations (6.16), (6.17), and (6.18), we can further derive that

$$
\begin{aligned}
\mathrm{PWald} &= \int_{\Theta} \left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right)' \left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right]\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right) p(\boldsymbol{\vartheta}|\mathbf{y}) d\boldsymbol{\vartheta} \\
&= \mathrm{tr}\left\{\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right] E\left[\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right)\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right)' |\mathbf{y}\right]\right\} \\
&= \mathrm{tr}\left\{\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right]\left[\mathbf{V}_{\vartheta\vartheta}\left(\overline{\boldsymbol{\theta}}\right) + \left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' + o_p(n^{-1})\right]\right\} \\
&= q_{\vartheta} + \mathrm{tr}\left\{\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right]\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)'\right\} + \mathrm{tr}\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\theta}\right) o_p(n^{-1})\right]
\end{aligned}
$$

$$
\begin{aligned}
&= q_\vartheta + \operatorname{tr}\left\{\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right]\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'\right\} + O_p(n)o_p(n^{-1}) \\
&= q_\vartheta + \operatorname{tr}\left\{\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right]\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'\right\} + o_p(1) \\
&= q_\vartheta + \operatorname{tr}\left\{\left[n\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right] n\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'\right\} + o_p(1) \\
&= q_\vartheta + \operatorname{tr}\left\{\left[n\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right] n\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'\right\} + o_p(1) \\
&= q_\vartheta + \operatorname{tr}\left\{\left[-[\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\theta}})]_{\vartheta\vartheta} + o_p\left(1\right)\right]^{-1} n\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'\right\} + o_p(1).
\end{aligned}
$$

From the above derivation, it is easy to show that

$$
\begin{aligned}
\text{PWald} - q_\vartheta &= \operatorname{tr}\left\{\left[-[\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\theta}})]_{\vartheta\vartheta}\right]^{-1} n\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'\right\} \\
&= n\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'\left[-[\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\theta}})]_{\vartheta\vartheta}\right]^{-1}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right) + o_p\left(1\right) = \text{Wald} + o_p(1).
\end{aligned}
$$

Hence, under the local hypothesis, we can show that

$$
\text{PWald} = \int \left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)'\left[\mathbf{V}_{\vartheta\vartheta}^{-1}(\overline{\boldsymbol{\theta}})\right]\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right) p(\boldsymbol{\vartheta}|\mathbf{y})d\boldsymbol{\vartheta} \xrightarrow{d} \chi^2(q_\vartheta, \tau). \tag{6.19}
$$

We only prove the limiting distribution of Wald under H_{L1}. First, we have

$$
\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_n^*\right) \xrightarrow{d} \mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{J}^{1/2}\left(\boldsymbol{\theta}^*\right) N\left(\mathbf{0}, \mathbf{I}_q\right) \overset{d}{=} \boldsymbol{\delta} + \mathbf{B}\mathbf{H}^{-1/2}\left(\boldsymbol{\theta}^*\right) N\left(\mathbf{0}, \mathbf{I}_q\right),
$$

where $\mathbf{H}\left(\boldsymbol{\theta}^*\right) = \lim_{n\to\infty}\mathbf{H}_n\left(\boldsymbol{\theta}_n^*\right)$ and $\mathbf{J}\left(\boldsymbol{\theta}^*\right) = \lim_{n\to\infty}\mathbf{J}_n\left(\boldsymbol{\theta}_n^*\right)$. Here we have used the fact that $\lim_{n\to\infty}[\mathbf{H}_n\left(\boldsymbol{\theta}_n^*\right) + \mathbf{J}_n\left(\boldsymbol{\theta}_n^*\right)] = 0$ when there is no model misspecification. It follows that

$$
\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right) = \sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_n^*\right) + \boldsymbol{\delta} \xrightarrow{d} N\left(\boldsymbol{\delta}, \mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{B}'\right).
$$

Then

$$
\begin{aligned}
\text{Wald} &= n\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'[\mathbf{B}\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\theta}}\right)\bar{\mathbf{J}}_n\left(\widehat{\boldsymbol{\theta}}\right)\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\theta}}\right)\mathbf{B}']^{-1}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right) \\
&= \sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'\left[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{B}'\right]^{-1}\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right) + o_p\left(1\right) \\
&= \mathbf{Z}_0'\mathbf{Z}_0 + o_p\left(1\right) \xrightarrow{d} \chi^2\left(q_\vartheta, \tau\right),
\end{aligned}
$$

where

$$
\mathbf{Z}_0 = \left[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{B}'\right]^{-1/2}\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right) \xrightarrow{d} N\left(\left[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{B}'\right]^{-1/2}\boldsymbol{\delta}, \mathbf{I}_q\right),
$$

and $\tau = \boldsymbol{\delta}'[\mathbf{B}\mathbf{H}^{-1}\left(\boldsymbol{\theta}^*\right)\mathbf{B}']^{-1}\boldsymbol{\delta}$.

Appendix 5: Proof of Theorem 3.2

First, when $\boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}_0 = O(n^{-1/2+a})$ for some $a > 0$, based on the ML theory, we have

$$\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 = \widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_n^* + \boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}_0 = O_p(n^{-1/2}) + O_p(n^{-1/2+a}) = O_p(n^{-1/2+a}). \tag{6.20}$$

By Remark 3.2, we can further show that

$$\begin{aligned}
\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' &= \left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}} + \widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}} + \widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' \\
&= \left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}}\right)\left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}}\right)' + \left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' \\
&\quad + \left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}}\right)\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)' + \left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \widehat{\boldsymbol{\vartheta}}\right)' \\
&= o_p(n^{-1/2})o_p(n^{-1/2}) + O_p(n^{-1/2+a})O_p(n^{-1/2+a}) + 2o_p(n^{-1/2})O_p(n^{-1/2+a}) \\
&= O_p(n^{-1+2a}).
\end{aligned}$$

Note that $E\left[(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)'|\mathbf{y}\right] = \mathbf{V}_{\vartheta\vartheta}\left(\overline{\boldsymbol{\theta}}\right) + \left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)'$, and $\mathbf{V}_{\vartheta\vartheta}\left(\overline{\boldsymbol{\theta}}\right) = O_p(n^{-1})$ in (6.16). Hence, we can derive that

$$\begin{aligned}
\text{PWald} &= \int_{\Theta_\vartheta} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)'\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right](\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)\, p(\boldsymbol{\vartheta}|\mathbf{y})d\boldsymbol{\vartheta} \\
&= \operatorname{tr}\left\{\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right] E\left[(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)'|\mathbf{y}\right]\right\} \\
&= \operatorname{tr}\left\{\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right]\left[\mathbf{V}_{\vartheta\vartheta}\left(\overline{\boldsymbol{\theta}}\right) + \left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)'\right]\right\} \\
&= q_\vartheta + \operatorname{tr}\left\{\left[\mathbf{V}_{\vartheta\vartheta}^{-1}\left(\overline{\boldsymbol{\theta}}\right)\right]\left[\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)\left(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right)'\right]\right\} \\
&= q_\vartheta + O_p(n)O_p(n^{-1+2a}) = q_\vartheta + O_p(n^{2a}) = O_p(n^{2a}).
\end{aligned}$$

As a consequence, we have

$$\ln \text{BF}_{01}(\text{PWald}) = \frac{q_\vartheta}{2}\ln(cn+1) - \left\{\frac{\text{PWald} - q_\vartheta}{2(cn+1)/(cn)}\right\}.$$

Hence, under H_0, we can show that $\text{PWald} = O_p(1)$, $\frac{cn}{cn+1} = O(1)$, $\frac{q_\vartheta}{2}\ln(cn+1) = O(\ln n)$, and $\text{lnBF}_{01}(\text{PWald}) = O_p(\ln n)$. Under H_{L1}, when $\boldsymbol{\vartheta}_n^* - \boldsymbol{\vartheta}_0 = O(n^{-1/2+a})$, it can be shown that $\text{PWald} = O_p(n^{2a})$ such that $\text{lnBF}_{01}(\text{PWald}) = O_p(\ln n) + O_p(-n^{2a}) = O_p(-n^{2a})$ with $a > 0$.

References

Abadie, A. (2020). Statistical non-significance in empirical economics. *American Economic Review: Insights* 2(2), 193-208.

Aitkin, M., Boys, R. J. and Chadwick, T. (2005). Bayesian point null hypothesis testing via the posterior likelihood ratio. *Statistics and Computing* 15(3), 217-230.

Amrhein, V., Greenland, S. and McShane, B. (2019). Scientists rise up against statistical significance. *Nature* 567(7748), 305-307.

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

Andrews, I. and Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review* 109(8), 2766-2794.

Baker, M. (2015). Over half of psychology studies fail reproducibility test. *Nature News*.

Benjamin, D. J., Berger, J. O., ... and Cesarini, D. (2018). Redefine statistical significance. *Nature Human Behaviour* 2(1), 6-10.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the Chi-square test. *Journal of the American Statistical Association* 33, 526-536.

Bernardo, J. M. and Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review* 7, 351-372.

Brodeur, A., Cook, N. and Heyes, A. G. (2018). Methods matter: p-hacking and causal inference in economics. *Working paper* No. 11796, IZA.

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

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

Davidson, J. (2000). *Econometric Theory.* Wiley-Blackwell.

Davidson, R. and MacKinnon, J. G. (1987). Implicit alternatives and the local power of test statistics. *Econometrica* 55(6), 1305-1329.

Dorsey, R.E. and Mayer, W. J. (2000). Detection of spurious maxima through random draw tests and specification test. *Computational Economics* 16, 237-256.

Elliott, G., Kudrin, N. and Wüthrich, K. (2022). Detecting p-hacking. *Econometrica* 90(2), 887-906.

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

Gallant, A. R. and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* Blackwell.

Goodman, S. N. (2008). A dirty dozen: Twelve p-value misconceptions. *Seminars in Hematology* 45(3), 135-140.

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *Journal of Finance* 72(4), 1399-1440.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T. and Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology* 13(3), e1002106.

Herbst, E. and Schorfheide, F. (2014). Sequential Monte Carlo sampling for DSGE models. *Journal of Applied Econometrics* 29(7), 1073-1098.

Hu, J. and Johnson, V. E. (2009). Bayesian model selection using test statistics. *Journal of the Royal Statistical Society: Series B* 71(1), 143-158.

Johnson, V. E. (2005). Bayes factors based on test statistics. *Journal of the Royal Statistical Society: Series B* 67(5), 689-701.

Johnson, V. E. (2008). Properties of Bayes factors based on test statistics. *Scandinavian Journal of Statistics* 35(2), 354-368.

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

Kim, J. H., Ahmed, K. and Ji, P. I. (2018). Significance testing in accounting research: A critical evaluation based on evidence. *Abacus* 54(4), 524-546.

Kim, J. H. and Ji, P. I. (2015). Significance testing in empirical finance: A critical review and assessment. *Journal of Empirical Finance* 34, 1-14.

Lee, L. F. (2005). Classical inference with ML and GMM estimates with various rates of convergence. *Working paper*, Ohio State University.

Li, Y. (2023). Hypothesis testing statistics based on posterior output with applications in financial econometrics. In *Financial Econometrics –Theory and Applications* edited by Shi, S., Wang, X. and Zeng, T., Cambridge University Press, forthcoming.

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

Li, Y., Wang, N. L. and Yu, J. (2023). Improved marginal likelihood estimation via power posteriors and importance sampling. *Journal of Econometrics*, forthcoming.

Li, Y., Wang, N. L., Yu, J. and Zeng, T. (2022). Deviance information criterion for Bayesian model selection: Justification and variation. *Working paper*, Singapore Management University.

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

Li, Y., Yu, J. and Zeng, T. (2018). Specification tests based on MCMC output. *Journal of Econometrics* 207, 237-260.

Li, Y., Yu, J. and Zeng, T. (2020). Deviance information criterion for latent variable models and misspecified models. *Journal of Econometrics* 216(2), 450-493.

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

Liu, X.B., Li, Y., Zeng, T. and Yu, J. (2022). Posterior-based Wald-type statistics for hypothesis testing. *Journal of Econometrics* 230(1), 83-113.

Marden, J. I. (2000). Hypothesis testing: From p values to Bayes factors. *Journal of the American Statistical Association* 95(452), 1316-1320.

Mergner, S. and Bulla, J. (2008). Time-varying beta risk of Pan-European industry portfolios: A comparison of alternative modelling techniques. *The European Journal of Finance* 14(8), 771-802.

Meyer, R. and Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal* 3, 198-215.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review* 16, 225-237.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6, 461-464.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19, 425-442.

Simmons, J. P., Nelson, L. D. and Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11), 1359-1366.

Simonsohn, U., Nelson,L.D. and Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology General* 143(2), 534-547.

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

Wasserstein, R. L. and Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician* 70, 129-133.

Wasserstein, R. L., Schirm, A. L. and Lazar, N. A. (2019). Moving to a world beyond "p <0.05". *The American Statistician* 73, 1-19.

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

Yu, J. (2005). On leverage in a stochastic volatility models. *Journal of Econometrics* 127, 165-178.

Online Supplement: Proofs of Lemmas

Supplement 1: Proof of Lemma 3.1

To prove this lemma, we first restate a lemma developed in Li et al. (2020) about the high-order stochastic expansion. This lemma is also used to prove Theorem 3.1.

Without any loss of generality, for any function $f(\boldsymbol{\theta})$, let $f^{(j)}(\boldsymbol{\theta})$ be the jth order derivative of $f(\boldsymbol{\theta})$ for $j = 1, 2, 3, 4, 5$. When there is no confusion, for convenience, we simply let $\hat{f}$ be the value of function f evaluated at $\widehat{\boldsymbol{\theta}}$, i.e., $\hat{f} := f\left(\widehat{\boldsymbol{\theta}}\right)$ and for convenience of exposition, we write $\frac{\partial^d}{\partial\theta_{j_1}\partial\theta_{j_2}\cdots\partial\theta_{j_d}} f(\boldsymbol{\theta})$ as $f_{j_1\cdots j_d}$, and let $\hat{f}_{j_1\cdots j_d} := f_{j_1\cdots j_d}\left(\widehat{\boldsymbol{\theta}}\right)$. The lemma developed in Li et al. (2020) states as below.

Lemma 6.1 *For some real-valued function $g(\boldsymbol{\theta})$, if both $(\{h_n\}, g \times b_D)$ and $(\{h_n\}, b_D)$ satisfy the analytical assumptions for the stochastic Laplace method, then we have*

$$\frac{\int g(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) \exp(-nh_n(\boldsymbol{\theta}))\, d\boldsymbol{\theta}}{\int b_D(\boldsymbol{\theta}) \exp(-nh_n(\boldsymbol{\theta}))\, d\boldsymbol{\theta}} = \hat{g} + \frac{1}{n}B + O_p\left(\frac{1}{n^2}\right),$$

where

$$B = \frac{1}{2} tr\left[\left(\hat{h}^{(2)}\right)^{-1} \hat{g}^{(2)}\right] + \left(\hat{g}^{(1)}\right)' \left(\hat{h}^{(2)}\right)^{-1} \frac{\hat{b}_D^{(1)}}{\hat{b}_D} - \frac{1}{2} vec\left(\left(\hat{h}^{(2)}\right)^{-1}\right) \hat{h}^{(3)} \left(\hat{h}^{(2)}\right)^{-1} \hat{g}^{(1)}.$$

For more details about this lemma, one can refer to Li et al. (2020).

The proof of Lemma 3.1 is similar to the proof of Lemma 3.2 in Li et al. (2020). Note that

$$\begin{aligned} \overline{\boldsymbol{\theta}} &= \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = \int \boldsymbol{\theta} \frac{p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y})} d\boldsymbol{\theta} \\ &= \frac{\int \boldsymbol{\theta} \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{p(\mathbf{y})} = \frac{\int \boldsymbol{\theta} \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{\int \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}. \end{aligned} \tag{6.21}$$

From this formula, we can let

$$g_i(\boldsymbol{\theta}) = \theta_i = a_i'\boldsymbol{\theta}, i = 1, \cdots, q, b_D(\boldsymbol{\theta}) = 1, h_n(\boldsymbol{\theta}) = -\frac{1}{n} \ln\left[p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right],$$

where a_i is an indicator vector with the ith element being one and the others being zero. Hence, according to Lemma 6.1, it is easy to show that for $i = 1, \cdots, q$,

$$\overline{\boldsymbol{\theta}}_i = \frac{\int g_i(\boldsymbol{\theta}) \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}}{\int \exp\left[\ln\left(p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\right)\right] d\boldsymbol{\theta}} = g_i(\widehat{\boldsymbol{\theta}}_m) + O_p(n^{-1})$$

$$= a_i'\widehat{\boldsymbol{\theta}}_m + O_p(n^{-1}) = \widehat{\theta}_{mi} + O_p(n^{-1}).$$

Similarly, under the null hypothesis, it is also easy to show that

$$\overline{\boldsymbol{\psi}}_0 = \widehat{\boldsymbol{\psi}}_{m0} + O_p(n^{-1}).$$

Supplement 2: Proof of Lemma 3.2

From (6.6) and (6.10) in the proof of Theorem 3.1, we get that

$$\begin{aligned} 2D_c(\mathbf{y}) &= \int \ln p(\mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \\ &= 2\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_m) - q + O_p(n^{-1}) = 2\ln p(\mathbf{y}, \overline{\boldsymbol{\theta}}) - q + O_p(n^{-1}). \end{aligned} \tag{6.22}$$

By Remark 3.3, the Taylor expansion, and the definition of MLE, we have

$$\begin{aligned} \ln p(\mathbf{y}|\overline{\boldsymbol{\theta}}) &= \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + \frac{\partial \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\overline{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) + \frac{1}{2}(\overline{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})' \frac{\partial^2 \ln p(\mathbf{y}, \widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\overline{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) \\ &= \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + 0 + O_p(n^{-1})O_p(n)O_p(n^{-1}) = \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + O_p(n^{-1}), \end{aligned} \tag{6.23}$$

where $\widetilde{\boldsymbol{\theta}}$ is an intermediate value between $\overline{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$. Furthermore, for the prior distribution, again using the Taylor expansion, we can also get that

$$\begin{aligned} \ln p(\overline{\boldsymbol{\theta}}) &= \ln p(\widehat{\boldsymbol{\theta}}) + \frac{\partial \ln p(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\overline{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) + \frac{1}{2}(\overline{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})' \frac{\partial^2 \ln p(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\overline{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) \\ &= \ln p(\widehat{\boldsymbol{\theta}}) + O_p(1)O_p(n^{-1}) + O_p(n^{-1})O_p(1)O_p(n^{-1}) = \ln p(\widehat{\boldsymbol{\theta}}) + O_p(n^{-1}). \end{aligned} \tag{6.24}$$

From Equations (6.23) and (6.24), we have

$$\begin{aligned} 2D_c(\mathbf{y}) &= 2\ln p(\mathbf{y}, \overline{\boldsymbol{\theta}}) - q + O_p(n^{-1}) \\ &= 2\left[\ln p(\mathbf{y}|\overline{\boldsymbol{\theta}}) + \ln p(\overline{\boldsymbol{\theta}})\right] - q + O_p(n^{-1}) \\ &= 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + \ln p(\widehat{\boldsymbol{\theta}})\right] - q + O_p(n^{-1}) \\ &= 2\ln p(\mathbf{y}, \widehat{\boldsymbol{\theta}}) - q + O_p(n^{-1}). \end{aligned} \tag{6.25}$$

Similar to the above derivation, according to Equations (6.9) and (6.25), we have

$$\begin{aligned} D_c(\mathbf{y}, \boldsymbol{\vartheta}_0) &= \int [\ln p(\mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\vartheta}_0)] p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi} \\ &= \int \left[\ln p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\vartheta}_0) + \ln p(\boldsymbol{\vartheta}_0)\right] p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi} \end{aligned}$$

$$
\begin{aligned}
&= \int [\ln p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\vartheta}_0)] p(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\vartheta}_0) d\boldsymbol{\psi} + \ln p(\boldsymbol{\vartheta}_0) \\
&= \ln p(\mathbf{y}, \widehat{\boldsymbol{\psi}}_{m0}|\boldsymbol{\vartheta}_0) + \ln p(\boldsymbol{\vartheta}_0) - \frac{q_\psi}{2} + O_p(n^{-1}) \\
&= \ln p(\mathbf{y}, \widehat{\boldsymbol{\psi}}_{m0}|\boldsymbol{\vartheta}_0) + \ln p(\boldsymbol{\vartheta}_0) - \frac{q_\psi}{2} + O_p(n^{-1}).
\end{aligned} \tag{6.26}
$$

From Equations (6.25) and (6.26), we can show that

$$
\begin{aligned}
\text{PLR1} &= 2\left[D_c(\mathbf{y}) - D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)\right] \\
&= \text{PLR2} - q_\vartheta + O_p(n^{-1}) \\
&= 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + \ln p(\widehat{\boldsymbol{\theta}})\right] - q + O_p(n^{-1}) - 2\left[\ln p(\mathbf{y}, \widehat{\boldsymbol{\psi}}_0|\boldsymbol{\vartheta}_0) + \ln p(\boldsymbol{\vartheta}_0) - \frac{q_\psi}{2} + O_p(n^{-1})\right] \\
&= 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + \ln p(\widehat{\boldsymbol{\theta}})\right] - 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\psi}}_0, \boldsymbol{\vartheta}_0) + \ln p(\widehat{\boldsymbol{\psi}}_0|\boldsymbol{\vartheta}_0) + \ln p(\boldsymbol{\vartheta}_0))\right] - q_\vartheta + O_p(n^{-1}) \\
&= 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + \ln p(\widehat{\boldsymbol{\theta}})\right] - 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\psi}}_0, \boldsymbol{\vartheta}_0) + \ln p(\widehat{\boldsymbol{\psi}}_0, \boldsymbol{\vartheta}_0)\right] - q_\vartheta + O_p(n^{-1}) \\
&= 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) + \ln p(\widehat{\boldsymbol{\theta}})\right] - 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0) + \ln p(\widehat{\boldsymbol{\theta}}_0)\right] - q_\vartheta + O_p(n^{-1}) \\
&= 2\left[\ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}) - \ln p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_0)\right] + 2\left[\ln p(\widehat{\boldsymbol{\theta}}) - \ln p(\widehat{\boldsymbol{\theta}}_0)\right] - q_\vartheta + O_p(n^{-1}) \\
&= \text{LR} + 2\left[\ln p(\widehat{\boldsymbol{\theta}}) - \ln p(\widehat{\boldsymbol{\theta}}_0)\right] - q_\vartheta + O_p(n^{-1}).
\end{aligned} \tag{6.27}
$$

Under H_0, based on the ML theory, we have $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_n^* + O_p(n^{-\frac{1}{2}})$ and $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_n^* + O_p(n^{-\frac{1}{2}})$ such that $\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_0 = O_p(n^{-\frac{1}{2}})$. Hence, using the Taylor expansion, we can also get that

$$
\begin{aligned}
\ln p(\widehat{\boldsymbol{\theta}}_0) &= \ln p(\widehat{\boldsymbol{\theta}}) + \frac{\partial \ln p(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_0 - \widehat{\boldsymbol{\theta}}) + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_0 - \widehat{\boldsymbol{\theta}})' \frac{\partial^2 \ln p(\widetilde{\boldsymbol{\theta}}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_0 - \widehat{\boldsymbol{\theta}}) \\
&= \ln p(\widehat{\boldsymbol{\theta}}) + O_p(1) O_p(n^{-\frac{1}{2}}) + O_p(n^{-\frac{1}{2}}) O_p(1) O_p(n^{-\frac{1}{2}}) = \ln p(\widehat{\boldsymbol{\theta}}) + O_p(n^{-\frac{1}{2}}),
\end{aligned} \tag{6.28}
$$

where $\widetilde{\boldsymbol{\theta}}_0$ is an intermediate value between $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_0$.

From Equations (6.27) and (6.28), under H_0, it is easy to show that

$$
\begin{aligned}
\text{PLR1} &= 2\left[D_c(\mathbf{y}) - D_c(\mathbf{y}, \boldsymbol{\vartheta}_0)\right] = \text{PLR2} - q_\vartheta + O_p(n^{-1}) \\
&= \text{LR} + 2\left[\ln p(\widehat{\boldsymbol{\theta}}) - \ln p(\widehat{\boldsymbol{\theta}}_0)\right] - q_\vartheta + O_p(n^{-1}) = \text{LR} - q_\vartheta + O_p(n^{-\frac{1}{2}}).
\end{aligned} \tag{6.29}
$$

Under the local alternative hypothesis of $\boldsymbol{\vartheta}_n^* = \boldsymbol{\vartheta}_0 + \frac{\boldsymbol{\delta}}{\sqrt{n}}$, based on the ML theory (see e.g. Proposition 4.2 in Lee, 2005), we have $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_n^* + O_p(n^{-\frac{1}{2}})$ and $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_n^* + O_p(n^{-\frac{1}{2}})$. Hence, $\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_0 = O_p(n^{-\frac{1}{2}})$. Naturally, Equation (6.29) holds.