Posterior-Based Wald-Type Statistics for Hypothesis Testing*

Xiaobin Liu
Zhejiang University

Yong Li
Renmin University

Jun Yu
Singapore Management University

Tao Zeng
Zhejiang University

November 21, 2020

Abstract

A new Wald-type statistic is proposed for hypothesis testing based on Bayesian posterior distributions under the correct model specification. The new statistic can be explained as a posterior version of the Wald statistic and has several nice properties. First, it is well-defined under improper prior distributions. Second, it avoids Jeffreys-Lindley-Bartlett's paradox. Third, under the null hypothesis and repeated sampling, it follows a χ^2 distribution asymptotically, offering an asymptotically pivotal test. Fourth, it only requires inverting the posterior covariance for parameters of interest. Fifth and perhaps most importantly, when a random sample from the posterior distribution (such as MCMC output) is available, the proposed statistic can be easily obtained as a by-product of posterior simulation. In addition, the numerical standard error of the estimated proposed statistic can be computed based on random samples. A robust version of the test statistic is developed under model misspecification and inherits many nice properties of the new posterior statistic. The finite sample performance of the statistics is examined in Monte Carlo studies. The method is applied to two latent variable models used in microeconometrics and financial econometrics.

JEL classification: C11, C12
Keywords: Decision theory; Hypothesis testing; Latent variable models; Posterior simulation; Wald test.

*We would like to thank Xiaohong Chen (the editor) and the two referees for their constructive comments. Liu gratefully acknowledges the financial support from the National Natural Science Foundation of China under No.72003171 and No.71973122. Li gratefully acknowledges the financial support of the Chinese Natural Science fund (No.71773130). Yu would like to acknowledge the financial support from Singapore Ministry of Education Academic Research Fund Tier 3 under the grant number MOE2013-T3-1-009. Zeng gratefully acknowledges the financial support from the Chinese Ministry of Education Project of Humanities and Social Sciences (No. 18YJC790005) and the National Natural Science Foundation of China (No. 72073121). All authors acknowledge Peter Phillips for helpful comments and financial support from the Lee Foundation via the Lee Kong Chian Professorship. Author for correspondence: Tao Zeng, School of Economics, Academy of Financial Research, and Institute for Fiscal Big-Data & Policy of Zhejiang University, Zhejiang, China 310058.

1 Introduction

This paper develops an approach to testing a point null hypothesis based on Bayesian posterior distributions under the correct model specification and also under model misspecification. The statistics can be understood as the posterior version of the well-known Wald statistic that has been used widely in practical applications. The Wald statistic is often based on the maximum likelihood estimator (MLE) or the classical extremum estimators (denoted by $\widehat{\boldsymbol{\theta}}$) of parameters of interest (denoted by $\boldsymbol{\theta}$). Typically some squared difference between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$ is shown to follow a χ^2 distribution asymptotically under the null hypothesis, producing an asymptotically pivotal test.

However, in many practical applications, the MLE and the classical extremum estimators may be too difficult to obtain computationally. For example, for the entire class of non-linear and non-Gaussian state-space models, the likelihood function is tough to calculate numerically, making the MLE nearly impossible to obtain. Not surprisingly, Bayesian MCMC methods have emerged as the leading estimation tool for non-linear and non-Gaussian state-space models. There are many other examples in economics and finance where the MLE and classical extremum estimators are too difficult to obtain; see, for example, Chernozhukov and Hong (2003), Imai et al. (2009) and Geweke et al. (2011). To circumvent the problems in frequentist methods, Chernozhukov and Hong (2003) introduce a class of quasi-Bayesian methods that allow users to employ MCMC to simulate a random sequence such that the marginal distribution of the sequence is the same as the quasi-posterior distribution.

The central question we ask in this paper is how to test a point null hypothesis when the posterior distribution but not classical extremum estimator is available. Testing a point null hypothesis is important for checking statistical evidence from data to support or to be against a particular theory because theory often can be reduced to a testable hypothesis. In many cases, the posterior distribution of parameters is available in the form of random samples (such as MCMC samples).

Broadly speaking, there are three posterior-based methods available in the literature for hypothesis testing. The first one is the Bayes factor (BF) that compares the marginal likelihoods of the two competing models corresponding to the null and alternative hypotheses (Kass and Raftery, 1995). Unfortunately, BFs are subject to a few theoretical and practical problems. First, BFs are not well-defined under improper priors. Second, BFs are subject to Jeffreys-Lindley-Bartlett's paradox. That is, they tend to choose the null hypothesis when a very vague prior is used for parameters in the null hypothesis; see Kass and Raftery (1995), Poirier (1995), Chapter 4 in Wakefield (2013). Third, in many cases, the evaluation of marginal likelihood is difficult. Several strategies have been proposed in the literature to address some of these difficulties. For example, to deal with the first two problems, when calculating BFs, one may use a prior that is data-dependent. To make the prior data-dependent, one may split the data into two parts, one as a training set, the other for statistical analysis. The training data can be used to update a prior (which can be improper) to generate a proper prior to analyze the remaining data. See the fractional BF of O'Hagan (1995) and the intrinsic BF of Berger (1985). To address the computational problem in BFs, one can use the methods of Friel and Pettitt (2008), Li et al. (2019), and Chib (1995).

The second posterior-based method uses credible intervals for point identified parameters and credible sets for partially identified parameters. This line of approaches has drawn a lot of attention among econometricians and statisticians in recent years; see Chernozhukov and Hong (2003), Moon and Schorfheide (2012), Kline and Tamer (2016), Liao and Simoni (2019), Chen et al. (2018). Most of these studies justify credible intervals and sets using a large sample theory under repeated sampling.

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

In this paper, following the third line of approach, we propose two Wald-type statistics for hypothesis testing based on posterior distributions, one for correctly specified models and the other for misspecified models. The new statistics are well-defined under improper prior distributions and avoid Jeffreys-Lindley-Bartlett's paradox. They are asymptotically equivalent to the Wald statistic under the null hypothesis, and hence, follow a χ^2 distribution asymptotically. They can be obtained as a by-product of posterior simulation under the alternative hypothesis, requiring almost no coding effort and incurring a low computational cost.

The paper is organized as follows. Section 2 reviews existing posterior-based statistics for hypothesis testing in the statistical decision framework. Section 3 develops the new statistic and establishes its large sample theory under the correct model specification. Section 4 develops a version of the test statistic that is robust under model misspecification and establishes its large sample theory. Section 5 investigates the finite-sample properties of the proposed statistic using simulated data. Section 6 gives two real data applications of the proposed method. Section 7 concludes the paper. The appendix collects proof of theoretical results.

Throughout the paper, let $\mathbf{I}_q$ denote the $q \times q$ identity matrix, $\mathbf{tr}(A)$ denote the trace of matrix A, $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$ denote the convergence in probability and convergence in distribution, $\otimes$ denote the Kronecker product, *vech* denote the column-wise vectorization of the lower triangular part of a symmetric matrix. The CPU time (in seconds) reported in the paper is from a laptop with an Intel i5 CPU and 8 GB memory.

2 Hypothesis Testing based on Statistical Decision Theory

It is assumed that a probability model $M \equiv \{p(\mathbf{y}|\boldsymbol{\vartheta})\}$ is used to fit data $\mathbf{y} := (y_1, \ldots y_n)'$ where $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \boldsymbol{\psi}')' \in \boldsymbol{\Theta}$. We are concerned with testing a point null hypothesis which may arise from the prediction of a particular theory. Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}_\theta$ denote a vector of q_θ-dimensional parameters of interest and $\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi$ a vector of q_ψ-dimensional nuisance parameters so that $\boldsymbol{\Theta} = \boldsymbol{\Theta}_\theta \times \boldsymbol{\Theta}_\psi$. The testing problem is given by

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

In the statistical decision framework, hypothesis testing may be understood as follows. There are two statistical decisions in the decision space, accepting H_0 (name it d_0) or rejecting H_0 (name it d_1). Let $\{\mathcal{L}(d_i, \boldsymbol{\theta}, \boldsymbol{\psi}), i = 0, 1\}$ be the loss function of the statistical decision associated with d_i. When the expected posterior loss of accepting H_0 is sufficiently larger than that of rejecting H_0, we reject H_0. That is, H_0 is rejected if

$$\begin{aligned} \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) &= \int_{\boldsymbol{\Theta}} \{\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) - \mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi})\} p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} \\ &= \int_{\boldsymbol{\Theta}} \Delta\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\boldsymbol{\psi} \\ &= E_{\boldsymbol{\vartheta}|\mathbf{y}} (\Delta\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})) > c \geq 0, \end{aligned}$$

where $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ is a posterior-based statistic, $p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})$ is the posterior distribution, c is a threshold value, $\Delta\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) - \mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi})$ is the net loss function.

BR show that if $p(\boldsymbol{\theta} = \boldsymbol{\theta}_0) = p(\boldsymbol{\theta} \neq \boldsymbol{\theta}_0) = \frac{1}{2}$, $c = 0$, and

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

then $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) > 0$ leads to the decision rule based on the BF. That is, reject H_0 if

$$\mathrm{BF}_{10} = \frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} = \frac{\int p(\mathbf{y}, \boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{\vartheta}}{\int p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}_0) \mathrm{d}\boldsymbol{\psi}} > 1.$$

While the BF serves as the gold standard for model comparison after posterior distributions are obtained for candidate models, it suffers from several theoretical and computational difficulties when it is used to test a point null hypothesis, as argued earlier.

In the statistical decision framework, several statistics have been proposed for testing a point null hypothesis. Poirier (1997) develops a loss function approach for hypothesis testing for models without latent variables. BR (2002) suggest choosing the loss function to be the KL divergence function. Given that the KL function is not analytically available for most latent variable models, Li and Yu (2012) suggest basing the loss function on the Q-function used in the EM algorithm. Li et al. (2014) suggest using the deviance function to be the loss function. While some of these statistics can avoid Jeffreys-Lindley-Bartlett's paradox in finite samples, unfortunately, no pivotal large sample theory is not available for any of them.

LLY (2015) propose the following quadratic net loss function

$$\Delta\mathcal{L}\left(H_0,\boldsymbol{\theta},\boldsymbol{\psi}\right)=\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)'C_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}_0\right)\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right),C\left(\boldsymbol{\vartheta}\right)=\left\{\frac{\partial\log p\left(\mathbf{y},\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\theta}}\right\}\left\{\frac{\partial\log p\left(\mathbf{y},\boldsymbol{\vartheta}\right)}{\partial\boldsymbol{\theta}}\right\}',$$

where $\bar{\boldsymbol{\vartheta}}=\left(\bar{\boldsymbol{\theta}}',\bar{\boldsymbol{\psi}}'\right)'$ and $\bar{\boldsymbol{\vartheta}}_0=\left(\boldsymbol{\theta}_0',\bar{\boldsymbol{\psi}}_0'\right)'$ are the posterior mean under H_1 and H_0, respectively, $C_{\theta\theta}$ is the submatrix of C corresponding to $\boldsymbol{\theta}$. The statistic corresponding to this net loss function is

$$\mathbf{T}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)=E_{\boldsymbol{\vartheta}|\mathbf{y}}\left(\Delta\mathcal{L}\left(H_0,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)=\int_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)'C_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}_0\right)\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)p(\boldsymbol{\vartheta}|\mathbf{y})\mathrm{d}\boldsymbol{\vartheta}. \quad (2)$$

Under repeated sampling, LLY (2015) show that $\mathbf{T}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ follows a χ^2 distribution asymptotically, providing an asymptotically pivotal quantity. This statistic is well-defined under improper priors and immune to Jeffreys-Lindley-Bartlett's paradox. Clearly, $\mathbf{T}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ requires evaluating the first-order derivative of the likelihood function. In many models, especially in latent variable models, this first-order derivative is not easy to evaluate since the observed-data likelihood function may not have an analytical expression.[1] Another feature of $\mathbf{T}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ is that it requires estimating both the null model and the alternative model although, under H_0, it is shown to be asymptotically equivalent to the Lagrange Multiplier (LM) test that requires estimating the null model only.

3 A Posterior Wald-type Statistic under the Correct Model Specification

Before proposing the new statistic, let us give a simple definition of the correct model specification. Assume that the data $\mathbf{y}=\left(y_1,\ldots,y_n\right)'$ come from a probability measure P_0

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

on the probability space (Ω, F, P_0). Let $P_{\boldsymbol{\vartheta}}$ be a collection of candidate models indexed by parameters $\boldsymbol{\vartheta}$. Following White (1987), if there exists $\boldsymbol{\vartheta}$ such that $P_0 \in P_{\boldsymbol{\vartheta}}$, we call that the model $P_{\boldsymbol{\vartheta}}$ is correctly specified. If for any $\boldsymbol{\vartheta}$, $P_0 \notin P_{\boldsymbol{\vartheta}}$, we say the model $P_{\boldsymbol{\vartheta}}$ is misspecified. By the information matrix equality (White, 1996), the assumption of the correct model specification implies Assumption 8 given later.

3.1 The statistic based on a quadratic loss function

For notational simplicity, let $\int$ denote $\int_{\boldsymbol{\Theta}}$ unless specified. For any $\tilde{\boldsymbol{\vartheta}} \in \boldsymbol{\Theta}$, denote

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

Let us introduce the following reasonable loss functions,

$$\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ \frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' \left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \tag{3}$$

$$\mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ -\frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' \left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}, \tag{4}$$

where $\bar{\boldsymbol{\vartheta}}$ is the posterior mean of $\boldsymbol{\vartheta}$ under H_1, $\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})$ the submatrix of $\mathbf{V}\left(\bar{\boldsymbol{\vartheta}}\right)$ corresponding to $\boldsymbol{\theta}$, $\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}$ the inverse of $\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})$.

Based on (3) and (4), we propose the following net loss function for hypothesis testing:

$$\Delta\mathcal{L}\left(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' \left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right).$$

Then, the new test statistic can be defined as:

$$\mathbf{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right) = \int \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' \left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta} = \mathbf{tr}\left[\left(\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right)^{-1} \mathbf{V}_{\theta}\left(\boldsymbol{\theta}_0\right)\right], \tag{5}$$

where $\mathbf{V}_{\theta}\left(\boldsymbol{\theta}_0\right) = \int \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' p(\boldsymbol{\vartheta}|\mathbf{y}) \mathrm{d}\boldsymbol{\vartheta}$. Clearly, $\mathbf{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ depends only on the alternative model.

Remark 3.1. *It is easy to show that $\mathbf{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ is well-defined under improper priors. An improper prior $p(\boldsymbol{\vartheta})$ satisfies that $p(\boldsymbol{\vartheta}) = af(\boldsymbol{\vartheta})$ where $f(\boldsymbol{\vartheta})$ is a non-integrable function and a is an arbitrary positive constant. Since the posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$ is independent of a, $\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})$, being the posterior covariance matrix of $\boldsymbol{\theta}$, is also independent of a. Hence, the proposed statistic does not depend on a.*

Remark 3.2. *To see how the new statistic can avoid Jeffreys-Lindley-Bartlett's paradox, consider the example in LLY (2015). Let $y_1, \ldots, y_n \sim N(\theta, \sigma^2)$ with a known σ^2, the null* *hypothesis be $H_0 : \theta = 0$, the prior distribution of θ be $N(0, \tau^2)$. Let $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. It is easy to show that the posterior distribution of θ is $N(\mu(\mathbf{y}), \omega^2)$ with*

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

and

$$2\log BF_{10} = \frac{n\tau^2}{n\tau^2 + \sigma^2}\frac{n\bar{y}^2}{\sigma^2} + \log\frac{\sigma^2}{n\tau^2 + \sigma^2},$$
$$\mathbf{T}(\mathbf{y}, \theta_0) = \frac{n\tau^2}{n\tau^2 + \sigma^2}\frac{n\bar{y}^2}{\sigma^2} + 1.$$

Thus, holding n constant, when $\tau^2 \to +\infty$ (the prior information becomes more and more uninformative), $\log BF_{10} \to -\infty$ which suggest that the BF supports H_0 regardless of $\bar{y}$. This is exactly what Jeffreys-Lindley-Bartlett's paradox predicts.[2] *On the other hand, $\mathbf{T}(\mathbf{y}, \theta_0) \to \frac{n\bar{y}^2}{\sigma^2} + 1$ as $\tau^2 \to +\infty$. Hence, $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ is distributed asymptotically as $\chi^2(1)$ when H_0 is true, suggesting that $\mathbf{T}(\mathbf{y}, \theta_0)$ is immune to Jeffreys-Lindley-Bartlett's paradox.*

3.2 Large sample theory for $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$

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

$$\mathbf{s}(\mathbf{y}^t, \boldsymbol{\vartheta}) := \frac{\partial \log p(\mathbf{y}^t|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \sum_{i=1}^{t} l_i^{(1)}(\boldsymbol{\vartheta}),\ \mathbf{h}(\mathbf{y}^t, \boldsymbol{\vartheta}) := \frac{\partial^2 \log p(\mathbf{y}^t|\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} = \sum_{i=1}^{t} l_i^{(2)}(\boldsymbol{\vartheta}),$$
$$\mathbf{s}_t(\boldsymbol{\vartheta}) := l_t^{(1)}(\boldsymbol{\vartheta}) = \mathbf{s}(\mathbf{y}^t, \boldsymbol{\vartheta}) - \mathbf{s}(\mathbf{y}^{t-1}, \boldsymbol{\vartheta}),\ \mathbf{h}_t(\boldsymbol{\vartheta}) := l_t^{(2)}(\boldsymbol{\vartheta}) = \mathbf{h}(\mathbf{y}^t, \boldsymbol{\vartheta}) - \mathbf{h}(\mathbf{y}^{t-1}, \boldsymbol{\vartheta}),$$
$$\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{h}_t(\boldsymbol{\vartheta}), \bar{\mathbf{J}}_n(\boldsymbol{\vartheta}) := \frac{1}{n}\sum_{t=1}^{n} [\mathbf{s}_t(\boldsymbol{\vartheta}) - \bar{\mathbf{s}}_t(\boldsymbol{\vartheta})][\mathbf{s}_t(\boldsymbol{\vartheta}) - \bar{\mathbf{s}}_t(\boldsymbol{\vartheta})]', \bar{\mathbf{s}}_t(\boldsymbol{\vartheta}) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{s}_t(\boldsymbol{\vartheta}),$$
$$\mathbf{B}_n(\boldsymbol{\vartheta}) := Var\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} l_t^{(1)}(\boldsymbol{\vartheta})\right], \mathcal{L}_n^{[j]}(\boldsymbol{\vartheta}) := \partial^j \log p(\boldsymbol{\vartheta}|\mathbf{y})/\partial \boldsymbol{\vartheta}^j,$$
$$\mathbf{H}_n(\boldsymbol{\vartheta}) := \int \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) g(\mathbf{y})\, d\mathbf{y},\ \mathbf{J}_n(\boldsymbol{\vartheta}) := \int \bar{\mathbf{J}}_n(\boldsymbol{\vartheta}) g(\mathbf{y})\, d\mathbf{y},$$

[2] Jeffreys-Lindley-Bartlett's paradox disappears if n is allowed to go to $+\infty$. The reason is as follows. No matter what τ is (as long as it is different from 0), when the true value of θ is 0, $\sqrt{n}\bar{y} \sim O_p(1)$. Hence, $\frac{n\tau^2}{n\tau^2+\sigma^2}\frac{n\bar{y}^2}{\sigma^2} \sim O_p(1)$. On the other hand, as $n \to +\infty$, $\log\frac{\sigma^2}{n\tau^2+\sigma^2} = \log\left(\frac{\sigma^2}{\tau^2+\sigma^2/n}\right) - \log(n) \to -\infty$. Hence, as $n \to +\infty$, $2\log BF_{10} \xrightarrow{p} -\infty$. When the true value of θ is different from 0, $n\bar{y}^2 \sim O_p(n)$. As a result, $\frac{n\tau^2}{n\tau^2+\sigma^2}\frac{n\bar{y}^2}{\sigma^2} \sim O_p(n)$, implying that $2\log BF_{10} = O_p(n) - \log(n) \xrightarrow{p} +\infty$.

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

In this paper, we first impose the following regularity conditions. Similar assumptions are used in Li et al. (2017, 2020). Li et al. (2020) discuss the importance of these assumptions.

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

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

Assumption 3: For all t, $l_t(\boldsymbol{\vartheta})$ is eight-times differentiable on $F_{-\infty}^t \times \boldsymbol{\Theta}$ where $F_{-\infty}^t = \sigma(y_t, y_{t-1}, \cdots)$.

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

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

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

Assumption 7: Let $\boldsymbol{\vartheta}_n^0$ be the minimizer of the KL divergence between the DGP $g(\mathbf{y})$ and the candidate model $p(\mathbf{y}|\boldsymbol{\vartheta})$, that is,

$$\boldsymbol{\vartheta}_n^0 = \arg\min_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} \frac{1}{n} \int \log \frac{g(\mathbf{y})}{p(\mathbf{y}|\boldsymbol{\vartheta})} g(\mathbf{y}) d\mathbf{y},$$

where $\left\{\boldsymbol{\vartheta}_n^0\right\}$ is the sequence of minimizers interior to $\boldsymbol{\Theta}$. For any $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup \sup_{\boldsymbol{\Theta}\setminus N\left(\boldsymbol{\vartheta}_{n,\varepsilon}^0\right)} \frac{1}{n} \sum_{t=1}^{n} \left\{ E\left[l_t(\boldsymbol{\vartheta})\right] - E\left[l_t\left(\boldsymbol{\vartheta}_n^0\right)\right] \right\} < 0, \tag{6}$$

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

Assumption 8: For all n, $\left\{\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right), \mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\}$ are negative definite and positive definite, respectively, and $\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right) + \mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right) = 0$.

Assumption 9: The prior density $p(\boldsymbol{\vartheta})$ is thrice continuously differentiable and $0 < p\left(\boldsymbol{\vartheta}_n^0\right) < \infty$. Moreover, there exists an n^* such that, for any $n > n^*$, the posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$ is proper and $\int \|\boldsymbol{\vartheta}\|^2 p(\boldsymbol{\vartheta}|\mathbf{y}) d\boldsymbol{\vartheta} < \infty$.

Let $\widehat{\boldsymbol{\vartheta}}$ be the MLE of $\boldsymbol{\vartheta}$, $\widehat{\boldsymbol{\theta}}$ be the subvector of $\widehat{\boldsymbol{\vartheta}}$ corresponding to $\boldsymbol{\theta}$. The frequentist Wald statistic is defined as

$$\mathbf{Wald} = n\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]^{-1} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right), \tag{7}$$

where $\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)$ is the submatrix of $\bar{\mathbf{H}}_{n}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)$ corresponding to $\boldsymbol{\theta}$ and $\bar{\mathbf{H}}_{n}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)$ is the inverse of $\bar{\mathbf{H}}_{n}\left(\widehat{\boldsymbol{\vartheta}}\right)$. Clearly, to compute **Wald**, one must invert $\bar{\mathbf{H}}_{n}\left(\widehat{\boldsymbol{\vartheta}}\right)$, a $(q_\theta+q_\psi)$-dimensional matrix.

By replacing $\widehat{\boldsymbol{\theta}}$ and $-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)$ with their posterior counterparts $\bar{\boldsymbol{\theta}}$ and $\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})$, a natural Bayesian version of **Wald** is given by,

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

It can be shown that

$$\begin{aligned}
\mathbf{T}\left(\mathbf{y},\boldsymbol{\theta}_0\right) &= \int\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)'\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)p(\boldsymbol{\vartheta}|\mathbf{y})d\boldsymbol{\theta}\\
&= \mathbf{tr}\left\{\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}E\left[\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)\left(\boldsymbol{\theta}-\boldsymbol{\theta}_0\right)'|\mathbf{y}\right]\right\}\\
&= \mathbf{tr}\left\{\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)+\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\right]\right\}\\
&= q_\theta+\mathbf{tr}\left\{\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\right\}\\
&= q_\theta+\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\\
&= q_\theta+\mathbf{W}.
\end{aligned}$$

Hence, $\mathbf{T}\left(\mathbf{y},\boldsymbol{\theta}_0\right)-q_\theta$ is equivalent to $\mathbf{W}$. The reason that $\mathbf{W}$ is not used as the test statistic in this paper is because it is difficult to specify a reasonable loss function for $\mathbf{W}$.

To show that $\mathbf{T}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ (and also $\mathbf{W}$) is a Bayesian version of **Wald**, we first need to show that $\bar{\boldsymbol{\vartheta}}$ and $\mathbf{V}\left(\widehat{\boldsymbol{\vartheta}}\right)$ are asymptotically equivalent to $\widehat{\boldsymbol{\vartheta}}$ and $-\frac{1}{n}\bar{\mathbf{H}}_{n}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)$. This is the implication of the Bernstein-von Mises (BvM) theorem, as reported in the next lemma.

Lemma 3.1. *Under Assumptions 1-9, it can be shown that*

$$\bar{\boldsymbol{\vartheta}} = E\left[\boldsymbol{\vartheta}|\mathbf{y},H_1\right]=\widehat{\boldsymbol{\vartheta}}+o_p(n^{-1/2}), \tag{8}$$

$$\mathbf{V}\left(\widehat{\boldsymbol{\vartheta}}\right) = E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)'|\mathbf{y},H_1\right]=-\frac{1}{n}\bar{\mathbf{H}}_{n}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)+o_p(n^{-1}). \tag{9}$$

Theorem 3.1. *If Assumptions 1-9 hold, and under H_0,*

$$\mathbf{T}\left(\mathbf{y},\boldsymbol{\theta}_0\right)-q_\theta=\mathbf{W}=\mathbf{Wald}+o_p(1)\stackrel{d}{\to}\chi^2(q_\theta).$$

Corollary 3.2. *Under the same conditions as in Theorem 3.1 we have*

$$\mathbf{T}\left(\mathbf{y},\boldsymbol{\theta}_0\right)-q_\theta=\mathbf{T}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)+o_p(1)\stackrel{d}{\to}\chi^2(q_\theta). \tag{10}$$

Remark 3.3. *LLY (2015) establish the relationship between $\mathbf{T}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ and* **LM** *under H_0, that is, $\mathbf{T}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)=$* **LM** *$+o_p(1)$. It is noted in Engle (1984) that under H_0,* **LM** $=$ **Wald**$+o_p(1)$. *Equation (10) is the posterior version of this asymptotic equivalence between*

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

Remark 3.4. *Theorem 3.1 suggests that the asymptotic distribution of* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ *is pivotal. To implement the proposed test, we can choose the threshold value, c, to be the critical value of* $\chi^2(q_\theta)$ *distribution such that we*

$$\text{accept } H_0 \text{ if } \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta \leq c; \text{ reject } H_0 \text{ if } \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta > c.$$

Remark 3.5. *It is obvious that* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ *only requires evaluating the inverse of the submatrix of the covariance matrix corresponding to* $\boldsymbol{\theta}$*. In contrast, the Wald statistic in (7) requires evaluating the inverse of the entire empirical Hessian matrix and then use the submatrix corresponding to* $\boldsymbol{\theta}$*. When* $\boldsymbol{\vartheta}$ *is high-dimensional but* $\boldsymbol{\theta}$ *is low-dimensional, the inversion in Wald is numerically more involved than that in* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$*. For example, consider the case where the dimension of* $\boldsymbol{\vartheta}$ *is* 100*, but the null hypothesis involves only one of the parameters. To use the Wald statistic, one has to evaluate the inverse of a* 100×100 *dimensional Hessian matrix. Whereas, to use* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$*, one only needs to evaluate the inverse of a scalar.*

Remark 3.6. *Unlike the Wald statistic,* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ *depends on the prior distribution and the prior information may help improve the performance of* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$*. To illustrate this advantage and also the advantage of the proposed test over the BF, we consider the following experiment. Let* $y_1, ..., y_n \sim N(\theta, 1)$ *and we test* $H_0 : \theta = 0$*. Consider two cases where the true value of* θ *is set at* $\theta_n^0 = 0.1$ *and* 0*, respectively. The prior distribution of* θ *is* $N(\mu_0, \tau^2)$*. It can be shown that*

$$\begin{aligned}
2\log BF_{10} &= \frac{n\tau^2}{1+n\tau^2}\left[\left(\sqrt{n}\bar{y}\right)^2 + 2\frac{\mu_0\bar{y}}{\tau^2} - \frac{\mu_0^2}{\tau^2}\right] + \log\frac{1}{1+n\tau^2}, \\
\mathbf{T}(\mathbf{y}, \theta_0) - 1 &= \frac{1}{\frac{1}{\tau^2}+n}\left(n\bar{y} + \frac{\mu_0}{\tau^2}\right)^2 = \frac{1}{\frac{1}{n\tau^2}+1}n\bar{y}^2 + \frac{1}{\frac{1}{\tau^2}+n}\left(2n\bar{y}\frac{\mu_0}{\tau^2} + \frac{\mu_0^2}{\tau^4}\right), \\
\mathbf{Wald} &= n\bar{y}^2.
\end{aligned}$$

Under H_0*, as* $n \to +\infty$*,*

$$\mathbf{T}(\mathbf{y}, \theta_0) - 1 - \mathbf{Wald} \xrightarrow{p} 0,$$

and the asymptotic distribution for both $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *and* **Wald** *is* $\chi^2(1)$*. For each case of* θ_n^0*, suppose two prior distributions are used, a highly informative prior* $N(\theta_n^0, 10^{-3})$ *and a very vague prior* $N(0, 10^{50})$*. We simulate n observations from the model, obtaining* $\bar{y}$ *and calculating* $2\log BF_{10}$*,* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$*, and* **Wald***. We repeat the experiment for*

5,000 times. Table 1 reports rejection rates of $2 \log BF_{10}$*,* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$*, and* **Wald** *when* $n = 10, 100, 1000, 10000$ *out of 5,000 replications. The top panel corresponds to the case of* $\theta_n^0 = 0.1$*. The bottom panel corresponds to the case of* $\theta_n^0 = 0$*. The left panel corresponds to the case of the informative prior* $N(\theta_n^0, 10^{-3})$*. The right panel corresponds to the case of the vague prior* $N(0, 10^{50})$*.*

When $\theta_n^0 = 0.1$*, a good test is expected to reject* H_0*. What we find from the top panel of Table 1 is that* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *rejects* H_0 *successfully regardless of prior as long as* n *is large enough and also that* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *rejects* H_0 *successfully regardless of* n *as long as the prior is highly informative.* **Wald** *performs much worse than* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *when the prior is highly information and* n *is very small. Moreover, under the vague prior,* $2 \log BF_{10}$ *tends to choose the wrong model even when the sample size is very large (but finite). This is the same as Jeffreys-Lindley-Bartlett's paradox.*

When $\theta_n^0 = 0$*, a good test is expected not to reject* H_0*. What we find from the bottom panel of Table 1 is the rejection rates for* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *and* **Wald** *are the same which are always very close to 5% under the vague prior. Under the informative prior, the rejection rates for* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *are much lower than 5% for small values of* n *and tend to 5% as* n *increases. While* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *is conservative in this case,* $2 \log BF_{10}$ *tends to choose the wrong model too often under the informative prior. For example, when* $n = 1,000$*, it chooses the wrong model 25% times.*

Table 1: The rejection rates of alternative test statistics for $H_0 : \theta = 0$ in 5,000 replications

	$\theta_n^0 = 0.1$							
Prior	$N\left(0.1, 10^{-3}\right)$				$N\left(0, 10^{50}\right)$			
n	10	100	$1,000$	$10,000$	10	100	$1,000$	$10,000$
$2 \log BF_{10}$	100	100	100	100	0	0	0	12.82
$\mathbf{T}(\mathbf{y}, \theta_0) - 1$	100	100	99.98	100	5.74	16.96	88.50	100
Wald	5.74	16.96	88.50	100	5.74	16.96	88.50	100
	$\theta_n^0 = 0$							
Prior	$N\left(0, 10^{-3}\right)$				$N\left(0, 10^{50}\right)$			
n	10	100	$1,000$	$10,000$	10	100	$1,000$	$10,000$
$2 \log BF_{10}$	31.86	30.56	24.94	10.12	0	0	0	0
$\mathbf{T}(\mathbf{y}, \theta_0) - 1$	0	0	0.74	4.24	4.70	4.76	5.02	5.20
Wald	4.70	4.76	5.02	5.20	4.70	4.76	5.02	5.20

3.3 Extension to hypotheses in a general form

In this subsection, we extend the point null hypothesis to the following non-linear form,

$$\begin{cases} H_0: & R\left(\boldsymbol{\theta}_n^0\right) = \boldsymbol{r} \\ H_1: & R\left(\boldsymbol{\theta}_n^0\right) \neq \boldsymbol{r} \end{cases}, \tag{11}$$

where $R(\cdot) : \boldsymbol{\Theta}_\theta \to \mathbb{R}^m$, $m \leq q_\theta$, and $\boldsymbol{r} \in \mathbb{R}^m$. Here R is a set of m non-linear functions/restrictions. We can test for a single hypothesis on multiple parameters, as well as a jointly multiple hypotheses on single/multiple parameters. While this hypothesis is in the standard form for the Wald test, as pointed out by McCulloch and Rossi (1992), the non-linear restrictions make BFs difficult to implement. To develop large sample properties of the proposed test, we need to impose the following assumption on $R(\boldsymbol{\theta})$.

Assumption 10: $R(\boldsymbol{\theta})$ is second-order continuously differentiable with respect to $\boldsymbol{\theta}$ on $\boldsymbol{\Theta}_\theta$ and full rank at $\boldsymbol{\theta}_n^0$.

For the hypothesis defined in (11), the classical Wald statistic and its asymptotic theory are

$$\mathbf{Wald} = \left[R\left(\widehat{\boldsymbol{\theta}}\right) - \mathbf{r}\right]' n \left\{ \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} \left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\right] \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}} \right\}^{-1} \left[R\left(\widehat{\boldsymbol{\theta}}\right) - \boldsymbol{r}\right] \xrightarrow{d} \chi^2(m).$$

Based on the statistical decision theory, we can define the following net loss function

$$\Delta\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = (R(\boldsymbol{\theta}) - \mathbf{r})' \left[\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} \mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right) \frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right]^{-1} (R(\boldsymbol{\theta}) - \mathbf{r}),$$

and introduce our test statistic as:

$$\begin{aligned}
\mathbf{T}(\mathbf{y}, \mathbf{r}) &= \int \Delta\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})\, p(\boldsymbol{\vartheta}|\mathbf{y})\, d\boldsymbol{\vartheta} \\
&= \int (R(\boldsymbol{\theta}) - \boldsymbol{r})' \left[\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} V_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right) \frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right]^{-1} (R(\boldsymbol{\theta}) - \mathbf{r})\, p(\boldsymbol{\vartheta}|\mathbf{y})\, d\boldsymbol{\vartheta} \\
&= \mathbf{tr}\left[\left(\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} V_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right) \frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right)^{-1} \mathbf{V}_\theta(\mathbf{r})\right],
\end{aligned} \tag{12}$$

where $\mathbf{V}_\theta(\mathbf{r}) = \int (R(\boldsymbol{\theta}) - \mathbf{r})(R(\boldsymbol{\theta}) - \mathbf{r})' p(\boldsymbol{\vartheta}|\mathbf{y})\, d\boldsymbol{\vartheta}$.

Theorem 3.3. *If Assumptions 1-10 hold, and under H_0,*

$$\mathbf{T}(\mathbf{y}, \mathbf{r}) - m = \mathbf{Wald} + o_p(1) \xrightarrow{d} \chi^2(m).$$

3.4 Calculating the proposed statistic

As noted in Sections 3.2 and 3.3, the proposed statistics are only dependent on the posterior mean and the posterior variance of $\boldsymbol{\vartheta}$, $\overline{\boldsymbol{\vartheta}}$ and $\mathbf{V}\left(\overline{\boldsymbol{\vartheta}}\right)$. In practice, $\overline{\boldsymbol{\vartheta}}$ and $\mathbf{V}\left(\overline{\boldsymbol{\vartheta}}\right)$ are often unknown analytically. Fortunately, when random samples from the posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$ are obtained via posterior simulation (such as MCMC or importance sampling), we can consistently estimate $\overline{\boldsymbol{\vartheta}}$ and $\mathbf{V}\left(\overline{\boldsymbol{\vartheta}}\right)$ arbitrarily well. Specifically, let $\left\{\boldsymbol{\vartheta}^{[j]}, j=1,\ldots,J\right\}$ be effective samples generated from $p(\boldsymbol{\vartheta}|\mathbf{y})$, and consistent estimates of $\overline{\boldsymbol{\vartheta}}$, $\overline{\boldsymbol{\theta}}$ and $\mathbf{V}\left(\overline{\boldsymbol{\vartheta}}\right)$ be given by

$$\bar{\bar{\boldsymbol{\vartheta}}} = \frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\vartheta}^{[j]}, \bar{\mathbf{v}}_1 = \frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\theta}^{[j]}, \bar{\mathbf{V}}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right) = \frac{1}{J}\sum_{j=1}^{J}\left(\boldsymbol{\vartheta}^{[j]} - \bar{\bar{\boldsymbol{\vartheta}}}\right)\left(\boldsymbol{\vartheta}^{[j]} - \bar{\bar{\boldsymbol{\vartheta}}}\right)'.$$

By plugging $\bar{\bar{\boldsymbol{\vartheta}}}$, $\bar{\mathbf{v}}_1$, and $\bar{\mathbf{V}}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)$ into the proposed statistics, we obtain consistent estimates of $\mathbf{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ and $\mathbf{T}\left(\mathbf{y}, \mathbf{r}\right)$ as

$$\begin{aligned}
\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{\theta}_0\right) &:= \mathbf{tr}\left[\left(\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\right)^{-1}\bar{\mathbf{V}}_{\theta}\left(\boldsymbol{\theta}_0\right)\right], \\
\widehat{\mathbf{T}}\left(\mathbf{y}, \mathbf{r}\right) &:= \mathbf{tr}\left[\left(\frac{\partial R\left(\bar{\mathbf{v}}_1\right)}{\partial \boldsymbol{\theta}'}\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\frac{\partial R\left(\bar{\mathbf{v}}_1\right)}{\partial \boldsymbol{\theta}}\right)^{-1}\bar{\mathbf{V}}_{\theta}(\mathbf{r})\right],
\end{aligned} \tag{13}$$

where

$$\begin{aligned}
\bar{\mathbf{V}}_{\theta}\left(\boldsymbol{\theta}_0\right) &= \frac{1}{J}\sum_{j=1}^{J}\left(\boldsymbol{\theta}^{[j]} - \boldsymbol{\theta}_0\right)\left(\boldsymbol{\theta}^{[j]} - \boldsymbol{\theta}_0\right)', \\
\bar{\mathbf{V}}_{\theta}(\mathbf{r}) &= \frac{1}{J}\sum_{j=1}^{J}\left(R\left(\boldsymbol{\theta}^{[j]}\right) - \boldsymbol{r}\right)\left(R\left(\boldsymbol{\theta}^{[j]}\right) - \mathbf{r}\right)'.
\end{aligned}$$

Remark 3.7. *Various approaches have been developed for posterior simulation. Examples include Monte Carlo (MC) integration, importance sampling, MCMC techniques such as the Gibbs sampler and the Metropolis-Hastings algorithm. For more details about posterior simulation, one can refer to Geweke (2005). All these approaches can be used to generate random observations from $p(\boldsymbol{\vartheta}|\mathbf{y})$. From (13), the proposed statistics are by-products of posterior simulation. Furthermore, the test statistics can be applied in a variety of models.*

When $\widehat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0)$ and $\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{r}\right)$ are calculated from posterior simulation, it is important to obtain their numerical standard error (NSE) that measures the magnitude of simulation errors. The following theorem shows how to calculate the NSE of $\widehat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0)$ and $\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{r}\right)$.

Theorem 3.4. *Let $\bar{\mathbf{V}}_2 = \frac{1}{J}\sum_{j=1}^{J}\left(\boldsymbol{\theta}^{[j]} - \bar{\mathbf{v}}_1\right)\left(\boldsymbol{\theta}^{[j]} - \bar{\mathbf{v}}_1\right)'$, $\bar{\mathbf{v}}_2 = vech\left(\bar{\mathbf{V}}_2\right)$, $\bar{\mathbf{v}} = \left(\bar{\mathbf{v}}_1', \bar{\mathbf{v}}_2'\right)'$, $Var\left(\bar{\mathbf{v}}\right)$ be the NSE of $\bar{\mathbf{v}}$. The NSE of $\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ is given by*

$$NSE\left(\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)\right) = \sqrt{\frac{\partial \widehat{\mathbf{T}}\left(\mathbf{y},\ \boldsymbol{\theta}_0\right)}{\partial \bar{\mathbf{v}}'}Var\left(\bar{\mathbf{v}}\right)\frac{\partial \widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)}{\partial \bar{\mathbf{v}}}},$$

where

$$\frac{\partial \widehat{\mathbf{T}}(\mathbf{y},\boldsymbol{\theta}_0)}{\partial \bar{\mathbf{v}}} = vech\left(\mathbf{I}_{q_\theta}\right)'\left[\left(\left((\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0)'\,\bar{\mathbf{V}}_2^{-1}\right)'\otimes \mathbf{I}_{q_\theta}+\bar{\mathbf{V}}_2^{-1}\otimes(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0)\right)\frac{\partial \bar{\mathbf{v}}_1}{\partial \bar{\mathbf{v}}}\right.$$
$$\left. - \left[\mathbf{I}_{q_\theta}\otimes(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0)(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0)'\right]\left(\bar{\mathbf{V}}_2^{-1}\otimes\bar{\mathbf{V}}_2^{-1}\right)\frac{\partial \bar{\mathbf{V}}_2}{\partial \bar{\mathbf{v}}}\right],$$

and

$$\frac{\partial \bar{\mathbf{v}}_1}{\partial \bar{\mathbf{v}}}=\frac{\partial \bar{\mathbf{v}}_1'}{\partial \bar{\mathbf{v}}}=\left[\mathbf{I}_{q_\theta},0_{q_\theta\times q^*}\right],\ \frac{\partial \bar{\mathbf{V}}_2}{\partial \bar{\mathbf{v}}}=\left[0_{q_\theta^2\times q_\theta},\left(\frac{\partial vech\left(\bar{\mathbf{V}}_2\right)}{\partial \bar{\mathbf{v}}_2}\right)_{q_\theta^2\times q^*}\right].$$

Furthermore, the NSE of $\widehat{\mathbf{T}}(\mathbf{y},\mathbf{r})$ *is given by*

$$NSE\left(\widehat{\mathbf{T}}(\mathbf{y},\mathbf{r})\right)=\sqrt{\frac{\partial \widehat{\mathbf{T}}(\mathbf{y},\mathbf{r})}{\partial \bar{\mathbf{v}}'}Var(\bar{\mathbf{v}})\frac{\partial \widehat{\mathbf{T}}(\mathbf{y},\mathbf{r})}{\partial \bar{\mathbf{v}}}},$$

where

$$\frac{\partial \widehat{\mathbf{T}}(\mathbf{y},\mathbf{r})}{\partial \bar{\mathbf{v}}} = vech\left(\mathbf{I}_m\right)'\left\{\left[\left((\bar{\mathbf{v}}_3-\mathbf{r})'\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\right)'\otimes\mathbf{I}_m\right]\frac{\partial \bar{\mathbf{v}}_3}{\partial \bar{\mathbf{v}}_1}\frac{\partial \bar{\mathbf{v}}_1}{\partial \bar{\mathbf{v}}}\right.$$
$$+\left[\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\otimes(\bar{\mathbf{v}}_3-\mathbf{r})\right]\frac{\partial \bar{\mathbf{v}}_3'}{\partial \bar{\mathbf{v}}_1}\frac{\partial \bar{\mathbf{v}}_1}{\partial \bar{\mathbf{v}}}$$
$$+\left[\mathbf{I}_m\otimes(\bar{\mathbf{v}}_3-\mathbf{r})(\bar{\mathbf{v}}_3-\mathbf{r})'\right]\left[\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\otimes\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\right]$$
$$\left.\times\frac{\partial vech\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)}{\partial \bar{\mathbf{v}}}\right\},$$

$$\bar{\mathbf{v}}_3=R(\bar{\mathbf{v}}_1),\ \bar{\mathbf{V}}_4=\frac{\partial R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\bar{\mathbf{v}}_1},$$

$$\frac{\partial vech\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)}{\partial \bar{\mathbf{v}}}=\left(\left(\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)'\otimes\mathbf{I}_m\right)\frac{\partial \bar{\mathbf{V}}_4'}{\partial \bar{\mathbf{v}}_1}\frac{\partial \bar{\mathbf{v}}_1}{\partial \bar{\mathbf{v}}}+\left(\bar{\mathbf{V}}_4\otimes\bar{\mathbf{V}}_4'\right)\frac{\partial \bar{\mathbf{V}}_2}{\partial \bar{\mathbf{v}}}$$
$$+\left(\mathbf{I}_m\otimes\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\right)\frac{\partial \bar{\mathbf{V}}_4}{\partial \bar{\mathbf{v}}_1}\frac{\partial \bar{\mathbf{v}}_1}{\partial \bar{\mathbf{v}}},$$

and the derivatives of $\bar{\mathbf{V}}_4$ *and* $\bar{\mathbf{v}}_3$ *depend on the form of* $R(\boldsymbol{\theta})$.

Remark 3.8. *Following Newey and West (1987), a consistent estimator of the NSE of* $\bar{\mathbf{v}}$ *is given by*

$$Var(\bar{\mathbf{v}})=\frac{1}{J}\left[\Omega_0+\sum_{k=1}^{K}\left(1-\frac{k}{K+1}\right)\left(\Omega_k+\Omega_k'\right)\right],$$

where

$$\Omega_k=J^{-1}\sum_{j=k+1}^{J}\left(\mathbf{v}^{[j]}-\bar{\mathbf{v}}\right)\left(\mathbf{v}^{[j]}-\widehat{\mathbf{v}}\right)'.$$

4 Posterior Statistic under Model Misspecification

In many applications, the assumption of the correct model specification is too strong. We now develop a version of posterior statistic that is robust under model misspecification. According to the information matrix equality, the correct model specification implies $\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)+\mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right)=0$. To allow for model misspecification, we replace Assumption 8 with the following assumption without imposing $\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)+\mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right)=0$.

Assumption 8B: For all n, $\left\{\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right),\mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\}$ are negative definite and positive definite, respectively.

Under Assumptions 1-7 and 8B,

$$\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)-\mathbf{H}_n^{-1}\left(\boldsymbol{\vartheta}_n^0\right)\xrightarrow{p}0. \tag{14}$$

According to Li et al. (2020), under Assumptions 1-7, 8B, 9,

$$\bar{\mathbf{H}}_n^{-1}\left(\bar{\boldsymbol{\vartheta}}\right)-\mathbf{H}_n^{-1}\left(\boldsymbol{\vartheta}_n^0\right)\xrightarrow{p}0. \tag{15}$$

Moreover, according to Newey and West (1987), a heteroskedasticity and autocorrelation consistent (HAC) estimator of $\mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right)$ can be constructed

$$\bar{\boldsymbol{\Omega}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)=\frac{1}{n}\sum_{t=1}^{n}\sum_{\tau=1}^{n}\mathbf{s}_t\left(\widehat{\boldsymbol{\vartheta}}\right)\mathbf{s}_\tau\left(\widehat{\boldsymbol{\vartheta}}\right)'k\left(\frac{t-\tau}{\gamma_n}\right),$$

where $k(\cdot)$ is a kernel function and γ_n is the bandwidth. To ensure consistency and positive semidefiniteness of $\bar{\boldsymbol{\Omega}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)$, following de Jong and Davidson (2000), we add three assumptions, the first two of which are about the kernel function and bandwidth, while the last of which is about the score function $\mathbf{s}_t\left(\boldsymbol{\vartheta}_n^0\right)$.

Assumption 11: Assume the kernel function $k(\cdot)\in\mathcal{H}$, where

$$\mathcal{H}=\left\{\begin{array}{c} k(\cdot):R\rightarrow[-1,1],k(x)=k(-x),\text{for any }x\in R,\\ \int_{-\infty}^{+\infty}|k(x)|\,dx<\infty,\ \int_{-\infty}^{+\infty}\psi(\xi)\,d\xi<\infty,\\ k(\cdot)\ \text{ is continuous at 0 and at all but a finite number of points in }R\end{array}\right\},$$

where

$$\psi(\xi)=(2\pi)^{-1}\int_{-\infty}^{+\infty}k(x)\,e^{i\xi x}dx.$$

Assumption 12: The bandwidth parameter γ_n is an increasing function of sample size n and $\gamma_n=o\left(n^{1/2}\right)$.

Assumption 13: The expectation of the score function $E\left(\mathbf{s}_t\left(\boldsymbol{\vartheta}_n^0\right)\right)=0$ for any t.

Remark 4.1. *In Assumption 11, the function class $\mathcal{H}$ includes many well-known kernel functions, such as Bartlett, Parzen, Quadratic Spectral, and Tukey-Hanning kernels. It ensures that $\bar{\boldsymbol{\Omega}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)$ is positive semidefinite with probability approaching one; see Andrews (1991). Note that $\mathcal{H}$ does not include truncated kernels.*

Remark 4.2. *From Assumption 1-7 and 8B, according to Gallant and White (1988),*

$$\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_n^0\right)\xrightarrow{d}N\left(0,\boldsymbol{\Sigma}_n^0\right),$$

where $\boldsymbol{\Sigma}_n^0$ *is the so-called "sandwich" covariance matrix given by*

$$\boldsymbol{\Sigma}_n^0=\mathbf{H}_n^{-1}\left(\boldsymbol{\vartheta}_n^0\right)\mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right)\mathbf{H}_n^{-1}\left(\boldsymbol{\vartheta}_n^0\right).$$

According to Li et al. (2020), under Assumptions 1-7, 8B and 11-13,

$$\bar{\boldsymbol{\Omega}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)-\mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right)\xrightarrow{p}0. \tag{16}$$

Then a consistent estimator of $\boldsymbol{\Sigma}_n^0$ *is*

$$\hat{\boldsymbol{\Sigma}}_S\left(\widehat{\boldsymbol{\vartheta}}\right)=\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\bar{\boldsymbol{\Omega}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right).$$

Remark 4.3. *According to Li et al. (2020), under Assumptions 1-7, 8B, 9, 11-12,*

$$\bar{\boldsymbol{\Omega}}_n\left(\bar{\boldsymbol{\vartheta}}\right)-\bar{\boldsymbol{\Omega}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\xrightarrow{p}0. \tag{17}$$

Hence, $\bar{\boldsymbol{\Omega}}_n\left(\bar{\boldsymbol{\vartheta}}\right)$ *is also a consistent estimator of* $\mathbf{B}_n\left(\boldsymbol{\vartheta}_n^0\right)$. *From (15) and (17), we have*

$$\hat{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)-\hat{\boldsymbol{\Sigma}}_S\left(\widehat{\boldsymbol{\vartheta}}\right)\xrightarrow{p}0.$$

That is, $\hat{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)$ *is also a consistent estimator of* $\boldsymbol{\Sigma}_n^0$. *If we further define*

$$\bar{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)=n\mathbf{V}\left(\bar{\boldsymbol{\vartheta}}\right)\bar{\boldsymbol{\Omega}}_n\left(\bar{\boldsymbol{\vartheta}}\right)n\mathbf{V}\left(\bar{\boldsymbol{\vartheta}}\right),$$

it can be shown that

$$\bar{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)-\hat{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)\xrightarrow{p}0,$$

which means that

$$\bar{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)-\boldsymbol{\Sigma}_n^0\xrightarrow{p}0. \tag{18}$$

It is well known that, whether the model is misspecified or not, the BvM theorem suggests that

$$\sqrt{n}\left(\boldsymbol{\vartheta}-\bar{\boldsymbol{\vartheta}}\right)|\mathbf{y}\overset{a}{\sim}N\left(\mathbf{0},-\mathbf{H}_n^{-1}\left(\boldsymbol{\vartheta}_n^0\right)\right).$$

Due to the discrepancy between $-\mathbf{H}_n^{-1}\left(\boldsymbol{\vartheta}_n^0\right)$ and $\boldsymbol{\Sigma}_n^0$ in misspecified models, Müller (2013) uses an artificial posterior distribution based on the "sandwich" covariance matrix to improve the Bayesian statistical inference. In particular, the posterior distribution of $\boldsymbol{\vartheta}$ is artificially constructed as $N\left(\bar{\boldsymbol{\vartheta}},\frac{1}{n}\boldsymbol{\Sigma}_n^0\right)$. Since $\boldsymbol{\Sigma}_n^0$ is unknown, it is substituted by the consistent estimator $\bar{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)$. Müller (2013) shows that for misspecified models, this new posterior distribution can yield a lower risk for parameter estimation compared with the original posterior distribution. We now show how to use $N\left(\bar{\boldsymbol{\vartheta}}, \frac{1}{n}\bar{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)\right)$ to construct our test statistic that is robust under model misspecification.

Let $p_S\left(\boldsymbol{\vartheta}|\mathbf{y}\right)$ denote the pdf of $N\left(\bar{\boldsymbol{\vartheta}}, \frac{1}{n}\bar{\boldsymbol{\Sigma}}_S\left(\bar{\boldsymbol{\vartheta}}\right)\right)$. Define the following loss function:

$$\mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ \frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases} \tag{19}$$

and

$$\mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ -\frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) & \text{if } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases}. \tag{20}$$

Hence, the net loss function is

$$\Delta\mathcal{L}\left(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right).$$

Based on this net loss function, we propose the following test statistic

$$\begin{aligned} \mathbf{T}_S\left(\mathbf{y}, \boldsymbol{\theta}_0\right) &= \int \Delta\mathcal{L}\left(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) p_S\left(\boldsymbol{\vartheta}|\mathbf{y}\right) \mathrm{d}\boldsymbol{\vartheta} \\ &= \int \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) p_S\left(\boldsymbol{\vartheta}|\mathbf{y}\right) \mathrm{d}\boldsymbol{\vartheta}. \end{aligned}$$

Let $\mathbf{W}_S := \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$. We have

$$\begin{aligned} \mathbf{T}_S\left(\mathbf{y}, \boldsymbol{\theta}_0\right) &= \mathbf{tr}\left\{n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1} \int \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' p_S\left(\boldsymbol{\theta}|\mathbf{y}\right) d\boldsymbol{\theta}\right\} \\ &= \mathbf{tr}\left\{n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1} \left[\frac{1}{n}\bar{\boldsymbol{\Sigma}}_{S,\,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right) + \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)'\right]\right\} \\ &= q_\theta + \mathbf{tr}\left\{n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)'\right\} \\ &= q_\theta + \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' n \left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \\ &= q_\theta + \mathbf{W}_S. \end{aligned}$$

Define the Wald statistic that is robust under model specification as

$$\mathbf{Wald}_S = \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' n \left[\hat{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right).$$

The following theorem establishes the asymptotic equivalence between $\mathbf{T}_S\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ and $\mathbf{Wald}_S$ under H_0.

Theorem 4.1. *If Assumptions 1-7, 8B, 9, 11-13 hold, and under* H_0*,*

$$\mathbf{T}_S\left(\mathbf{y}, \boldsymbol{\theta}_0\right) - q_\theta = \mathbf{Wald}_S + o_p(1).$$

We can extend the results to the hypothesis defined in (11) under model misspecification. Note that the classical Wald statistic under misspecification and its null asymptotic theory are

$$\mathbf{Wald}_S^R = \left[R\left(\widehat{\boldsymbol{\theta}}\right) - \mathbf{r}\right]' n \left\{\frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} \hat{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\widehat{\boldsymbol{\vartheta}}\right) \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right\}^{-1} \left[R\left(\widehat{\boldsymbol{\theta}}\right) - \boldsymbol{r}\right] \xrightarrow{d} \chi^2(m).$$

Define the net loss function

$$\Delta\mathcal{L}\left(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) = \left(R\left(\boldsymbol{\theta}\right) - \mathbf{r}\right)' n \left[\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} \bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right) \frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right]^{-1} \left(R\left(\boldsymbol{\theta}\right) - \mathbf{r}\right),$$

and the test statistic

$$\begin{aligned}
\mathbf{T}_S\left(\mathbf{y}, \mathbf{r}\right) &= \int \Delta\mathcal{L}\left(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}\right) p_S\left(\boldsymbol{\vartheta}|\mathbf{y}\right) d\boldsymbol{\vartheta} \\
&= \int \left(R\left(\boldsymbol{\theta}\right) - \boldsymbol{r}\right)' n \left[\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} \bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right) \frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right]^{-1} \left(R\left(\boldsymbol{\theta}\right) - \mathbf{r}\right) p_S\left(\boldsymbol{\vartheta}|\mathbf{y}\right) d\boldsymbol{\vartheta}.
\end{aligned}$$

Theorem 4.2. *If Assumptions 1-7, 8B, 9-13 hold, under H_0,*

$$\mathbf{T}_S\left(\mathbf{y}, \mathbf{r}\right) - m = \mathbf{Wald}_S^R + o_p(1) \xrightarrow{d} \chi^2(m).$$

5 Simulation Studies

We design three experiments to examine the finite-sample performance of the proposed test with simulated data. In the first experiment, we test different null hypotheses in a linear regression model. This study aims to compare the finite-sample behavior of $\mathbf{T}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ and that of Wald in terms of size and power. In the second experiment, we study the finite-sample properties of $\mathbf{T}_S\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ in a misspecified linear regression model. In the third experiment, we test the point null hypothesis in a discrete choice model. It is a simultaneous equation model with ordered probit and two-limit censored regression. Li (2006) applied this microeconometric model to study the relationship between high school completion and future youth unemployment.

5.1 Hypothesis testing in a linear regression model

The linear regression model we consider is specified as

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \epsilon_i \sim N\left(0, \sigma^2\right), i = 1, \ldots, n.$$

with $x_{i1} = 1$. Let $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\right)'$. We consider two different null hypotheses, both concerning $\boldsymbol{\beta}_1$. The first one is to test $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ against $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^*$. The other is to test

$H_0 : R\boldsymbol{\beta}_1 = \mathbf{r}$ against $H_1 : R\boldsymbol{\beta}_1 \neq \mathbf{r}$. To do Bayesian analysis, the conjugate priors for $\boldsymbol{\beta}$ and σ^2 can be specified as the normal distribution and the inverse gamma distribution, respectively,

$$\boldsymbol{\beta}|\sigma^2 \sim N\left(\boldsymbol{\mu}_0, \sigma^2 \mathbf{V}_0\right), \sigma^2 \sim IG\left(a, b\right),$$

where $\boldsymbol{\mu}_0$, $\mathbf{V}_0$ and a, b are hyperparameters. As a result, the posterior distributions are available analytically.

For simplicity, we consider the case in which $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$, where $x_{i1} = 1$, $x_{i2}, x_{i3}, x_{i4} \sim N(0, 1)$. The true parameter values used to simulate data are given as $\sigma^2 = 0.01, \beta_1 = 0.3, \beta_2 = 0.2, \beta_3 = 0.1C, \beta_4 = 0.5C$ for $C = 0, 0.1, 0.3, 0.5$, where C is used to control the difference between the true value and zero. The number of replications is set at $1,000$ while three sample sizes are considered, $n = 50, 100, 150$. Each of four null hypotheses is tested, $\beta_3 = 0$, or $\beta_4 = 0$, or $\beta_3 = \beta_4 = 0$, or $\beta_3 + \beta_4 = 0$, in every replication. To make the priors vague, the hyperparameters are specified at

$$\boldsymbol{\mu}_0 = (0, 0, 0, 0)', \mathbf{V}_0 = 1000 \times \mathbf{I}_4, a = 0.0001, b = 0.0001.$$

In each replication, we draw $5,000$ i.i.d. random samples from the posterior distribution and then use the posterior samples to compute $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$. Also computed is **Wald** for the purpose of comparison. The Wald test is feasible because MLE is easy to obtain in this application.

Table 2 reports the size and the power of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ and **Wald** for a nominal size of 5%. In all cases, the size distortion for $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ is very small and the two tests perform similarly in terms of size. The size approaches 5% as the sample size increases. Moreover, in all cases, the power of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ is comparable to that of **Wald**. As C increases, the power of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ approaches 100%. Similarly, as the sample size increases, the power of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ approaches 100%.

5.2 Hypothesis testing in a misspecified linear regression model

To examine the performance of $\mathbf{T}_S(\mathbf{y}, \boldsymbol{\theta}_0)$, following Zhou et al. (2012), we consider the point-null hypothesis testing problem in a linear regression model with heteroskedastic errors. In particular, we adopt the design of Zhou et al. (2012) by simulating data from

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \sigma_i \epsilon_i, \epsilon_i \sim N(0, 1), i = 1, \ldots, n,$$

where $\beta_1 = 0.1C$, $\beta_2 = 0.5C$, $C = 0, 0.1, 0.3, 0.5$, $x_{i1} = 1$, $x_{i2} \sim U(-5, 5)$, σ_i is the ith diagonal element in matrix $\sqrt{X(X'X)^{-1}X}$ with $X = (x_1, x_2)$, $x_i = (x_{1j}, \ldots, x_{nj})'$, $j = 1, 2$. We then obtain the size and power of $\mathbf{T}_S(\mathbf{y}, \boldsymbol{\theta}_0)$. We also obtain the size and power $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ and **Wald** without taking heteroskedasticity into account.

Table 2: The size and power of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ and **Wald** in the linear regression model

		Empirical Size		Empirical Power					
		$C = 0$		$C = 0.1$		$C = 0.3$		$C = 0.5$	
n	H_0	$\mathbf{T}(\mathbf{y}, \boldsymbol{\beta}_{10})$	**Wald**	$\mathbf{T}(\mathbf{y}, \boldsymbol{\beta}_{10})$	**Wald**	$\mathbf{T}(\mathbf{y}, \boldsymbol{\beta}_{10})$	**Wald**	$\mathbf{T}(\mathbf{y}, \boldsymbol{\beta}_{10})$	**Wald**
50	$\beta_3 = 0$	4.50	5.10	10.40	11.00	55.80	57.30	92.00	92.20
	$\beta_4 = 0$	6.50	7.10	92.00	92.5%	100	100	100	100
	$\beta_3 = \beta_4 = 0$	6.60	7.50	88.80	89.70	100	100	100	100
	$\beta_3 + \beta_4 = 0$	6.20	6.70	83.30	84.00	100	100	100	100
100	$\beta_3 = 0$	5.50	5.80	20.20	20.40	82.00	82.80	99.90	100
	$\beta_4 = 0$	4.60	5.00	99.70	99.70	100	100	100	100
	$\beta_3 = \beta_4 = 0$	5.70	6.00	99.50	99.50	100	100	100	100
	$\beta_3 + \beta_4 = 0$	6.00	6.20	98.60	98.60	100	100	100	100
150	$\beta_3 = 0$	5.30	5.40	24.40	24.60	95.90	95.90	100	100
	$\beta_4 = 0$	5.20	5.30	100	100	100	100	100	100
	$\beta_3 = \beta_4 = 0$	5.40	5.60	100	100	100	100	100	100
	$\beta_3 + \beta_4 = 0$	4.20	4.20	99.80	99.80	100	100	100	100

The estimation and hypothesis testing problems are the same as those in Section 5.1. The empirical size and power are reported in Table 3. It is clear that the empirical size of $\mathbf{T}_S(\mathbf{y}, \boldsymbol{\theta}_0)$ in all cases are close to the nominal level 5%. In particular, as the sample size increases, the size distortion becomes smaller. In a contrast, the size of both $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ and **Wald** does not seem to converge to the nominal level even when $n = 200$. Moreover, the empirical power of $\mathbf{T}_S(\mathbf{y}, \boldsymbol{\theta}_0)$ is good in all cases.

Table 3: The size and power of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$, $\mathbf{T}_S(\mathbf{y}, \boldsymbol{\theta}_0)$ and **Wald** in the misspecified linear regression model.

		Empirical Size			Empirical Power								
		$C = 0$			$C = 0.1$			$C = 0.3$			$C = 0.5$		
n	H_0	$\mathbf{T}$	$\mathbf{T}_S$	**Wald**	$\mathbf{T}$	$\mathbf{T}_S$	**Wald**	$\mathbf{T}$	$\mathbf{T}_S$	**Wald**	$\mathbf{T}$	$\mathbf{T}_S$	**Wald**
100	$\beta_3 = 0$	5.40	6.60	5.50	10.60	12.30	11.00	54.40	57.90	54.90	93.20	93.90	93.40
	$\beta_4 = 0$	8.90	6.80	9.10	100	100	100	100	100	100	100	100	100
	$\beta_3 = \beta_4 = 0$	9.40	7.50	9.90	100	100	100	100	100	100	100	100	100
	$\beta_3 + \beta_4 = 0$	6.30	6.80	6.40	96.80	100	96.80	100	100	100	100	100	100
150	$\beta_3 = 0$	5.40	5.70	5.40	18.00	20.00	18.20	88.40	89.40	88.40	99.90	100	100
	$\beta_4 = 0$	9.00	5.50	9.20	100	100	100	100	100	100	100	100	100
	$\beta_3 = \beta_4 = 0$	7.20	5.60	7.70	100	100	100	100	100	100	100	100	100
	$\beta_3 + \beta_4 = 0$	6.20	5.50	6.50	100	100	100	100	100	100	100	100	100
200	$\beta_3 = 0$	4.10	4.50	4.30	29.50	31.30	29.70	99.00	99.10	99.00	100	100	100
	$\beta_4 = 0$	8.50	4.70	9.10	100	100	100	100	100	100	100	100	100
	$\beta_3 = \beta_4 = 0$	7.20	5.00	7.30	100	100	100	100	100	100	100	100	100
	$\beta_3 + \beta_4 = 0$	4.70	4.70	4.80	100	100	100	100	100	100	100	100	100

5.3 Hypothesis testing in a discrete choice model

The third model in the simulation study is a simplified version of the model of Li (2006), where the effects of attendance on high school completion and future youth unemployment were studied. As noted in Li (2006), the likelihood function involves multiple integrals and discrete and censored variables. Consequently, the likelihood function and the corresponding derivatives are not easy to evaluate. That is why Li (2006) introduces an MCMC approach for statistical analysis. We perform hypothesis testing in the discrete choice model with latent variables.

Let $z_i = 1, 2, 3, 4$ denote the high school grade completed by individual i which is, by definition, an ordered integer. Let y_i denote the latent outcome corresponding to z_i. The first part of the model is an ordered probit defined as

$$\begin{cases} y_i = \beta_0 + \beta_1 x_i + \epsilon_i, & \epsilon_i \sim N\left(0, \sigma^2\right), \gamma_{z_i} < y_i < \gamma_{z_i+1}, \\ \gamma_1 = -\infty, \gamma_2 = 0, & \gamma_2 < \gamma_3 < \gamma_4, \gamma_4 = 1, \gamma_5 = \infty, \end{cases}$$

where $i = 1, \ldots, n$ with n being the total number of individuals, ϵ_i is an individual level random error term, σ^2 is the variance of the error term, $\{\gamma_j\}_{j=1}^5$ are the cutoff points, x_i contains some covariates which are assumed to be exogenous. For the purpose of simulating data, we simply assume x_i is univariate and $x_i \sim N(0, 1)$.

Furthermore, let ω_i denote the proportion of time during which individual i is unemployed, $\tilde{y}_i$ is the latent outcome corresponding to ω_i, and $\tilde{y}_i$ is limited as,

$$\tilde{y}_i \begin{cases} \leq 0, & \omega_i = 0, \\ = \omega_i, & 0 < \omega_i < 1, \\ \geq 1, & \omega_i = 1. \end{cases}$$

Then the censored regression is,

$$\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x + \tilde{\epsilon}_i, \tilde{\epsilon}_i \sim N\left(0, \tilde{\sigma}^2\right). \tag{21}$$

The two error terms are correlated, that is,

$$\begin{pmatrix} \epsilon_i \\ \tilde{\epsilon}_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & \tilde{\sigma}^2 \end{pmatrix} \right) := N(0, \Sigma).$$

In the simulation study, the null and alternative hypotheses are,

$$H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0.$$

To calculate the size and power of the proposed statistic, three sample sizes are considered, $n = 100$, 250, and 500. In each case, we compute the empirical size when $\beta_1 = 0$ at a

Table 4: The size and power of $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ in the discrete choice model

	Empirical Size	Empirical Power		
	$\beta_1 = 0$	$\beta_1 = 0.1$	$\beta_1 = 0.2$	$\beta_1 = 0.4$
$n = 100$	4.20	11.00	23.00	75.80
$n = 250$	5.20	24.00	65.00	100
$n = 500$	4.60	49.40	97.20	100

nominal size of 5%. We also compute the power when $\beta_1 = 0.1, 0.2$ and 0.4. The number of replications is 500. The true values of other parameters are set at,

$$\beta_0 = 1, \tilde{\beta}_0 = 0.01, \tilde{\beta}_1 = 0.1, \Sigma = \begin{pmatrix} 1 & -0.01 \\ -0.01 & 0.1 \end{pmatrix}, \gamma_3 = 0.67.$$

These values are close to those reported in Li (2006) based on actual data.

Following Li (2006), we use the following vague priors to do Bayesian analysis,

$$\boldsymbol{\beta} = \left(\beta_0, \beta_1.\tilde{\beta}_0, \tilde{\beta}_1\right)' \sim N\left(0, 1000 \times \mathbf{I}_4\right), \ \Sigma \sim IW\left(6, 6 \times \mathbf{I}_2\right), \gamma_3 \sim Beta\left(1, 1\right),$$

where IW denotes the inverted Wishart distribution and $Beta$ denotes the Beta distribution.

We run MCMC to obtain 10,000 random samples. After dropping the first 4,000 samples, we treat the remaining 6,000 sample as effective draws from the posterior distribution. Let $\left\{\beta_1^{[j]}\right\}_{j=1}^{J}$ denote the effective posterior draws. From (13), the proposed statistic can be simply calculated as

$$\widehat{\mathbf{T}}\left(\mathbf{y}, \beta_1 = 0\right) = \frac{\frac{1}{J}\sum_{j=1}^{J}\left(\beta_1^{[j]}\right)^2}{\frac{1}{J}\sum_{j=1}^{J}\left(\beta_1^{[j]} - \overline{\overline{\beta}}_1\right)^2}, \ \overline{\overline{\beta}}_1 = \frac{1}{J}\sum_{j=1}^{J}\beta_1^{[j]}.$$

Other test statistics, such as BFs and the Wald statistic, are harder to obtain due to the presence of latent variables.

The empirical size and power of the proposed test are reported in Table 4 for a nominal size of 5%. It is obvious that the empirical size is close to the nominal size in all cases, even when the sample size is only 100. When β_1 becomes further and further away from 0, the power increases and approaches 100. Furthermore, as the sample size increases, the power increases in all cases.

6 Empirical Examples

We then consider two empirical studies using real data. The first model is the full version of the discrete choice model of Li (2006). The second model is the stochastic volatility model with leverage effect. For both models, it is well-known that the observed-data likelihood function is intractable due to latent variables. As a result, the observed-data likelihood function and its derivatives are very difficult to evaluate, and hence it is advantageous to use the proposed statistic over existing statistics for hypothesis testing.

6.1 Hypothesis testing in a discrete choice model

In the first empirical study, we consider the same model and use the same data set as in Li (2006). Let z_{hi} denote the high school grade completed by individual i, and y_{hi} denote the latent outcome corresponding to z_{hi}, where h is the schooling outcome. Let $z_{hi} = 1$ if individual i drops out of high school after completing the ninth grade, $z_{hi} = 2$ if he drops out after completing the tenth grade, $z_{hi} = 3$ if he drops out after completing the eleventh grade, and $z_{hi} = 4$ if he completes high school. An ordered probit is specified as

$$\begin{cases} y_{hi} = \boldsymbol{\beta}_h' \boldsymbol{x}_{hi} + \epsilon_{hi}, & \epsilon_{hi} \sim N\left(0, \sigma_h^2\right), \gamma_{z_{hi}} < y_{hi} < \gamma_{z_{hi}+1} \\ \gamma_1 = -\infty, \gamma_2 = 0, & \gamma_2 < \gamma_3 < \gamma_4, \gamma_4 = 1, \gamma_5 = \infty \end{cases}, \tag{22}$$

where $\boldsymbol{x}_{hi}$ is a $k_h \times 1$ vector incorporating individual-level variables, including base year cognitive test score, parental income, parental education, number of siblings, gender, race, county-level employment growth rate between 1980 and 1982, a fourth-order polynomial in age and a fourth-order polynomial in the time eligible to drop out.

Furthermore, let ω_{ui} represent the proportion of time when individual i is unemployed, y_{ui} the latent outcome corresponding to ω_{ui}, and y_{ui} is limited as,

$$y_{ui} \begin{cases} \leq 0, & \omega_{ui} = 0, \\ = \omega_{ui}, & 0 < \omega_{ui} < 1, \\ \geq 1, & \omega_{ui} = 1. \end{cases} \tag{23}$$

Thus, the censored regression is,

$$y_{ui} = \boldsymbol{\beta}_u' \boldsymbol{x}_{ui} + \boldsymbol{s}_i' \boldsymbol{\eta} + \epsilon_{ui}, \epsilon_{ui} \sim N\left(0, \sigma_u^2\right), \tag{24}$$

where x_{ui} is a $k_u \times 1$ vector incorporating observed variables, including base year cognitive test score, parental income, parental education, number of siblings, gender, race, age and a dummy variable indicating any post-secondary education.

In Equation (24), $\boldsymbol{s}_i$ is a 4×1 vector consisting of dummy variables indicating the high school grade completed by individual i. In other words, $\boldsymbol{s}_i = (s_{i,1}, s_{i,2}, s_{i,3}, s_{i,4})'$, and if $s_{i,z_{hi}} = 1$ then $s_{i,j} = 0$, $j \neq z_{hi}$. Besides, $\boldsymbol{\eta}$ indicates the 4×1 vector of the effect of high school completion on unemployment. For simplicity, $\boldsymbol{\eta}$ is assumed to be the same across schools. This assumption is different from that in Li (2006) although our empirical results are almost the same as those in Li. The random terms are assumed to be correlated,

$$\begin{pmatrix} \epsilon_{hi} \\ \epsilon_{ui} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_h^2 & \sigma_{hu} \\ \sigma_{hu} & \sigma_u^2 \end{pmatrix}\right) := N(0, \Sigma).$$

In total, there are 34 parameters in the model.

As noted in Li (2006), the MLE is difficult to obtain. Hence, the MCMC technique is implemented. We adopted the same priors as Li which are listed in the following,

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_h', \boldsymbol{\beta}_u'\right)' \sim N\left(0_{k\times 1}, 1000 \times \mathbf{I}_k\right), \ \Sigma \sim IW\left(6, 6 \times \mathbf{I}_2\right),$$

$$\boldsymbol{\eta} \sim N\left(0, \mathbf{I}_4\right), \ \gamma_3 \sim Beta\left(1, 1\right),$$

where $k = k_h + k_u$.

The dataset contains 5,238 students from 871 schools. For more details about the data, one can refer to Li (2006). We run MCMC for 20,000 times. After dropping the first 4,000 samples, we treat the remaining 16,000 as effective draws. Posterior means and posterior standard errors are reported in Table 5, all of which are very close to those reported in Li (2006).

Suppose one is interested in testing that the marginal effects of father's education level and mother's education level on the completion of high school can be ignored or not. The null hypothesis can be written as $H_0 : \beta_{4h} = \beta_{5h} = 0$. With the MCMC output, we can very easily compute the statistic. We also compute $\widehat{\log BF}_{10}$ and $\widehat{\mathbf{T}}_{LLY}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$. The three test statistics and their numerical standard errors are reported in Table 6.[3]

According to Table 6, both $\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{\theta}_0\right) - 2$ and $\widehat{\mathbf{T}}_{LLY}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ take very large values, indicating that the null hypothesis is overwhelmingly rejected. This conclusion is consistent with that by $\widehat{\log BF}_{10}$, which strongly supports the alternative hypothesis. Furthermore, their numerical standard errors are all small relative to the values of the statistics. Finally, in spite of the same conclusion reached, the CPU time required to compute the test statistics is vastly different. The proposed statistic is more than 1700 times and nearly 13000 times faster to compute than $\widehat{\mathbf{T}}_{LLY}\left(\mathbf{y}, \boldsymbol{\theta}_0\right)$ and $\widehat{\log BF}_{10}$ after MCMC outputs are available. An additional advantage that does not reflect in the CPU time is that the proposed statistic only needs MCMC output from the alternative model while the other two statistics require MCMC output for both the null and alternative models.

6.2 Hypothesis testing in a stochastic volatility model

Stochastic volatility (SV) models with leverage effect have been widely used in finance; see Harvey and Shephard (1996) and Aït-Sahalia et al. (2017). Following Yu (2005), the SV model with leverage effect is defined as,

$$\begin{cases} r_t = \exp\left(h_t/2\right)\epsilon_t, \\ h_{t+1} = \mu + \phi\left(h_t - \mu\right) + \sigma\varepsilon_{t+1}, h_0 = \mu, \end{cases}$$

[3] We use the marginal likelihood method of Chib (1995) to compute the BF and its NSE.

Table 5: Posterior means and standard errors of parameters in the discrete choice model

	$E(\cdot\|Data)$	$SE(\cdot\|Data)$
	High school completion y_h	
Constant	0.9474	0.2119
Parental income	0.0110	0.0262
Base year cognitive test	0.4413	0.0370
Father's education	0.0456	0.0131
Mother's education	0.0627	0.0159
Number of siblings	-0.0370	0.0153
Female	-0.0694	0.0534
Minority	0.3840	0.0664
County employment growth	-0.0132	0.0047
Age	-0.4150	0.0853
Age^2	-0.1887	0.0766
Age^3	-0.0333	0.0468
Age^4	0.0311	0.0148
Time eligible to drop out	0.0932	0.0696
Time^2	0.0905	0.0473
Time^3	-0.0090	0.0106
Time^4	-0.0094	0.0053
	Proportion of time unemployed ω_u	
Parental income	-0.0275	0.0056
Base year cognitive test	-0.0392	0.0071
Father's education	-0.0020	0.0025
Mother's education	-0.0043	0.0030
Number of siblings	0.0049	0.0034
Post-secondary education	-0.0113	0.0138
Female	0.0621	0.0112
Minority	0.0826	0.0131
Age	-0.0058	0.0126
Completing ninth grade($\boldsymbol{\eta}_1$)	0.1925	0.0705
Completing tenth grade($\boldsymbol{\eta}_2$)	0.1211	0.0530
Completing eleventh grade($\boldsymbol{\eta}_3$)	0.1187	0.0492
Completing high school($\boldsymbol{\eta}_4$)	0.0083	0.0416
	Covariance matrix Σ	
σ_h^2	0.9450	0.0914
σ_u^2	0.1215	0.0039
σ_{hu}	-0.0099	0.0191
	Cutoff point	
γ_3	0.6684	0.0220

Table 6: The test statistics, $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)$, $\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\boldsymbol{\theta}_0)$, $\widehat{\log BF_{10}}$, their NSE and their computational time in the discrete choice model

	$\beta_4=\beta_5=0$		
	Value	NSE	CPU Time (seconds)†
$\widehat{\mathbf{T}}(\mathbf{y},\boldsymbol{\theta}_0)-2$	43.39	1.59	22.54
$\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\boldsymbol{\theta}_0)$	2502.00	89.57	39,096.79
$\widehat{\log BF_{10}}$	5.2019	1.03	292,886.45

with

$$\begin{pmatrix}\epsilon_t\\ \varepsilon_{t+1}\end{pmatrix}\overset{i.i.d.}{\sim}N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&\rho\\ \rho&1\end{pmatrix}\right),$$

where r_t is the return at time t, h_t the latent volatility at period t. In this model, ρ is the parameter that captures the leverage effect when it is negative. Hence, we test $H_0:\rho=0$ against $H_1:\rho\neq 0$. In this example, we use two different datasets for hypothesis testing. For each dataset, we compute the proposed statistic, $\mathbf{T}_{LLY}(\mathbf{y},\boldsymbol{\theta}_0)$ and $\widehat{\log BF_{10}}$.[4]

Let $\left\{\rho^{[j]}\right\}_{j=1}^{J}$ denote the effective posterior draws for ρ under H_1. The proposed statistic is simply calculated as

$$\widehat{\mathbf{T}}(\mathbf{y},\rho=0)=\frac{\frac{1}{J}\sum_{j=1}^{J}\left(\rho^{[j]}\right)^2}{\frac{1}{J}\sum_{j=1}^{J}\left(\rho^{[j]}-\overline{\overline{\rho}}\right)^2},\ \overline{\overline{\rho}}=\frac{1}{J}\sum_{j=1}^{J}\rho^{[j]}.$$

On the contrary, computing $\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\boldsymbol{\theta}_0)$ and $\widehat{\log BF_{10}}$ require substantially higher coding efforts and extra CPU time.

The first dataset consists of daily returns on Pound/Dollar exchange rates from October 1, 1981 to June 28, 1985 with sample size 945. The series r_t is the daily mean-corrected returns. The following vague priors are used:

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

We draw 50,000 from the posterior distribution and discard the first 20,000 as burn-in samples. Then we store every 5th value of the remaining samples as effective random samples. The estimation results are reported in Table 7.

Table 8 reports the proposed statistic, $\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\boldsymbol{\theta}_0)$ and $\widehat{\log BF_{10}}$ and the NSEs for the first two statistics. Since the observed-data likelihood function is expensive to compute, the NSE of BF is too difficult to obtain and not reported. $\widehat{\log BF_{10}}$ strongly supports the null hypothesis, that is, the SV model without leverage effect. $\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\boldsymbol{\theta}_0)$ takes a very small value, suggesting that we cannot reject the null hypothesis. When the null

[4]Again we use the marginal likelihood method of Chib (1995) to compute the BF.

Table 7: Posterior means and standard errors of parameters in the SV model

	H_1		H_0	
Parameter	Mean	SE	Mean	SE
μ	-0.5776	0.3487	-0.6608	0.3164
ϕ	0.9849	0.0097	0.9793	0.0127
ρ	-0.0941	0.1507	-	-
τ	0.1553	0.0243	0.1618	0.0360

Table 8: The test statistics, $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$, $\widehat{\mathbf{T}}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$, $\log \widehat{BF}_{10}$, their NSE, and the CPU time in the SV model

	$\widehat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0) - 1$	$\widehat{\mathbf{T}}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$	$\log \widehat{BF}_{10}$
Value	0.3893	0.2883	-10.1235
NSE	0.0255	0.2028	-
CPU Time (seconds)	0.9411	549.0631	3,701.2241

hypothesis is true, we know that $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) - 1 \xrightarrow{d} \chi^2(1)$. It can be found that $\widehat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0) - 1$ is very closed to $\widehat{\mathbf{T}}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$, also suggesting that we cannot reject the null hypothesis. Finally, our proposed statistic has a smaller NSE than $\widehat{\mathbf{T}}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$.

The second dataset contains 1,822 daily returns of the Standard & Poor (S&P) 500 index, covering the period between January 3, 2005 and March 28, 2012. We use the same priors and method as before to estimate the model with and without leverage effect. The estimation results are reported in Table 9.

Table 9: The estimates of the SV models for the S&P500 returns.

	H_1		H_0	
Parameter	Mean	SE	Mean	SE
μ	-10.8800	0.1751	-11.2200	0.3349
ϕ	0.9804	0.0039	0.9897	0.0042
ρ	-0.7151	0.0422	-	-
τ	0.2057	0.0178	0.1705	0.0169

The three test statistics and the NSEs for the first two statistics are reported in Table 10. Contrary to the case of Pound/Dollar returns, all three statistics strongly support the alternative hypothesis. Both $\widehat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0) - 1$ and $\widehat{\mathbf{T}}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ reject the null hypothesis under the 99% significance level. Similarly, $\log \widehat{BF}_{10}$ strongly supports the alternative hypothesis. However, the proposed statistic is nearly 1000 times and more than 6000 times faster to compute than $\widehat{\mathbf{T}}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ and $\log \widehat{BF}_{10}$ after MCMC outputs are available.

Table 10: The performance of alternative statistics in the SV model.

	$\widehat{\mathbf{T}}(\mathbf{y},\boldsymbol{\theta}_0)-1$	$\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\boldsymbol{\theta}_0)$	$\log\widehat{BF_{10}}$
Value	286.7944	8.2419	51.9582
NSE	0.6915	0.6849	-
CPU Time (seconds)	1.2922	1,256.7768	7,785.6888

7 Conclusion

In this paper, a new statistic based on posterior distribution is proposed to test for a point null hypothesis under the correct model specification. It can be treated as the posterior version of the Wald test. Compared with existing methods, the proposed statistic has many important advantages. First, it is well-defined under improper prior distributions. Second, it avoids Jeffreys-Lindley-Bartlett's paradox. Third, its asymptotic distribution is a χ^2 distribution under the null hypothesis and repeated sampling. This property is the same as the Wald statistic so that the critical values can be easily obtained. Fourth, it is very easy to compute as it is based on the posterior mean and posterior variance of the parameters of interest. Fifth, it can be used to test hypotheses that impose non-linear relationships among the parameters of interest, for which the BF is difficult to use. Sixth, for latent variable models for which the MLE and the Wald test are more difficult to obtain, the proposed statistic is the by-product of posterior sampling. Finally, only posterior sampling for the alternative hypothesis is needed for the proposed statistic. We also propose a test statistic based on an artificial posterior distribution that is robust under model misspecification that inherits many nice properties of the first test statistic.

The finite-sample properties of the proposed statistic under the correct model specification is examined in a linear regression model and in a discrete choice model with latent variables. In the linear regression models, the Wald statistic is feasible and compared with the proposed test. Simulation results show that the proposed test has little size distortion even when the sample size is small and its size and power are very similar to those of the Wald test when a vague prior is used. In the discrete choice model, the proposed test continues to enjoy small size distortions even when the sample size is small. The power increases rapidly when the sample size increases or when the difference between the null and alternative hypotheses increases. We also check the finite-sample behavior of the proposed robust test in a linear regression model with heteroskedastic errors. The simulation results also suggest that the proposed test has good finite-sample properties.

We apply the method to two models using real data. The first one is a discrete choice model, and the second is an SV model. In both models, there are latent variables. Due to the presence of latent variables, the Wald statistic is very difficult to obtain and because the maximum likelihood method is difficult to use. While both the BF and the test proposed by LLY (2015) are feasible to compute based on MCMC output, they are much more expensive to compute than the proposed statistic with longer CPU time after MCMC output is available. The empirical conclusion obtained by these three methods is the same in both empirical applications.

8 Appendix

Before we prove Lemma 3.1, we need to introduce and prove three lemmas.

Lemma 8.1. *Let* $\mathbf{N}_0(\delta) = \left\{\boldsymbol{\vartheta} : \left\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_n^0\right\| \leq \delta\right\}$. *If Assumptions 1-7 hold true, then for any* $\varepsilon > 0$, *there exists* $\delta(\varepsilon) > 0$ *such that*

$$P\left(\sup_{\boldsymbol{\vartheta}\in N_0(\delta(\varepsilon))}\left|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right| < \varepsilon\right) \to 1, \tag{25}$$

and

$$P\left(\sup_{\boldsymbol{\vartheta}\in N_0(\delta(\varepsilon)),\|\mathbf{r}_0\|=1}\left|1 - \mathbf{r}_0'\bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right)\bar{\mathbf{H}}_n(\boldsymbol{\vartheta})\bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right)\mathbf{r}_0\right| < \varepsilon\right) \to 1.$$

where $\mathbf{r}_0$ *is a* q*-dimensional vector.*

Proof. First, we can show that

$$\begin{aligned}
& \left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\| \\
= & \left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right) + \bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right) - \mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right) + \mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\| \\
\leq & \left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\| + \left\|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right) - \mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\| + \left\|\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\|.
\end{aligned} \tag{26}$$

For any ε, there exists a $\delta(\varepsilon) > 0$ such that

$$P\left(\sup_{\boldsymbol{\vartheta}\in N(\boldsymbol{\vartheta}_n^0,\delta(\varepsilon))}\left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\| < \frac{\varepsilon}{3}\right) \to 1. \tag{27}$$

by Assumptions 1-7, we have the uniform convergence of $l_t^{(2)}(\boldsymbol{\vartheta})$ and $\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_n^0 \xrightarrow{p} 0$ (Gallant and White, 1988). Hence, by Assumption 3 that $l_t^{(2)}(\boldsymbol{\vartheta})$ is almost surely continuous at $\boldsymbol{\vartheta}_n^0$, we have

$$P\left(\left\|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right) - \mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\| < \frac{\varepsilon}{3}\right) \to 1,\ P\left(\left\|\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\| < \frac{\varepsilon}{3}\right) \to 1. \tag{28}$$

Let $A_n(\varepsilon) = \left\{ \sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon))} \left\| \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}_n^0) \right\| < \frac{\varepsilon}{3} \right\}$, $B_n(\varepsilon) = \left\{ \left\| \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}_n^0) - \mathbf{H}_n(\boldsymbol{\vartheta}_n^0) \right\| < \frac{\varepsilon}{3} \right\}$, and $C_n(\varepsilon) = \left\{ \left\| \mathbf{H}_n(\boldsymbol{\vartheta}_n^0) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right) \right\| < \frac{\varepsilon}{3} \right\}$. Then, we have

$$\begin{aligned} & P\left(\sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon))} \left\{ \left\| \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}_n^0) \right\| + \left\| \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}_n^0) - \mathbf{H}_n(\boldsymbol{\vartheta}_n^0) \right\| + \left\| \mathbf{H}_n(\boldsymbol{\vartheta}_n^0) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right) \right\| \right\} < \varepsilon \right) \\ \geq\ & P\left(A_n(\varepsilon) \cap B_n(\varepsilon) \cap C_n(\varepsilon) \right). \end{aligned}$$

From (27) and (28), the probability of the complementary of $A_n(\varepsilon) \cap A_n(\varepsilon) \cap A_n(\varepsilon)$ is

$$\begin{aligned} & P\left(\left(A_n(\varepsilon) \cap B_n(\varepsilon) \cap C_n(\varepsilon) \right)^c \right) \\ =\ & P\left(A_n(\varepsilon)^c \cup B_n(\varepsilon)^c \cup C_n(\varepsilon)^c \right) \leq P\left(A_n(\varepsilon)^c \right) + P\left(B_n(\varepsilon)^c \right) + P\left(C_n(\varepsilon)^c \right) \to 0, \end{aligned}$$

which implies

$$P\left(A_n(\varepsilon) \cap B_n(\varepsilon) \cap C_n(\varepsilon) \right) \to 1.$$

Hence, by (26), for any $\varepsilon > 0$

$$\begin{aligned} & P\left(\sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon))} \left\| \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right) \right\| < \varepsilon \right) \\ \geq\ & P\left(\sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon))} \left\| \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}_n^0) \right\| + \left\| \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}_n^0) - \mathbf{H}_n(\boldsymbol{\vartheta}_n^0) \right\| + \left\| \mathbf{H}_n(\boldsymbol{\vartheta}_n^0) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right) \right\| < \varepsilon \right) \\ \to\ & 1. \end{aligned} \tag{29}$$

Note that

$$\begin{aligned} & \sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)), \|\mathbf{r}_0\|=1} \left| 1 - \mathbf{r}_0' \bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right) \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) \bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right) \mathbf{r}_0 \right| \\ =\ & \sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)), \|\mathbf{r}_0\|=1} \left| 1 + \mathbf{r}_0' \left(-\bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right) \right) \left(-\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) \right) \left(-\bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right) \right) \mathbf{r}_0 \right| \\ =\ & \sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)), \|\mathbf{r}_0\|=1} \left| \mathbf{r}_0' \left(-\bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right) \right) \left[-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right) + \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) \right] \left(-\bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right) \right) \mathbf{r}_0 \right| \\ \leq\ & \lambda_n \sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)), \|\mathbf{r}_0\|=1} \left| \mathbf{r}_0' \left(\bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right) \right) \mathbf{r}_0 \right| \\ \leq\ & \lambda_n \sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)), \|\mathbf{r}_0\|=1} \left\| \mathbf{r}_0' \right\| \left\| \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right) \right\| \left\| \mathbf{r}_0 \right\| \\ =\ & \lambda_n \sup_{N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)), \|\mathbf{r}_0\|=1} \left\| \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right) - \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) \right\|, \end{aligned}$$

where λ_n is the smallest eigenvalue of $-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)$. Then, from (29), for any $\varepsilon > 0$,

$$P\left(\sup_{\boldsymbol{\vartheta} \in N(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)), \|\mathbf{r}_0\|=1} \left| 1 - \mathbf{r}_0' \bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right) \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) \bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right) \mathbf{r}_0 \right| < \varepsilon \right) \to 1. \tag{30}$$

■

Lemma 8.2. *Let $X_1, ..., X_q$ be i.i.d. random variables. If $E|X_1|^{2k} < \infty$ for $k > 0$, then*

$$E\left[\left(\max_{i\in\{1,...,q\}} |X_i|\right)^k\right] < \sqrt{2}e^{5/3}\frac{q+1}{\sqrt{q}}\left[E|X_1|^{2k}\right]^{1/2}.$$

Proof. Let $\delta = k\rho^{-1}$, $0 < \rho \leq 1/2$. From Gribkova (1995), the following inequality

$$E\left[\left|\max_{i\in\{1,...,q\}} X_i\right|^k\right] < C(\rho)\left\{E|X_1|^\delta g^{-1}\left(\frac{q}{q+1}\right)\right\}^\rho,$$

holds for $q \geq 2\rho + 1$, where $C(\rho) = 2\sqrt{\rho}\exp(\rho + 7/6)$ and $g(u) = u(1-u)$. By setting $\rho = 0.5$, it can be shown that, for $q \geq 2$,

$$E\left[\left|\max_{i\in\{1,...,q\}} X_i\right|^k\right] < C(0.5)\left\{E|X_1|^\delta g^{-1}\left(\frac{q}{q+1}\right)\right\}^{1/2} = \sqrt{2}e^{5/3}\frac{q+1}{\sqrt{q}}\left[E|X_1|^{2k}\right]^{1/2}. \tag{31}$$

For $q = 1$, by Jensen's Inequality,

$$E\left[\left|\max_{i\in\{1,...,q\}} X_i\right|^k\right] = E\left[|X_1|^k\right] \leq \left[E|X_1|^{2k}\right]^{1/2}.$$

Then,

$$E\left[\left|\max_{i\in\{1,...,q\}} X_i\right|^k\right] < \sqrt{2}e^{5/3}\frac{1+1}{\sqrt{1}}\left[E|X_1|^{2k}\right]^{1/2}. \tag{32}$$

From (31) and (32), for $k > 0$ and $q \geq 1$, we get,

$$E\left[\left|\max_{i\in\{1,...,q\}} X_i\right|^k\right] < \sqrt{2}e^{5/3}\frac{q+1}{\sqrt{q}}\left[E|X_1|^{2k}\right]^{1/2}. \tag{33}$$

Let $Y_i = |X_i|$. Then by (33) we have

$$\begin{aligned} E\left[\left(\max_{i\in\{1,...,q\}} |X_i|\right)^k\right] &= E\left[\left|\max_{i\in\{1,...,q\}} |X_i|\right|^k\right] = E\left[\left|\max_{i\in\{1,...,q\}} Y_i\right|^k\right] \\ &< \sqrt{2}e^{5/3}\frac{q+1}{\sqrt{q}}\left[E|Y_1|^{2k}\right]^{1/2} = \sqrt{2}e^{5/3}\frac{q+1}{\sqrt{q}}\left[E|X_1|^{2k}\right]^{1/2}. \end{aligned}$$

■

Lemma 8.3. *Let $\boldsymbol{\Sigma}_n = -\frac{1}{n}\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\vartheta}})$, $\mathbf{z}_n = \boldsymbol{\Sigma}_n^{-1/2}\left(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}}\right)$, $p(\boldsymbol{\vartheta}|\mathbf{y})$ and $p(\mathbf{y})$ be the posterior density and the marginal likelihood, respectively. Then, under Assumptions 1-9*

$$\lim_{n\to\infty} P\left(\int_{A_n} \|\mathbf{z}_n\|^2 \left|p(\mathbf{z}_n|\mathbf{y}) - (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right| d\mathbf{z}_n > \varepsilon\right) = 0, \tag{34}$$

where $A_n = \left\{\mathbf{z}_n : \widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n \in \boldsymbol{\Theta}\right\}$ is the support of $\mathbf{z}_n$.

Proof. The posterior density of $\mathbf{z}_n$ can be written as

$$p(\mathbf{z}_n|\mathbf{y}) = \frac{|\boldsymbol{\Sigma}_n|^{1/2} p(\mathbf{y}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta})}{p(\mathbf{y})} = \frac{|\boldsymbol{\Sigma}_n|^{1/2} p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) p\left(\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)}{p(\mathbf{y})}. \tag{35}$$

Applying the Taylor expansion to $\log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)$ at $\widehat{\boldsymbol{\vartheta}}$, we have

$$\begin{aligned}
&\log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) \\
= \;& \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) + \frac{1}{2}\mathbf{z}_n'\boldsymbol{\Sigma}_n^{1/2}\frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n \\
= \;& \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) - \frac{1}{2}\mathbf{z}_n'\boldsymbol{\Sigma}_n^{1/2}\left[-\frac{\partial^2 \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} - \frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} + \frac{\partial^2 \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right]\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n \\
= \;& \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) - \frac{1}{2}\mathbf{z}_n'\boldsymbol{\Sigma}_n^{1/2}\left[\boldsymbol{\Sigma}_n^{-1} - \frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} - \boldsymbol{\Sigma}_n^{-1}\right]\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n \\
= \;& \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) - \frac{1}{2}\mathbf{z}_n'\left[\mathbf{I}_q - \mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})\right]\mathbf{z}_n,
\end{aligned} \tag{36}$$

where

$$\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y}) = \mathbf{I}_q + \boldsymbol{\Sigma}_n^{1/2}\frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\boldsymbol{\Sigma}_n^{1/2},$$

with $\widetilde{\boldsymbol{\vartheta}}_1$ lying between $\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n$ and $\widehat{\boldsymbol{\vartheta}}$.

To prove (34), by Chen (1985) and Schervish (2012), we have

$$\begin{aligned}
&p(\mathbf{z}_n|\mathbf{y}) - (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) \\
= \;& p(\mathbf{y})^{-1}|\boldsymbol{\Sigma}_n|^{1/2} p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \Sigma_n^{1/2}\mathbf{z}_n\right) p\left(\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) - (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) \\
= \;& p(\mathbf{y})^{-1}|\boldsymbol{\Sigma}_n|^{1/2} p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) p\left(\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)\frac{p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)}{p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}_n\right)} - (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right),
\end{aligned}$$

and

$$p(\mathbf{y})^{-1}|\boldsymbol{\Sigma}_n|^{1/2} p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) \xrightarrow{p} \frac{(2\pi)^{-q/2}}{p\left(\boldsymbol{\vartheta}_n^0\right)}.$$

To verify (34), according to (36), it is sufficient to show

$$P\left(\int_{A_n}\|\mathbf{z}_n\|^2\left|\frac{p\left(\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)}{p\left(\boldsymbol{\vartheta}_n^0\right)}\exp\left[-\frac{\mathbf{z}_n'[\mathbf{I}_q - \mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})]\mathbf{z}_n}{2}\right] - \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right| d\mathbf{z}_n < \varepsilon\right) \to 1. \tag{37}$$

To ensure (37), by Assumption 9, it is enough to prove

$$P\left(\int_{A_n}\|\mathbf{z}_n\|^2\left|p\left(\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)\exp\left[-\frac{\mathbf{z}_n'[\mathbf{I}_q-\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})]\mathbf{z}_n}{2}\right]-p\left(\boldsymbol{\vartheta}_n^0\right)\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|d\mathbf{z}_n<\varepsilon\right)\to 1. \tag{38}$$

We now prove (38). Since the prior density is continuous at $\boldsymbol{\vartheta}_n^0$, given any $\varepsilon>0$, for any $\eta\in(0,1)$ satisfying

$$\varepsilon\geq\eta\left(\frac{q^2(1+\eta)\sqrt{(2k+1)(2k+3)}}{2(1-\eta)^{\frac{q+k+2}{2}}}+1\right),$$

$\exists\delta_1>0$, so that, for any $\boldsymbol{\vartheta}\in\mathbf{N}_0(\delta_1)$,

$$\left|p(\boldsymbol{\vartheta})-p\left(\boldsymbol{\vartheta}_n^0\right)\right|=\left|p\left(\widehat{\boldsymbol{\vartheta}}+\Sigma_n^{1/2}\mathbf{z}_n\right)-p\left(\boldsymbol{\vartheta}_n^0\right)\right|\leq\eta p\left(\boldsymbol{\vartheta}_n^0\right). \tag{39}$$

Furthermore, by Lemma 8.1, $\forall\eta>0$, $\exists\delta_2>0$, so that,

$$\lim_{n\to\infty}P\left(\sup_{\boldsymbol{\vartheta}\in N_0(\delta_2),\|\mathbf{r}_0\|=1}\left|1+\mathbf{r}_0'\boldsymbol{\Sigma}_n^{1/2}\frac{\partial^2\log p(\mathbf{y}|\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\boldsymbol{\Sigma}_n^{1/2}\mathbf{r}_0\right|<\eta\right)=1. \tag{40}$$

Let $\delta=\min\{\delta_1,\delta_2\}$ and define

$$A_{1n}=\left\{\mathbf{z}_n:\widehat{\boldsymbol{\vartheta}}+\Sigma_n^{1/2}\mathbf{z}_n\in\mathbf{N}_0(\delta)\right\},A_{2n}=\left\{\mathbf{z}_n:\widehat{\boldsymbol{\vartheta}}+\Sigma_n^{1/2}\mathbf{z}_n\in\boldsymbol{\Theta}\backslash\mathbf{N}_0(\delta)\right\},$$

and

$$C_n=\|\mathbf{z}_n\|^2\left|p\left(\widehat{\boldsymbol{\vartheta}}+\Sigma_n^{1/2}\mathbf{z}_n\right)\exp\left[-\frac{1}{2}\mathbf{z}_n'\left[\mathbf{I}_q-\mathbf{R}_n(\tilde{\boldsymbol{\vartheta}}_1)\right]\mathbf{z}_n\right]-p\left(\boldsymbol{\vartheta}_n^0\right)\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|. \tag{41}$$

The integration of C_n over the support A_n can be decomposed into two areas, A_{1n} and A_{2n}, that is,

$$J=\int_{A_n}C_nd\mathbf{z}_n=\int_{A_{1n}}C_nd\mathbf{z}_n+\int_{A_{2n}}C_nd\mathbf{z}_n:=J_1+J_2.$$

In the following, we will show

$$J_1=\int_{A_{1n}}C_nd\mathbf{z}_n\xrightarrow{p}0,\ \ J_2=\int_{A_{2n}}C_nd\mathbf{z}_n\xrightarrow{p}0.$$

For J_1, note that

$$C_n\leq C_{1n}+C_{2n},$$

where

$$C_{1n}=\|\mathbf{z}_n\|^2\left|p\left(\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)\right|\left|\exp\left[-\frac{1}{2}\mathbf{z}_n'\left[\mathbf{I}_q-\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})\right]\mathbf{z}_n\right]-\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|,$$

$$C_{2n}=\|\mathbf{z}_n\|^2\left|p\left(\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)-p\left(\boldsymbol{\vartheta}_n^0\right)\right|\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right).$$

Then, we have

$$0 \leq J_1 \leq J_{11} + J_{12},$$

where

$$J_{11} = \int_{A_{1n}} C_{1n} d\mathbf{z}_n,\ J_{12} = \int_{A_{1n}} C_{2n} d\mathbf{z}_n.$$

Note that since $\delta \leq \delta_1$, from (39), $\left| p\left(\widehat{\boldsymbol{\vartheta}} + \Sigma_n^{1/2}\mathbf{z}_n\right)\right| \leq (1+\eta)\, p\left(\boldsymbol{\vartheta}_n^0\right)$. Hence, we have

$$J_{11} \leq (1+\eta)\, p\left(\boldsymbol{\vartheta}_n^0\right) \int_{A_{1n}} \|\mathbf{z}_n\|^2 \left| \exp\left[-\frac{\mathbf{z}_n' \left[\mathbf{I}_q - \mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\right] \mathbf{z}_n}{2} \right] - \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2} \right) \right| d\mathbf{z}_n.$$

Let $\mathbf{r}_0 = \mathbf{z}_n / \|\mathbf{z}_n\|$ so that $\|\mathbf{r}_0\| = 1$. Hence,

$$\mathbf{r}_0'\mathbf{R}_n\left(\tilde{\boldsymbol{\vartheta}}_1\right)\mathbf{r}_0 = \mathbf{r}_0'\mathbf{r}_0 + \mathbf{r}_0'\boldsymbol{\Sigma}^{1/2} \frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} \boldsymbol{\Sigma}^{1/2}\mathbf{r}_0 = 1 + \mathbf{r}_0'\boldsymbol{\Sigma}^{1/2} \frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} \boldsymbol{\Sigma}^{1/2}\mathbf{r}_0,$$

where $\widetilde{\boldsymbol{\vartheta}}_1$ lies between $\boldsymbol{\vartheta}$ and $\widehat{\boldsymbol{\vartheta}}$. Since $\widehat{\boldsymbol{\vartheta}} \xrightarrow{p} \boldsymbol{\vartheta}_n^0$, with probability approaching one, $\hat{\boldsymbol{\vartheta}} \in N_0(\delta)$ and hence $\tilde{\boldsymbol{\vartheta}}_1 \in N_0(\delta)$.

Following (40), with probability approaching one, when $\boldsymbol{\vartheta} \in N_0(\delta)$, we have

$$\begin{aligned}
& \|\mathbf{z}_n\|^2 \left| \exp\left(-\frac{1}{2}\mathbf{z}_n' \left[\mathbf{I}_q - \mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\right] \mathbf{z}_n \right) - \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2} \right) \right| \\
= & \|\mathbf{z}_n\|^2 \left| \exp\left(\frac{1}{2}\mathbf{z}_n'\mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\mathbf{z}_n \right) - 1 \right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2} \right) \\
\leq & \|\mathbf{z}_n\|^2 \exp\left(\left| \frac{1}{2}\mathbf{z}_n'\mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\mathbf{z}_n \right| \right) \left| \frac{1}{2}\mathbf{z}_n'\mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\mathbf{z}_n \right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2} \right) \\
= & \|\mathbf{z}_n\|^2 \exp\left(\left| \frac{1}{2}\mathbf{z}_n'\mathbf{z}_n \right| \left| \mathbf{r}_0'\mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\mathbf{r}_0 \right| \right) \left| \frac{1}{2}\mathbf{z}_n'\mathbf{z}_n \right| \left| \mathbf{r}_0'\mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\mathbf{r}_0 \right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2} \right) \\
\leq & \frac{\eta}{2} \|\mathbf{z}_n\|^2 \exp\left(\left| \frac{\eta}{2}\mathbf{z}_n'\mathbf{z}_n \right| \right) \left| \mathbf{z}_n'\mathbf{z}_n \right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2} \right) \\
= & \frac{\eta}{2} \|\mathbf{z}_n\|^4 \exp\left(-\frac{(1-\eta)\mathbf{z}_n'\mathbf{z}_n}{2} \right). \qquad (42)
\end{aligned}$$

Let

$$J_{11}^* = \int_{A_{1n}} \|\mathbf{z}_n\|^2 \left| \exp\left[-\frac{1}{2}\mathbf{z}_n' \left[\mathbf{I}_q - \mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\right] \mathbf{z}_n \right] - \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2} \right) \right| d\mathbf{z}_n.$$

Following (42), we have

$$\lim_{n\to\infty} P\left\{ J_{11}^* \leq \frac{\eta}{2} \int_{A_{1n}} \|\mathbf{z}_n\|^4 \exp\left(-\frac{(1-\eta)\,\mathbf{z}_n'\mathbf{z}_n}{2} \right) d\mathbf{z}_n \right\} = 1. \qquad (43)$$

By Lemma 8.2, we have

$$\int_{A_{1n}} \|\mathbf{z}_n\|^4 \exp\left(-\frac{(1-\eta)\,\mathbf{z}_n'\mathbf{z}_n}{2} \right) d\mathbf{z}_n$$

$$
\begin{aligned}
&\le \int_{\mathbb{R}^q} \|\mathbf{z}_n\|^4 \exp\left(-\frac{1-\eta}{2}\mathbf{z}_n'\mathbf{z}_n\right) d\mathbf{z}_n \le \int_{\mathbb{R}^q} \left(\sum_{i=1}^q |\mathbf{z}_{ni}|^2\right)^2 \exp\left(-\frac{1-\eta}{2}\mathbf{z}_n'\mathbf{z}_n\right) d\mathbf{z}_n \\
&\le (2\pi)^{q/2}(1-\eta)^{-q/2} q^4 \int_{\mathbb{R}^q} \left(\max_{i\in\{1,\ldots,q\}} |\mathbf{z}_{ni}|\right)^4 (2\pi)^{-q/2}(1-\eta)^{q/2} \exp\left(-\frac{1-\eta}{2}\mathbf{z}_n'\mathbf{z}_n\right) d\mathbf{z}_n \\
&\le \sqrt{2}e^{5/3}\left(\frac{q+1}{\sqrt{q}}\right) q^4 (2\pi)^{q/2}(1-\eta)^{-q/2}\left[\int_{\mathbb{R}} |t|^8 \sqrt{\frac{1-\eta}{2\pi}} \exp\left(-\frac{1-\eta}{2}t^2\right) dt\right]^{1/2} \\
&= \sqrt{2}e^{5/3}(q+1)\, q^{k+\frac{3}{2}} (2\pi)^{q/2}(1-\eta)^{-q/2}(1-\eta)^{-2}\, 2^2 \left(\frac{\Gamma\left(\frac{9}{2}\right)}{\sqrt{\pi}}\right)^{1/2} \\
&= 2^{\frac{5+q}{2}} e^{5/3}(q+1)\, q^{2+\frac{3}{2}} \sqrt{\Gamma\left(\frac{9}{2}\right)}\pi^{\frac{2q-1}{4}} \left(\frac{1}{1-\eta}\right)^{(4+q)/2} \\
&= 2^{\frac{5+q}{2}} e^{5/3}(q+1)\, q^{2+\frac{3}{2}} \pi^{\frac{2q-1}{4}} \sqrt{\frac{35}{4}\Gamma\left(\frac{5}{2}\right)} \left(\frac{1}{1-\eta}\right)^{(4+q)/2},
\end{aligned}
$$

where $\mathbf{z}_{ni}$ is the ith element of $\mathbf{z}_n$ and the third last equality follows from the fact that

$$
E\{|X-\mu|^\nu\} = \sigma^\nu 2^{\nu/2}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}}, \text{ if } X \sim N(\mu,\sigma^2).
$$

Hence,

$$
\lim_{n\to\infty} P\left(\frac{J_{11}}{C_{J_1}} \le \frac{q^2\eta(1+\eta)\sqrt{35}}{2(1-\eta)^{\frac{q+4}{2}}}\right) = 1, \tag{44}
$$

where

$$
C_{J_1} = e^{5/3} p\left(\boldsymbol{\vartheta}_n^0\right) 2^{\frac{q+3}{2}} \pi^{\frac{2q-1}{4}} (q+1)\, q^{\frac{3}{2}} \sqrt{\Gamma\left(\frac{5}{2}\right)}.
$$

We now deal with J_{12}. From (39) and Lemma 8.2, we have

$$
\begin{aligned}
J_{12} &\le \int_{A_{1n}} \|\mathbf{z}_n\|^2 \left|p\left(\widehat{\boldsymbol{\vartheta}} + \Sigma_n^{1/2}\mathbf{z}_n\right) - p\left(\boldsymbol{\vartheta}_n^0\right)\right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
&\le \eta p\left(\boldsymbol{\vartheta}_n^0\right) \int_{A_{1n}} \|\mathbf{z}_n\|^2 \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
&\le \eta p\left(\boldsymbol{\vartheta}_n^0\right) \int_{\mathbb{R}^q} \|\mathbf{z}_n\|^2 \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
&= \eta p\left(\boldsymbol{\vartheta}_n^0\right) (2\pi)^{q/2} \int_{\mathbb{R}^q} \|\mathbf{z}_n\|^2 (2\pi)^{-q/2} \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
&\le \eta p\left(\boldsymbol{\vartheta}_n^0\right) (2\pi)^{q/2} q^2 \int_{\mathbb{R}^q} \left(\max_i |\mathbf{z}_{ni}|\right)^2 (2\pi)^{-q/2} \exp\left(-\frac{1-\eta}{2}\mathbf{z}_n'\mathbf{z}_n\right) d\mathbf{z}_n \\
&\le \sqrt{2}e^{5/3}\left(\frac{q+1}{\sqrt{q}}\right) \eta p\left(\boldsymbol{\vartheta}_n^0\right) (2\pi)^{q/2} q^2 \left[\int_{\mathbb{R}} |t|^4 (2\pi)^{-1/2} \exp\left(-\frac{t^2}{2}\right) d\mathbf{t}\right]^{1/2} \\
&= \eta\sqrt{2}e^{5/3} p\left(\boldsymbol{\vartheta}_n^0\right) (2\pi)^{q/2} (q+1)\, q^{\frac{3}{2}} 2 \left(\frac{\Gamma\left(\frac{5}{2}\right)}{\sqrt{\pi}}\right)^{1/2}
\end{aligned}
$$

$$= \quad \eta e^{5/3} p\left(\boldsymbol{\vartheta}_n^0\right) 2^{\frac{q+3}{2}} \pi^{\frac{2q-1}{4}} (q+1) q^{\frac{3}{2}} \sqrt{\Gamma\left(\frac{5}{2}\right)} = C_{J_1}\eta.$$

Similarly, we have

$$\lim_{n\to\infty} P\left\{\frac{J_{12}}{C_{J_1}} \leq \eta\right\} = 1. \tag{45}$$

From (44) and (45), we have

$$\lim_{n\to\infty} P\left\{\frac{J_{11}+J_{12}}{C_{J_1}} \leq \eta\left(\frac{q^2(1+\eta)\sqrt{35}}{2(1-\eta)^{\frac{q+4}{2}}}+1\right)\right\} = 1. \tag{46}$$

By the way how η and ε are chosen, from (46), we have

$$\lim_{n\to\infty} P\left\{\frac{J_1}{C_{J_1}} \leq \varepsilon\right\} = 1. \tag{47}$$

Since ε is chosen arbitrarily and $J_1 \geq 0$, we have

$$J_1 \xrightarrow{p} 0.$$

Next we show that

$$J_2 \xrightarrow{p} 0. \tag{48}$$

Using (41), we can write

$$0 \leq J_2 = \int_{A_{2n}} C_n d\mathbf{z}_n \leq J_{21} + J_{22},$$

where

$$J_{21} = \int_{A_{2n}} \|\mathbf{z}_n\|^2 p\left(\widehat{\boldsymbol{\vartheta}} + \Sigma_n^{1/2}\mathbf{z}_n\right) \exp\left[-\frac{1}{2}\mathbf{z}_n'\left[\mathbf{I}_q - \mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})\right]\mathbf{z}_n\right] d\mathbf{z}_n,$$

$$J_{22} = \int_{A_{2n}} \|\mathbf{z}_n\|^2 p\left(\boldsymbol{\vartheta}_n^0\right) \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n.$$

For J_{21}, by (36), we have

$$\begin{aligned} J_{21} &= \int_{A_{2n}} \|\mathbf{z}_n\|^2 p\left(\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) \exp\left[\log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) - \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)\right] d\mathbf{z}_n \\ &= \int_{A_{2n}} \|\mathbf{z}_n\|^2 p\left(\widehat{\boldsymbol{\vartheta}} + \Sigma_n^{1/2}\mathbf{z}_n\right) \exp\left[\log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) - \log p\left(\mathbf{y}|\boldsymbol{\vartheta}_n^0\right)\right] d\mathbf{z}_n \\ &\quad \times \exp\left[\log p\left(\mathbf{y}|\boldsymbol{\vartheta}_n^0\right) - \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)\right]. \end{aligned} \tag{49}$$

According to Lemma 3.1 in Li et al. (2017), if $\mathbf{z}_n \in A_{2n}$, $\log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \Sigma^{1/2}\mathbf{z}_n\right) - \log p\left(\mathbf{y}|\boldsymbol{\vartheta}_n^0\right) < -nK(\delta)$ with probability approaching one. Note that $\exp\left[\log p\left(\mathbf{y}|\boldsymbol{\vartheta}_n^0\right) - \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)\right] \leq 1$. Hence, the integral on the right-hand side of (49) is less than

$$\exp\left[-nK(\delta)\right] \int_{A_{2n}} \|\mathbf{z}_n\|^2 p\left(\widehat{\boldsymbol{\vartheta}} + \Sigma_n^{1/2}\mathbf{z}_n\right) d\mathbf{z}_n,$$

with probability approaching one. Then, we have

$$
\begin{aligned}
& \exp\left[-nK(\delta)\right]\int_{A_{2n}} \|\mathbf{z}_n\|^2 p\left(\widehat{\boldsymbol{\vartheta}}+\Sigma_n^{1/2}\mathbf{z}_n\right) d\mathbf{z}_n \\
= \;& \exp\left[-nK(\delta)\right]\int_{\boldsymbol{\Theta}\setminus\mathbf{N}_0(\delta)} \left\|\boldsymbol{\Sigma}_n^{-1/2}\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\right\|^2 p(\boldsymbol{\vartheta})\left|\boldsymbol{\Sigma}_n\right|^{-1/2} d\boldsymbol{\vartheta} \\
\leq \;& \exp\left[-nK(\delta)\right]\int_{\boldsymbol{\Theta}\setminus\mathbf{N}_0(\delta)} \left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^2\left\|\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right\|^2 p(\boldsymbol{\vartheta})\left|\boldsymbol{\Sigma}_n\right|^{-1/2} d\boldsymbol{\vartheta} \\
\leq \;& \exp\left[-nK(\delta)\right]\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^2\left|\boldsymbol{\Sigma}_n\right|^{-1/2}\int \left\|\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right\|^2 p(\boldsymbol{\vartheta})\, d\boldsymbol{\vartheta} \\
\leq \;& \exp\left[-nK(\delta)\right]\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^2\left|\boldsymbol{\Sigma}_n\right|^{-1/2}\int \left(\|\boldsymbol{\vartheta}\|+\left\|\widehat{\boldsymbol{\vartheta}}\right\|\right)^2 p(\boldsymbol{\vartheta})\, d\boldsymbol{\vartheta} \\
\leq \;& \exp\left[-nK(\delta)\right]\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^2\left|\boldsymbol{\Sigma}_n\right|^{-1/2}\sum_{s=0}^{2}\binom{2}{s}\left\|\widehat{\boldsymbol{\vartheta}}\right\|^{2-s}\int \|\boldsymbol{\vartheta}\|^s p(\boldsymbol{\vartheta})\, d\boldsymbol{\vartheta}.
\end{aligned}
$$

Note that

$$
\exp\left[-nK(\delta)\right]\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^2\left|\boldsymbol{\Sigma}_n\right|^{-1/2} = \exp\left[-nK(\delta)\right] n^{3/2}\left\|-\bar{\mathbf{H}}_n^{-1/2}\right\|^2\left|\bar{\mathbf{H}}_n\right|^{-1/2} \xrightarrow{p} 0.
$$

Furthermore, $\int \|\boldsymbol{\vartheta}\|^2 p(\boldsymbol{\vartheta})\, d\boldsymbol{\vartheta} < \infty$ and $\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_n^0 \xrightarrow{p} 0$ by Assumptions 1-8. Then, we have

$$
J_{21} \xrightarrow{p} 0. \tag{50}
$$

For J_{22}, we can show that

$$
\begin{aligned}
J_{22} \;=\;& p\left(\boldsymbol{\vartheta}_n^p\right)\int_{A_{2n}} \|\mathbf{z}_n\|^2 \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
\leq\;& p\left(\boldsymbol{\vartheta}_n^0\right)\int_{\|\mathbf{z}_n\|>\sqrt{n\lambda_n}\delta} \|\mathbf{z}_n\|^2 \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
\leq\;& (2\pi)^{q/2} p\left(\boldsymbol{\vartheta}_n^0\right)\int_{\cap_{i=1}^q\left\{|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right\}} \|\mathbf{z}_n\|^2 (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
\leq\;& (2\pi)^{q/2} p\left(\boldsymbol{\vartheta}_n^0\right) q^2\int_{\cap_{i=1}^q\left\{|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right\}} \left(\max_{i\in\{1,\dots,q\}}|\mathbf{z}_{ni}|\right)^2 (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
=\;& (2\pi)^{q/2} p\left(\boldsymbol{\vartheta}_n^0\right) q^2\int_{R^q} \left(\max_{i\in\{1,\dots,q\}}|\mathbf{z}_{ni}|\right)^2 1\left(\cap_{i=1}^q\left\{|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right\}\right)(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
=\;& (2\pi)^{q/2} p\left(\boldsymbol{\vartheta}_n^0\right) q^2\int_{R^q} \left(\max_{i\in\{1,\dots,q\}}|\mathbf{z}_{ni}|\right)^2 \prod_{i=1}^q 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
\leq\;& (2\pi)^{q/2} p\left(\boldsymbol{\vartheta}_n^0\right) q^2\left[\int_{R^q} \left(\max_{i\in\{1,\dots,q\}}|\mathbf{z}_{ni}|\right)^4 (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n\right]^{1/2} \\
&\times\left\{\int_{R^q}\left[\prod_{i=1}^q 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)\right]^2 (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n\right\}^{1/2}
\end{aligned}
$$

where λ_n is the smallest eigenvalue of $-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)$.

From (33), we have

$$\int_{R^q}\left(\max_{i\in\{1,\ldots,q\}}|\mathbf{z}_{ni}|\right)^4(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)d\mathbf{z}_n<\infty. \tag{51}$$

It can be shown that

$$\begin{aligned}
&\int_{R^q}\left[\prod\nolimits_{i=1}^q 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)\right]^2(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)d\mathbf{z}_n\\
=&\int_{R^q}\prod\nolimits_{i=1}^q 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)d\mathbf{z}_n\\
=&\prod\nolimits_{i=1}^q\left[\int_R 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)(2\pi)^{-1/2}\exp\left(-\frac{\mathbf{z}_{ni}^2}{2}\right)d\mathbf{z}_{ni}\right]\\
=&\prod\nolimits_{i=1}^q\left[\int_{|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta}(2\pi)^{-1/2}\exp\left(-\frac{\mathbf{z}_{ni}^2}{2}\right)d\mathbf{z}_{ni}\right]\\
\le&\left(\sqrt{q+1}\frac{\exp\left(-n\lambda_n\delta^2/2(q+1)\right)}{\sqrt{n\lambda_n 2\pi}\delta}\right)^q\\
=&\ 2^{-\frac{q}{2}}(q+1)^{\frac{q}{2}}\left(\frac{1}{\sqrt{\pi}\delta}\right)^q(n\lambda_n)^{-\frac{q}{2}}\exp\left(-\frac{n\lambda_n q\delta^2}{q+1}\right)\xrightarrow{p}0,
\end{aligned} \tag{52}$$

where the last inequality is due to

$$\int_x^\infty\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\le\int_x^\infty\frac{1}{\sqrt{2\pi}}\frac{t}{x}e^{-\frac{t^2}{2}}dt=\frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}.$$

From (51) and (52), we have

$$J_{22}\xrightarrow{p}0. \tag{53}$$

From (50) and (53), we get (48). And from (47) and (48), we have

$$J\xrightarrow{p}0.$$

■

8.1 Proof of Lemma 3.1

To prove $E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)'|\mathbf{y}\right]-\boldsymbol{\Sigma}_n=o_p\left(\frac{1}{n}\right)$, it is sufficient to show that, for any $\varepsilon>0$,

$$\lim_{n\to\infty}P\left(\int_{A_n}\left\|\mathbf{z}_n\mathbf{z}_n'\right\|\left|p\left(\mathbf{z}_n|\mathbf{y}\right)-(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|d\mathbf{z}_n>\varepsilon\right)=0, \tag{54}$$

where $\|A_{n\times n}\| = \sup_{\{x:\|x\|=1,x\in R^n\}} \|Ax\|$ is known as the matrix norm. By (54),

$$\int_{A_n} \left\|\mathbf{z}_n\mathbf{z}_n'\right\| \left| p\left(\mathbf{z}_n|\mathbf{y}\right) - (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right| d\mathbf{z}_n \xrightarrow{p} 0.$$

Thus, $\left|\int_{A_n} \mathbf{z}_n\mathbf{z}_n' \left[p\left(\mathbf{z}_n|\mathbf{y}\right) - (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right] d\mathbf{z}_n\right| \xrightarrow{p} 0$, which implies that

$$\int_{A_n} \mathbf{z}_n\mathbf{z}_n' p\left(\mathbf{z}_n|\mathbf{y}\right) d\mathbf{z}_n - \int_{A_n} \mathbf{z}_n\mathbf{z}_n' (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \xrightarrow{p} 0. \tag{55}$$

From (35) we get

$$\begin{aligned}
\int_{A_n} \mathbf{z}_n\mathbf{z}_n' p\left(\mathbf{z}_n|\mathbf{y}\right) d\mathbf{z}_n &= \int_{A_n} \mathbf{z}_n\mathbf{z}_n' p\left(\mathbf{y}\right)^{-1} \left|\boldsymbol{\Sigma}_n\right|^{1/2} p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) p\left(\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) d\mathbf{z}_n \\
&= \int \boldsymbol{\Sigma}_n^{-1/2}\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' \boldsymbol{\Sigma}_n^{-1/2} p\left(\mathbf{y}\right)^{-1} \left|\boldsymbol{\Sigma}_n\right|^{1/2} p\left(\mathbf{y}|\boldsymbol{\vartheta}\right) p\left(\boldsymbol{\vartheta}\right) \left|\boldsymbol{\Sigma}_n\right|^{-1/2} d\boldsymbol{\vartheta} \\
&= \int \boldsymbol{\Sigma}_n^{-1/2}\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' \boldsymbol{\Sigma}_n^{-1/2} p\left(\boldsymbol{\vartheta}|\mathbf{y}\right) d\boldsymbol{\vartheta} \\
&= \boldsymbol{\Sigma}_n^{-1/2} E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' |\mathbf{y}\right] \boldsymbol{\Sigma}_n^{-1/2},
\end{aligned} \tag{56}$$

by the changing-of-variable technique. From (55) and (56), by Assumptions 1-9, we have

$$\boldsymbol{\Sigma}_n^{-1/2} E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' |\mathbf{y}\right] \boldsymbol{\Sigma}_n^{-1/2} - \mathbf{I}_q \xrightarrow{p} 0_{q\times q}.$$

Hence, we have

$$E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' |\mathbf{y}\right] - \boldsymbol{\Sigma}_n = E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' |\mathbf{y}\right] + \left(\frac{\partial^2 \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right)^{-1} = o_p\left(n^{-1}\right).$$

Since $\|\mathbf{z}_n\mathbf{z}_n'\| \leq \|\mathbf{z}_n\|^2$, Equation (34) holds so that (54) also holds. Similarly, it can be derived that $\sqrt{n}\left(\bar{\boldsymbol{\vartheta}}-\widehat{\boldsymbol{\vartheta}}\right) \xrightarrow{p} 0$. This completes the proof.

8.2 Proof of Theorem 3.1

According to Lemma 3.1, we have

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

Hence, based on Lemma 3.1, we have

$$\mathbf{V}\left(\bar{\boldsymbol{\vartheta}}\right) = E\left[(\boldsymbol{\vartheta}-\bar{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta}-\bar{\boldsymbol{\vartheta}})'|\mathbf{y}\right]$$

$$
\begin{aligned}
&= E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}+\widehat{\boldsymbol{\vartheta}}-\bar{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}+\widehat{\boldsymbol{\vartheta}}-\bar{\boldsymbol{\vartheta}}\right)'|\mathbf{y}\right] \\
&= E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)'|\mathbf{y}\right]+2E\left[\left(\widehat{\boldsymbol{\vartheta}}-\bar{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)'|\mathbf{y}\right]+E\left[\left(\widehat{\boldsymbol{\vartheta}}-\bar{\boldsymbol{\vartheta}}\right)\left(\widehat{\boldsymbol{\vartheta}}-\bar{\boldsymbol{\vartheta}}\right)'|\mathbf{y}\right] \\
&= \mathbf{V}\left(\widehat{\boldsymbol{\vartheta}}\right)-E\left[\left(\widehat{\boldsymbol{\vartheta}}-\bar{\boldsymbol{\vartheta}}\right)\left(\widehat{\boldsymbol{\vartheta}}-\bar{\boldsymbol{\vartheta}}\right)'|\mathbf{y}\right] \\
&= \mathbf{V}\left(\widehat{\boldsymbol{\vartheta}}\right)+o_p(n^{-1/2})o_p(n^{-1/2}) \\
&= -\frac{1}{n}\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\vartheta}})+o_p\left(n^{-1}\right)=O_p(n^{-1}). \qquad (57)
\end{aligned}
$$

According to the classical asymptotic theory for MLE (White, 1996), $\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0=O_p(n^{-1/2})$ under H_0. Thus,

$$
\begin{aligned}
&(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[-n^{-1}\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}}_n)+o_p(n^{-1})\right]^{-1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
&= \sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}})+o_p(1)\right]^{-1}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
&= \sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}})\right]^{-1}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(1)\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
&= \sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}})\right]^{-1}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(1)\sqrt{n}O_p(n^{-1/2})\sqrt{n}O_p(n^{-1/2}) \\
&= \sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}})\right]^{-1}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(1) \\
&= \mathbf{Wald}+o_p(1). \qquad (58)
\end{aligned}
$$

Under H_0, we can further prove that

$$
\begin{aligned}
\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0) &= q_\theta+\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}+\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}+\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) \\
&= q_\theta+\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) \\
&\quad +2\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}\right)'\left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}\right)'\left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1}\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}\right) \\
&= q_\theta+\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+2o_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(n\right)O_p\left(\frac{1}{\sqrt{n}}\right) \\
&\quad +o_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(n\right)o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= q_\theta+\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+o_p(1). \qquad (59)
\end{aligned}
$$

Thus, under H_0, from (58) and (59), we have

$$\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)-q_\theta=\mathbf{Wald}+o_p(1).$$

Since $\mathbf{T}(\mathbf{y},\boldsymbol{\theta}_0)-q_\theta=\mathbf{W}$, we have under H_0,

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

8.3 Proof of Theorem 3.3

Note that

$$
\begin{aligned}
& \mathbf{T}(\mathbf{y}, \boldsymbol{r}) = \int \Delta\mathcal{L}(H_0, \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}|\mathbf{y}) d\boldsymbol{\vartheta} \\
= & \int (R(\boldsymbol{\theta}) - \boldsymbol{r})' \left[\frac{\partial R(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} \mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}}) \frac{\partial R(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right]^{-1} (R(\boldsymbol{\theta}) - \mathbf{r}) p(\boldsymbol{\vartheta}|\mathbf{y}) d\boldsymbol{\vartheta} \\
= & \mathbf{tr}\left\{\int [R(\boldsymbol{\theta}) - \mathbf{r}][R(\boldsymbol{\theta}) - \mathbf{r}]' p(\boldsymbol{\vartheta}|\mathbf{y}) d\boldsymbol{\vartheta} \left[\frac{\partial R(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} \mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}}) \frac{\partial R(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right]^{-1}\right\} \\
= & \mathbf{tr}\left\{E\left[n(R(\boldsymbol{\theta}) - \mathbf{r})(R(\boldsymbol{\theta}) - \mathbf{r})' \middle| \mathbf{y}, H_1\right] \left[\frac{\partial R(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} n\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}}) \frac{\partial R(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right]^{-1}\right\}.
\end{aligned}
$$

We have

$$
\begin{aligned}
& E\left[n(R(\boldsymbol{\theta}) - \boldsymbol{r})(R(\boldsymbol{\theta}) - \mathbf{r})' \middle| \mathbf{y}, H_1\right] \\
= & E\left[n\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right) + R\left(\widehat{\boldsymbol{\theta}}\right) - \mathbf{r}\right)\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right) + R\left(\widehat{\boldsymbol{\theta}}\right) - \mathbf{r}\right)' \middle| \mathbf{y}, H_1\right] \\
= & E\left[n\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right)' \middle| \mathbf{y}, H_1\right] \\
& + 2E\left[n\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R\left(\widehat{\boldsymbol{\theta}}\right) - \boldsymbol{r}\right)' \middle| \mathbf{y}, H_1\right] \\
& + n\left(R\left(\widehat{\boldsymbol{\theta}}\right) - \mathbf{r}\right)\left(R\left(\widehat{\boldsymbol{\theta}}\right) - \mathbf{r}\right)'. \qquad (60)
\end{aligned}
$$

Apply the Taylor expansion to $R(\boldsymbol{\theta})$, we have

$$
\sqrt{n}\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right) = \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}} \sqrt{n}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) + \left[\sqrt{n}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)' \otimes \mathbf{I}_m\right] \frac{\partial R^2\left(\widetilde{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right),
$$

where $\widetilde{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$. Note that $\frac{\partial^2 R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$ is continuous and $\boldsymbol{\Theta}$ is compact. Thus, we have

$$
\left\|\frac{\partial^2 R\left(\widetilde{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right\| \leq M', \qquad (61)
$$

for some $0 < M' < \infty$. Furthermore, by the BvM theorem, $\sqrt{n}\left(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}\right) = O_p(1)$. Hence, from (61), we can further derive that

$$
\left[\sqrt{n}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)' \otimes \mathbf{I}_m\right] \frac{\partial R^2\left(\widetilde{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) = O_p(1)O(1)\frac{1}{\sqrt{n}} = o_p(1). \qquad (62)
$$

By Lemma 3.1, we have $\int \sqrt{n}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right) p\left(\boldsymbol{\vartheta}|\mathbf{y}\right) \mathrm{d}\boldsymbol{\vartheta} = \sqrt{n}(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}) = o_p\left(1\right)$. By the Delta method and the consistency of MLE, we have$\sqrt{n}\left(R\left(\widehat{\boldsymbol{\theta}}\right)-R\left(\boldsymbol{\theta}_0\right)\right) = O_p\left(1\right)$. Consequently, the second term of (60) is

$$
\begin{aligned}
& E\left[n\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R\left(\widehat{\boldsymbol{\theta}}\right)-\boldsymbol{r}\right)'\middle|\mathbf{y},H_1\right] \\
= & \int \sqrt{n}\left[R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right] p\left(\boldsymbol{\vartheta}|\mathbf{y}\right) d\boldsymbol{\vartheta} \times \sqrt{n}\left(R\left(\widehat{\boldsymbol{\theta}}\right)-R\left(\boldsymbol{\theta}_0\right)\right)' \\
= & \int \sqrt{n}\left[R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right] p\left(\boldsymbol{\vartheta}|\mathbf{y}\right) d\boldsymbol{\vartheta} \times \sqrt{n}\left(R\left(\widehat{\boldsymbol{\theta}}\right)-R\left(\boldsymbol{\theta}_0\right)\right)' \\
= & \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}} \int \sqrt{n}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right) p\left(\boldsymbol{\vartheta}|\mathbf{y}\right) d\boldsymbol{\vartheta}\sqrt{n}\left(R\left(\widehat{\boldsymbol{\theta}}\right)-R\left(\boldsymbol{\theta}_0\right)\right)+o_p\left(1\right) \\
= & O_p(1)o_p\left(1\right)O_p(1)+o_p\left(1\right)=o_p(1).
\end{aligned}
$$

For the first term of (60), after integrating out the nuisance parameters, by the Taylor expansion, we have,

$$
\begin{aligned}
& E\left[n\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)'\middle|\mathbf{y},H_1\right] \\
= & \int_{\boldsymbol{\Theta}_{\boldsymbol{\theta}}} n\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)' p\left(\boldsymbol{\theta}|\mathbf{y}\right) d\boldsymbol{\theta} \\
= & \frac{\partial R\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}'} \int_{\boldsymbol{\Theta}_{\boldsymbol{\theta}}} n\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)' p\left(\boldsymbol{\theta}|\mathbf{y}\right) d\boldsymbol{\theta}\frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}+o_p\left(1\right).
\end{aligned}
$$

By Lemma 3.1, we have

$$
\int_{\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)' p\left(\boldsymbol{\theta}|\mathbf{y}\right) d\boldsymbol{\theta} = -\frac{1}{n}\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)+o_p\left(n^{-1}\right).
$$

Therefore,

$$
E\left[n\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)'\middle|\mathbf{y},H_1\right] = \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]\frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}+o_p\left(1\right).
$$

Since $V_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right) = -\frac{1}{n}\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)+o_p\left(n^{-1}\right)$, $\bar{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}+o_p\left(n^{-1/2}\right) = \boldsymbol{\theta}_0+O_p\left(n^{-1/2}\right)$, by (62), we have

$$
\begin{aligned}
& \mathbf{tr}\left\{E\left[n\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R\left(\boldsymbol{\theta}\right)-R\left(\widehat{\boldsymbol{\theta}}\right)\right)'\middle|\mathbf{y},H_1\right]\left[\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}n\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right]^{-1}\right\} \\
= & \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]\frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\left[\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}n\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right]^{-1}+o_p\left(1\right) \\
= & \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]\frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\left[\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}n\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\frac{\partial R\left(\bar{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right]^{-1}+o_p\left(1\right)\stackrel{p}{\to} m.
\end{aligned}
$$

Finally, the third term of (60) can be expressed as

$$\begin{aligned}
& \mathbf{tr}\left\{n\left(R\left(\widehat{\boldsymbol{\theta}}\right)-\mathbf{r}\right)\left(R\left(\widehat{\boldsymbol{\theta}}\right)-\mathbf{r}\right)'\left[\frac{\partial R\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} n\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\frac{\partial R\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right]^{-1}\right\}+o_p\left(1\right) \\
= & \left[R\left(\widehat{\boldsymbol{\theta}}\right)-\boldsymbol{r}\right]'\left\{\frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'}\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]\frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}}\right\}^{-1}\left[R\left(\widehat{\boldsymbol{\theta}}\right)-\mathbf{r}\right]+o_p\left(1\right) \\
= & \mathbf{Wald}+o_p\left(1\right).
\end{aligned}$$

Therefore, under H_0, we have

$$\boldsymbol{T}\left(\mathbf{y},\mathbf{r}\right)-m=\mathbf{Wald}+o_p\left(1\right)\xrightarrow{d}\chi^2\left(m\right).$$

8.4 Proof of Theorem 3.4

Let $\left\{\boldsymbol{\vartheta}^{[j]}, j=1,\cdots,J\right\}$ be effective random draws from $p\left(\boldsymbol{\vartheta}|\mathbf{y}\right)$. And let

$$\bar{\mathbf{V}}_2=\frac{1}{J}\sum_{j=1}^{J}\left(\boldsymbol{\theta}^{[j]}-\bar{\mathbf{v}}_1\right)\left(\boldsymbol{\theta}^{[j]}-\bar{\mathbf{v}}_1\right)'=\frac{1}{J}\sum_{j=1}^{J}\mathbf{V}_2^{[j]}=\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right).$$

Hence, $\widehat{\mathbf{T}}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ in (13) can be rewritten as

$$\begin{aligned}
\widehat{\mathbf{T}}\left(\mathbf{y},\boldsymbol{\theta}_0\right) &= \mathbf{tr}\left[\left(\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\right)^{-1}\bar{\mathbf{V}}_{\theta}(\boldsymbol{\theta}_0)\right] \\
&= \mathbf{tr}\left\{\left[\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\right]^{-1}\left[\frac{1}{J}\sum_{j=1}^{J}\left(\boldsymbol{\theta}^{[j]}-\boldsymbol{\theta}_0\right)\left(\boldsymbol{\theta}^{[j]}-\boldsymbol{\theta}_0\right)\right]\right\} \\
&= \mathbf{tr}\left\{\left[\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\right]^{-1}\left[\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)+\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)'\right]\right\} \\
&= q_\theta+\mathbf{tr}\left[\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)'\bar{\mathbf{V}}_2^{-1}\right],
\end{aligned}$$

which is a consistent estimator of $\mathbf{T}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$.

Following the notations of Magnus and Neudecker (2002) about matrix derivatives, let

$$\mathbf{v}_2^{(j)}=vech\left(\mathbf{V}_2^{[j]}\right),\ \mathbf{v}_1^{[j]}=\boldsymbol{\theta}^{[j]},\bar{\mathbf{v}}_2=vech\left(\bar{\mathbf{V}}_2\right),\ \bar{\mathbf{v}}=\left(\bar{\mathbf{v}}_1',\bar{\mathbf{v}}_2'\right)'.$$

Note that the dimension of $\bar{\mathbf{v}}_2$ is $q^*\times 1$ where $q^*=q_\theta\left(q_\theta+1\right)/2$. Hence, we have

$$\begin{aligned}
\frac{\partial\widehat{\mathbf{T}}\left(\mathbf{y},\boldsymbol{\theta}_0\right)}{\partial\bar{\mathbf{v}}}=&vec\left(\mathbf{I}_{q_\theta}\right)'\left\{\left[\left(\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)'\bar{\mathbf{V}}_2^{-1}\right)'\otimes\mathbf{I}_{q_\theta}\right]\frac{\partial\bar{\mathbf{v}}_1}{\partial\bar{\mathbf{v}}}+\left[\bar{\mathbf{V}}_2^{-1}\otimes\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)\right]\frac{\partial\bar{\mathbf{v}}_1'}{\partial\bar{\mathbf{v}}}\right. \\
&\left.-\left[\mathbf{I}_{q_\theta}\otimes\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)'\right]\left(\bar{\mathbf{V}}_2^{-1}\otimes\bar{\mathbf{V}}_2^{-1}\right)\frac{\partial vec\left(\bar{\mathbf{V}}_2\right)}{\partial\bar{\mathbf{v}}}\right\}
\end{aligned}$$

$$=vec\left(\mathbf{I}_{q\theta}\right)'\left[\left(\left(\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)'\bar{\mathbf{V}}_2^{-1}\right)'\otimes\mathbf{I}_{q\theta}+\bar{\mathbf{V}}_2^{-1}\otimes\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)\right)\frac{\partial\bar{\mathbf{v}}_1}{\partial\bar{\mathbf{v}}}\right.$$
$$\left.-\left[\mathbf{I}_{q\theta}\otimes\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)\left(\bar{\mathbf{v}}_1-\boldsymbol{\theta}_0\right)'\right]\left(\bar{\mathbf{V}}_2^{-1}\otimes\bar{\mathbf{V}}_2^{-1}\right)\frac{\partial\bar{\mathbf{V}}_2}{\partial\bar{\mathbf{v}}}\right],$$

where

$$\frac{\partial\bar{\mathbf{v}}_1}{\partial\bar{\mathbf{v}}}=\frac{\partial\bar{\mathbf{v}}_1'}{\partial\bar{\mathbf{v}}}=\left[\mathbf{I}_{q_\theta},0_{q_\theta\times q^*}\right],\ \frac{\partial\bar{\mathbf{V}}_2}{\partial\bar{\mathbf{v}}}=\left[0_{q_\theta^2\times q_\theta},\left(\frac{\partial vec\left(\bar{\mathbf{V}}_2\right)}{\partial\bar{\mathbf{v}}_2}\right)_{q_\theta^2\times q^*}\right].$$

By the Delta method,

$$Var\left(\widehat{\mathbf{T}}\left(\mathbf{y},\boldsymbol{\theta}_0\right)\right)=\frac{\partial\widehat{\mathbf{T}}\left(\mathbf{y},\boldsymbol{\theta}_0\right)}{\partial\bar{\mathbf{v}}}Var\left(\bar{\mathbf{v}}\right)\left(\frac{\partial\widehat{\mathbf{T}}\left(\mathbf{y},\boldsymbol{\theta}_0\right)}{\partial\bar{\mathbf{v}}}\right)'.$$

Similarly, the expression of the NSE for $\widehat{\mathbf{T}}\left(\mathbf{y},\mathbf{r}\right)$ can be obtained as in (12). By the Delta method,

$$Var\left(\widehat{\mathbf{T}}\left(\mathbf{y},\mathbf{r}\right)\right)=\frac{\partial\widehat{\mathbf{T}}\left(\mathbf{y},\mathbf{r}\right)}{\partial\bar{\mathbf{v}}}Var\left(\bar{\mathbf{v}}\right)\frac{\partial\widehat{\mathbf{T}}\left(\mathbf{y},\mathbf{r}\right)}{\partial\bar{\mathbf{v}}'}.$$

8.5 Proof of Theorem 4.1

Under H_0, we have

$$\begin{aligned}
\mathbf{T}_S\left(\mathbf{y},\boldsymbol{\theta}_0\right) &= q_\theta+\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}+\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'n\left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}+\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\\
&= q_\theta+\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'n\left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\\
&\quad +2\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}\right)'n\left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}\right)'n\left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}\right)\\
&= q_\theta+\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'n\left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\\
&\quad +2\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}\right)'n\left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+o_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(n\right)o_p\left(\frac{1}{\sqrt{n}}\right)\\
&= q_\theta+\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)'n\left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+o_p(1),
\end{aligned}$$

since

$$\left(\bar{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}\right)'n\left[\bar{\boldsymbol{\Sigma}}_{S,\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\boldsymbol{\theta}_n^0-\boldsymbol{\theta}_0\right)=o_p\left(1\right).$$

Therefore, under H_0, we have

$$\mathbf{T}_S\left(\mathbf{y},\boldsymbol{\theta}_0\right)-q_\theta=\mathbf{W}_S=\mathbf{Wald}_S+o_p(1)\xrightarrow{d}\chi^2(q).$$

This completes the proof. Theorem 4.2 can be proved similarly.

References

Aït-Sahalia, Y., Fan, J., Laeven, R. J. A., Wang, C.D., and Yang, X. (2017). Estimation of the continuous and discontinuous leverage effects. *Journal of American Statistical Association*, 112, 1744-1758.

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, 59(3), 817-858.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis.* 2nd edition. Springer-Verlag.

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

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

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

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

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

De Jong, R. M., and Davidson, J. (2000). Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica*, 68(2), 407-423.

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

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

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

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics.* (Vol. 537). John Wiley and Sons.

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

Gribkova, N. (1995). Bounds for absolute moments of order statistics. *Exploring Stochastic Laws: Festschrift in honor of the 70th birthday of Acad. V.S. Korolyuk*, 129-134. VSP, Utrecht.

Harvey, A. C., and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, 14(4), 429-434.

Imai, S., Jain, N., and Ching, A. (2009). Bayesian estimation of dynamic discrete choice models. *Econometrica*, 77, 1865-1899.

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

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

Li, M. (2006). High school completion and future youth unemployment: New evidence from high school and beyond. *Journal of Applied Econometrics*, 21, 23-53.

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

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

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

Li, Y., Wang, N., and Yu, J. (2019). Improved marginal likelihood estimation via power posteriors and importance sampling. Working Paper, Singapore Management University.

Li, Y., Yu, J., and Zeng, T. (2017). Deviance information criterion for Bayesian model selection: Justification and variation. Working Paper, Singapore Management University.

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.

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

McCulloch, R., and Rossi, P. (1992). Bayes factors for non-linear hypotheses and likelihood distributions. *Biometrika*, 79(4), 663-676.

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

Müller, U. K. (2013). Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica*, 81(5), 1805-1849.

Newey, W. K., and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777-787.

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

Poirier, D. J. (1995). Intermediate statistics and econometrics: A comparative approach. MIT Press, Cambridge, MA.

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

Schervish, M. J. (2012). *Theory of Statistics.* Springer Science & Business Media.

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

Wakefield, J. (2013). *Bayesian and Frequentist Regression Methods.* Springer Science & Business Media.

White, H. (1987). Specification testing in dynamic models. *Advances in Econometrics*, Fifth World Congress of Econometric Society. 1, 1-58.

White, H. (1996). *Estimation, inference and specification analysis.* Cambridge University Press.

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

Zhou, Q. M., Song, P. X. K., and Thompson, M. E. (2012). Information ratio test for model misspecification in quasi-likelihood inference. *Journal of the American Statistical Association*, 107(497), 205-213.