9-2022

Posterior-based Wald-type statistic for hypothesis testing

Xiaobin LIU

Yong LI

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

Tao ZENG

Published in Journal of Econometrics, 2022, 230 (1), 83-113. DOI: 10.1016/j.jeconom.2021.11.003
Submitted version

A Posterior-Based Wald-Type Statistic for Hypothesis Testing*

Yong Li
Renmin University

Xiaobin Liu, Tao Zeng
Zhejiang University

Jun Yu
Singapore Management University

May 11, 2018

Abstract

A new Wald-type statistic is proposed for hypothesis testing based on Bayesian posterior distributions. The new statistic can be explained as a posterior version of Wald test and have several nice properties. First, it is well-defined under improper prior distributions. Second, it avoids Jeffreys-Lindley'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 the parameters of interest. Fifth and perhaps most importantly, when a random sample from the posterior distribution (such as an 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 the random sample. The finite sample performance of the statistic 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.

*Li gratefully acknowledges the financial support of the Chinese Natural Science fund (No.71773130) and the hospitality during his research visits to Singapore Management University. 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. All authors acknowledge Peter Phillips for helpful comments and financial support from Lee Foundation via the Lee Kong Chian Professorship. Yong Li, Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing, 1000872, China. Xiaobin Liu and Tao Zeng, School of Economics, Zhejiang University, Zhejiang, 310058, China. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yujun@smu.edu.sg. Co-authors for correspondence: Xiaobin Liu and Tao Zeng.

1 Introduction

This paper develops an approach to test a point null hypothesis based on the Bayesian posterior distribution. The statistic 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 the parameter(s) of interest (denoted $\boldsymbol{\theta}$). Typically one kind of 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 or 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 very hard to calculate numerically, making the MLE nearly impossible to obtain. Not surprisingly, Bayesian MCMC methods have emerged as the leading estimation tool to deal with non-linear and non-Gaussian state space models. There are many other examples in economics where the classical extremum estimators are subject to the curse of dimensionality in computation and some numerical problems. To circumvents this problem, Chernozhukov and Hong (2003) introduced a class of quasi-Bayesian methods that allow users to employ MCMC to simulate a random sequence of draws such that the marginal distribution of the sequence is the same as the quasi-posterior distribution of parameters.

The central question we ask in this paper is how to test a point null hypothesis with the posterior distribution of parameters being 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 a random sample (such as MCMC sample).

Broadly speaking, there are three posterior-based methods available in the literature for hypothesis testing. The first one is the Bayes factor (BF) which compares the posterior odds of the two competing theories 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'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). Third, the calculation of BFs generally involves evaluation of marginal likelihood. In many cases, 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 highly informative prior which is data-dependent. To make it 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 (whether it is improper or vague) to generate a proper informative prior which is subsequently used 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, one can use the methods of Chib (1995) and Chib and Jeliazkov (2001) to compute BFs.

The second posterior-based method is to use credible intervals for point identified parameters and credible sets for partially identified parameters. This line of approaches has drawn a great deal of attentions among econometricians and statisticians in recent years; see Chernozhukov and Hong (2003), Moon and Schorfheide (2012), Norets and Tang (2013), Kline and Tamer (2016), Liao and Simoni (2015), Chen, et al (2016). Except Chernozhukov and Hong (2003), all the other studies focus in developing credible sets in partially identified models. Most of these studies justify credible sets using large-sample theory under repeated sampling.

The third method is based on the statistical decision theory. The idea begins with Bernardo and Rueda (2002, BR hereafter) where they demonstrated that the BF can be regarded as a decision problem with a simple zero-one loss function when it is used for point hypothesis testing. It is this zero-one loss that leads to Jeffreys-Lindley's paradox. BR further suggested using the continuous Kullback-Leibler (KL) divergence function as the loss functions to replace the zero-one loss. Subsequent extensions include Li and Yu (2012), Li, Zeng and Yu (2014) and Li, Liu and Yu (2015, LLY hereafter) where different continuous loss functions or net loss functions were used. The justification of these extensions is made by large-sample theory under repeated sampling.

In this paper, following the third line of approach, we propose a Wald-type statistic for hypothesis testing based on posterior distributions. The new statistic is well-defined under improper prior distributions and avoids Jeffreys-Lindley's paradox. It is asymptotically equivalent to the Wald statistic under the null hypothesis, and hence, follows a χ^2 distribution asymptotically. It is a by-product of posterior simulation, requiring almost no coding effort and little 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. Section 4 explains how to implement the proposed test for an important class of models – latent variable models – where posterior analysis is routinely used. Section 5 investigates 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. Appendix collects the proof of theoretical results.

2 Hypothesis Testing based on Statistical Decision

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 $\boldsymbol{\vartheta} := (\boldsymbol{\theta}', \boldsymbol{\psi}')' \in \Theta$. We are concerned with testing a point null hypothesis which may arise from the prediction of a particular theory. Let $\boldsymbol{\theta} \in \Theta_\theta$ denote a vector of q_θ-dimensional parameters of interest and $\boldsymbol{\psi} \in \Theta_\psi$ a vector of q_ψ-dimensional nuisance parameters, where $\Theta = \Theta_\theta \times \Theta_\psi$. 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 the expected posterior loss of rejecting H_0, we reject H_0. That is, H_0 is rejected if

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

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, $\triangle\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi}) := \mathcal{L}(d_0, \boldsymbol{\theta}, \boldsymbol{\psi}) - \mathcal{L}(d_1, \boldsymbol{\theta}, \boldsymbol{\psi})$ is the net loss function.

BR showed that when the equal prior $p(\boldsymbol{\theta} = \boldsymbol{\theta}_0) = p(\boldsymbol{\theta} \neq \boldsymbol{\theta}_0) = \frac{1}{2}$ is used, $c = 0$, and the net loss function is taken as

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

then $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) > 0$ is equivalent to the following decision rule based on the BF: 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. First, it is not well-defined under improper priors. Second, it leads to Jeffreys-Lindley's paradox when a very vague prior is used. Third, BF_{10} requires evaluating the two marginal likelihood functions, $p(\mathbf{y}|H_i), i = 0, 1$. Clearly, this involves marginalizations over $\boldsymbol{\psi}$ and over $\boldsymbol{\vartheta}$.

Fourth, if $\boldsymbol{\vartheta}$ is high-dimensional so that the integration is a high-dimensional problem, calculating $p(\mathbf{y}|H_i), i = 0, 1$ will be difficult numerically although there have been several interesting methods proposed in the literature to compute the BF from MCMC output; see, for example, Chib (1995), and Chib and Jeliazkov (2001).

In the statistical decision framework, several statistics have been proposed for testing a point null hypothesis. Poirier (1997) developed a loss function approach for hypothesis testing for models without latent variables. BR (2002) suggested choosing the loss function to be the KL divergence function. The large-sample theory of the test statistics of BR has not been developed although it is well-defined under improper priors and can solve Jeffreys-Lindley's paradox.[1]

In a recent paper, LLY (2015) proposed 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_0 and H_1, respectively, $C_{\theta\theta}$ is the submatrix of C corresponding to $\boldsymbol{\theta}$. The statistic corresponding to this net loss function is given by

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

Under repeated sampling, LLY showed that $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ follows a χ^2 distribution asymptotically, providing an asymptotically pivotal quantity. This statistic is well-defined under improper priors and immune to Jeffreys-Lindley's paradox. Clearly, $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ requires evaluating the first-order derivative of the (observed-data) likelihood function. In some 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. Another feature of $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ is that it requires estimating both the null model and the alternative model although, under H_0, it was shown to be asymptotically equivalent to the Lagrange Multiplier (LM) test which requires estimating the null model only.

[1]Given that the KL function is not analytically available for most latent variable models, Li and Yu (2012) suggested basing the loss function on the Q-function used in the EM algorithm. However, its large-sample theory has not been developed. On the other hand, Li, Zeng and Yu (2014) suggested using the deviance function to be the loss function. large-sample theory of the test statistic is derived. Unfortunately, in general the asymptotic distribution depends on some unknown population parameters and hence the test is not pivotal asymptotically.

3 A Posterior Wald-type Statistic

3.1 The statistic based on a quadratic loss function

For any $\tilde{\boldsymbol{\vartheta}} \in \Theta$, denote

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

We propose the following net loss function for hypothesis testing:

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

where $\mathbf{V}_{\theta\theta}\left(\tilde{\boldsymbol{\vartheta}}\right)$ is the submatrix of $\mathbf{V}\left(\tilde{\boldsymbol{\vartheta}}\right)$ corresponding to $\boldsymbol{\theta}$, $\left[\mathbf{V}_{\theta\theta}\left(\tilde{\boldsymbol{\vartheta}}\right)\right]^{-1}$ the inverse of $\mathbf{V}_{\theta\theta}\left(\tilde{\boldsymbol{\vartheta}}\right)$, and $\bar{\boldsymbol{\vartheta}}$ the posterior mean of $\boldsymbol{\vartheta}$ under H_1. Then, the new test statistic can be defined as:

$$\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) = \int (\boldsymbol{\theta}-\boldsymbol{\theta}_0)' \left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\, p(\boldsymbol{\vartheta}|\mathbf{y})\mathrm{d}\boldsymbol{\vartheta} = \mathbf{tr}\left[\left[\mathbf{V}_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right)\right]^{-1}\mathbf{V}_{\theta}(\boldsymbol{\theta}_0)\right], \quad (3)$$

where $\mathbf{V}_{\theta}(\boldsymbol{\theta}_0) := \int (\boldsymbol{\theta}-\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)' p(\boldsymbol{\vartheta}|\mathbf{y})\mathrm{d}\boldsymbol{\vartheta}$.

Remark 3.1. *It is easy to see show that $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ 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, $\mathbf{V}_{\theta\theta}\left(\tilde{\boldsymbol{\vartheta}}\right)$, 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's paradox, consider the example used in LLY (2015). Let $y_1, y_2, \ldots, y_n \sim N(\theta, \sigma^2)$ with a known σ^2, the null hypothesis be $H_0 : \theta = 0$, the prior distribution of θ be $N(0, \tau^2)$. Denote $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. 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, 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 how much $\bar{y}$ is. This is exactly what Jeffreys-Lindley's paradox predicts. 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, $\mathbf{T}(\mathbf{y}, \theta_0)$ is distributed asymptotically as $\chi^2(1) + 1$ when H_0 is true, suggesting that $\mathbf{T}(\mathbf{y}, \theta_0)$ is immune to Jeffreys-Lindley'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 $0 \leq t \leq n$ 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 $1 \leq t \leq n$. 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

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

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

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

$$\mathcal{L}_n^{[j]}(\boldsymbol{\vartheta}) := \partial^j \log p(\boldsymbol{\vartheta}|\mathbf{y})/\partial \boldsymbol{\vartheta}^j, \mathbf{H}_n(\boldsymbol{\vartheta}) := \int \bar{\mathbf{H}}_n(\boldsymbol{\vartheta}) g(\mathbf{y})\, d\mathbf{y},\ \mathbf{J}_n(\boldsymbol{\vartheta}) := \int \bar{\mathbf{J}}_n(\boldsymbol{\vartheta}) g(\mathbf{y})\, d\mathbf{y}.$$

$\mathbf{H}_n(\boldsymbol{\vartheta})$ 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. A similar set of assumptions was used in Li, et al (2017).

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

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

Assumption 3: For all t, $l_t(\boldsymbol{\vartheta})$ satisfies the standard measurability and continuity condition, and the eight-times differentiability condition on $F_{-\infty}^{t} \times \Theta$ where $F_{-\infty}^{t} = \sigma(y_t, y_{t-1}, \cdots)$.

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

Assumption 5: For $j = 0, 1, 2, 3$, there exists a function $M_t(\mathbf{y}^t)$ such that for all $\boldsymbol{\vartheta} \in \Theta$, $l_t^{(j)}(\boldsymbol{\vartheta})$ exists, $\sup_{\boldsymbol{\vartheta} \in \Theta} \left\| l_t^{(j)}(\boldsymbol{\vartheta}) \right\| \leqslant M_t(\mathbf{y}^t)$, and $\sup_t E \left\| M_t(\mathbf{y}^t) \right\|^{r+\delta} \leq M < \infty$ for some $\delta > 0$, where r is the same as that in Assumption 2.

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

Assumption 7: Let $\boldsymbol{\vartheta}_n^0$ be the value that minimizes the KL loss between the DGP and the candidate model

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

where $\left\{\boldsymbol{\vartheta}_n^0\right\}$ is the sequence of minimizers interior to Θ uniformly in n. For all $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup \sup_{\Theta\backslash N\left(\boldsymbol{\vartheta}_n^0,\varepsilon\right)} \frac{1}{n} \sum_{t=1}^{n} \left\{ E\left[l_t\left(\boldsymbol{\vartheta}\right)\right] - E\left[l_t\left(\boldsymbol{\vartheta}_n^0\right)\right]\right\} < 0, \tag{4}$$

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

Assumption 8: The sequence $\left\{\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\}$ is negative definite.

Assumption 9: The prior density $p(\boldsymbol{\vartheta})$ is three-times continuously differentiable, $p\left(\boldsymbol{\vartheta}_n^0\right) > 0$ and $\int \|\boldsymbol{\vartheta}\|^2 p(\boldsymbol{\vartheta}) d\boldsymbol{\vartheta} < \infty$.

Remark 3.3. *An important condition for the asymptotic posterior normality is the consistency condition which means that, for each $\varepsilon > 0$, there exists $K(\varepsilon) > 0$ such that*

$$\lim_{n\to\infty} P\left(\sup_{\boldsymbol{\Theta}\backslash N\left(\boldsymbol{\vartheta}_n^0,\varepsilon\right)} \frac{1}{n} \sum_{t=1}^{n} \left[l_t\left(\boldsymbol{\vartheta}\right) - l_t\left(\boldsymbol{\vartheta}_n^0\right)\right] < -K\left(\varepsilon\right)\right) = 1; \tag{5}$$

see Heyde and Johnstone (1979), Schervish (2012), Ghosh and Ramamoorthi (2003). If Assumptions 1-7 hold true, then (5) holds, as shown in Li et al. (2017).

Remark 3.4. *According to Li et al. (2017), if Assumptions 1-9 hold true, then for each $\varepsilon > 0$, there exists $K(\varepsilon) > 0$ such that*

$$\lim_{n\to\infty} P\left(\sup_{\boldsymbol{\Theta}\backslash N\left(\widehat{\boldsymbol{\vartheta}}_m,\varepsilon\right)} \frac{1}{n} \left[\sum_{t=1}^{n} l_t\left(\boldsymbol{\vartheta}\right) - \sum_{t=1}^{n} l_t\left(\boldsymbol{\vartheta}_n^0\right)\right] < -K\left(\varepsilon\right)\right) = 1, \tag{6}$$

where $\widehat{\boldsymbol{\vartheta}}_m$ is the posterior mode of $\boldsymbol{\vartheta}$. Li et al. (2017) showed that this is sufficient to ensure that the concentration condition around the posterior mode given by Chen (1985).

Lemma 3.1. *Let $\widehat{\boldsymbol{\vartheta}}$ be the MLE of $\boldsymbol{\vartheta}$ and $\mathbf{N}_0(\delta) = \left\{\boldsymbol{\vartheta} : \left\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_n^0\right\| \leq \delta\right\}$. If Assumptions 1-7 hold true, then for any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that*

$$P\left(\sup_{N_0(\delta(\varepsilon))} \left|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}\right) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right| < \varepsilon\right) \to 1. \tag{7}$$

and

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

where $\mathbf{r}_0$ is q-dimension vector.

Let $\mathbf{\Sigma}_n = -\frac{1}{n}\bar{\mathbf{H}}_n^{-1}(\widehat{\boldsymbol{\vartheta}})$ and $\mathbf{z}_n = \mathbf{\Sigma}_n^{-1/2}\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)$. Lemma 3.2 below gives the order of the difference between the first k moments of the posterior distribution of $\mathbf{z}_n$ under H_1 and those of a standard multivariate normal distribution. To establish the closeness of higher order moments between the two distribution, we have to strengthen Assumption 9 by Assumption 9B. In Assumption 9B, the k-th order moment of the prior distribution is assumed to be finite.

Assumption 9B: The prior density $p(\boldsymbol{\vartheta})$ is three-times continuously differentiable, $p\left(\boldsymbol{\vartheta}_n^0\right) > 0$ and $\int \|\boldsymbol{\vartheta}\|^k p\left(\boldsymbol{\vartheta}\right) d\boldsymbol{\vartheta} < \infty$ for integer some $k \geq 1$.

Lemma 3.2. *Under Assumptions 1-8 and Assumption 9B, it can be shown that*

$$E\left[\mathbf{z}_n^{\{k\}}|\mathbf{y}, H_1\right] = \mathbf{MN}_q^{\{k\}} + o_p(1),$$

where $E\left[\mathbf{z}_n^{\{k\}}|\mathbf{y}, H_1\right]$ *is the k-th order moments of the posterior distribution of* $\mathbf{z}_n$ *under* H_1 *(i.e.* $\mathbf{z}_n|\mathbf{y}, H_1$*), and* $\mathbf{MN}_q^{\{k\}}$ *is the k-th order moments of a standard multivariate normal distribution with dimension q. When* $k = 1, 2$*, i.e., Assumption 9 holds, we can have*

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

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

Remark 3.5. *Under different regularity conditions, the Bernstein-von Mises theorem shows that the posterior distribution converges to a normal distribution with the MLE as its mean and the inverse of the empirical Hessian matrix evaluated at the MLE as its covariance. Based on the Bernstein-von Mises theorem, when the parameter is one-dimension, Ghosh and Ramamoorthi (2003) developed the same results as Lemma 3.2 for the i.i.d. case. Hence, Lemma 3.2 extends the results of Ghosh and Ramamoorthi (2003) in three aspects: (1) to the weakly dependent case; (2) to the multivariate case; (3) to show that the order of the difference in high-order moments between the posterior distribution and a normal distribution.*

Remark 3.6. *Assumptions 1-9 are weaker than those used in Li, et al. (2017) where a high order Laplace expansion was developed. With the high order Laplace expansion, Li, et al. (2018) derived the exact order for the difference in the first and second moments*

$$\bar{\boldsymbol{\vartheta}} = E\left[\boldsymbol{\vartheta}|\mathbf{y}, H_1\right] = \widehat{\boldsymbol{\vartheta}} + O_p(n^{-1}), \tag{10}$$

$$\mathbf{V}\left(\widehat{\boldsymbol{\vartheta}}\right) = E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)'|\mathbf{y}, H_1\right] = -\frac{1}{n}\bar{\mathbf{H}}_n^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right) + O_p(n^{-2}). \tag{11}$$

Clearly, (10) and (11) are a stronger set of results than (8) and (9). Lemma 3.2 is sufficient to develop large-sample properties of the proposed statistic. Hence, we can relax the assumptions of Li, et al (2017).

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

$$\text{Wald} = n\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right), \tag{12}$$

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 $\bar{\mathbf{H}}_{n}\left(\widehat{\boldsymbol{\vartheta}}\right)$.

Theorem 3.1. *Under Assumptions 1-9, we can show that, under the null hypothesis,*

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

and

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

Remark 3.7. *From Theorem 3.1,* $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta$ *may be regarded as the posterior version of the Wald statistic. It shares the same asymptotic distribution as the Wald test under the null hypothesis. However, the Wald statistic is based on the MLE of the alternative model, whereas the proposed test is based on the posterior mean and variance under the alternative hypothesis.*

Corollary 3.2. *Under Assumptions 1-9, we have, under the null hypothesis,*

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

Remark 3.8. *LLY (2015) has established the relationship between* $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0)$ *and the LM test statistic, i.e.,* $\mathbf{T}_{LLY}(\mathbf{y}, \boldsymbol{\theta}_0) = LM + o_p(1)$ *under the null hypothesis. It is noted in Engle (1984) that under the null hypothesis* $LM = Wald + o_p(1)$*. So Corollary 3.2 is the posterior version of this asymptotic equivalence between the Wald and LM statistics.*

Remark 3.9. *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, i.e.,*

$$\textit{Accept}\ \ H_0\ \textit{if}\ \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta \le c;\ \ \textit{Reject}\ \ H_0\ \ \textit{if}\ \mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0) - q_\theta > c.$$

Remark 3.10. *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* $\boldsymbol{\theta}$ *and, thus, it is very easy to compute. In contrast, the Wald statistic in (12) 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, this inversion is numerically more involved than the inversion of the submatrix. 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 3.11. *Compared with the Wald statistic, the proposed statistic can incorporate the prior information through the posterior distribution. To illustrate the influence of prior distribution, let* $y_1, ..., y_n \sim N(\theta, \sigma^2)$ *with a known variance* $\sigma^2 = 1$*. The true value of* θ *is set at* $\theta_0 = 0.10$*. The prior distribution of* θ *is set as* $N(\mu_0, \tau^2)$*. We wish to test* $H_0 : \theta = 0$*. It can be shown that*

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

where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$*. When* $n \to \infty$, $\mathbf{T}(\mathbf{y}, \theta_0) - 1 - Wald \xrightarrow{p} 0$ *and the asymptotic distribution for both* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *and the Wald statistic is* $\chi^2(1)$*. Suppose two prior distributions are used, a highly informative prior* $N(0.10, 10^{-3})$ *and a very vague prior* $N(0, 10^{50})$*. Table 1 reports* $2\log BF_{10}$, $\mathbf{T}(\mathbf{y}, \theta_0) - 1$*, and Wald when* $n = 10, 100, 1000, 10000$ *under these two priors. It can be seen that* $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *and Wald take identical values when the vague prior is used. It is consistent with the prediction of our asymptotic theory. Moreover, both the BF and the new statistic depend on the prior (although the BFs tend to choose the wrong model under the vague prior even when the sample size is very large) while the Wald test is independent of the prior. When* $n = 10, 100$, $\mathbf{T}(\mathbf{y}, \theta_0) - 1$ *correctly rejects the null hypothesis when the prior is informative but fails to reject it when the prior is vague under the 5% significance level. In this case, the Wald test fails to reject the null hypothesis.*

Table 1: Comparison of $2\log BF_{10}$, $\mathbf{T}(\mathbf{y}, \theta_0) - 1$, and the Wald statistic

Prior	$N(0.10, 10^{-3})$				$N(0, 10^{50})$			
n	10	100	1000	10000	10	100	1000	10000
$2\log BF_{10}$	9.96	11.12	20.60	93.58	-117.42	-118.50	-110.72	-38.00
$\mathbf{T}(\mathbf{y}, \theta_0) - 1$	9.96	11.22	21.30	95.98	0.01	1.23	11.32	86.03
Wald	0.01	1.23	11.32	86.03	0.01	1.23	11.32	86.03

Remark 3.12. *Assumption 9 requires finiteness of the first and second moments of the posterior distribution. When improper priors satisfies this assumption, Theorem 3.1 holds. In practice, however, many improper priors do not have finite first and second moments and hence Assumption is violated. In addition, Assumption 9 excludes the Jeffreys prior (Jeffreys, 1961) since the Jeffreys prior depends on the sample size* n*. If informative priors are not available, we suggest using vague noninformative priors (a prior with large variance spread) to implement our proposed tests. For more details about vague noninformative priors, one can refer to Kass and Raftery (1995).*

3.3 Extension to hypotheses in a general form

In this subsection, we extend the point null hypothesis to the following nonlinear form,

$$\begin{cases} H_0: & R(\boldsymbol{\theta}_0) = \boldsymbol{r} \\ H_1: & R(\boldsymbol{\theta}_0) \neq \boldsymbol{r} \end{cases}, \tag{13}$$

where $R(\cdot) : \boldsymbol{\Theta}_\theta \to \mathbb{R}^m$, $m \leq q$, and $\boldsymbol{r} \in \mathbb{R}^m$. Here R is a set of m nonlinear 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 problem is in the standard form for the Wald test, it makes BFs difficult to implement due to nonlinear relationships among parameters. 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 $\boldsymbol{\Theta}$ and full rank at $\boldsymbol{\theta}_0$.

For the hypothesis defined in (13), the classical Wald statistic and its asymptotic theory are

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

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

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

and introduce our test statistic as:

$$\begin{aligned} \mathbf{T}(\mathbf{y}, \mathbf{r}) &= \int_{\boldsymbol{\Theta}} \Delta\mathcal{L}(H_0, \boldsymbol{\theta}, \boldsymbol{\psi})\, p(\boldsymbol{\vartheta}|\mathbf{y})\, d\boldsymbol{\vartheta} \\ &= \int_{\boldsymbol{\Theta}} (R(\boldsymbol{\theta}) - \boldsymbol{r})' \left[\frac{\partial R\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} V_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}\right) \frac{\partial R\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}} \right]^{-1} (R(\boldsymbol{\theta}) - \mathbf{r})\, p(\boldsymbol{\vartheta}|\mathbf{y})\, d\boldsymbol{\vartheta} \\ &= \mathbf{tr}\left[\left(\frac{\partial R\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} V_{\theta\theta}\left(\overline{\boldsymbol{\vartheta}}\right) \frac{\partial R\left(\overline{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}} \right)^{-1} \mathbf{V}_\theta(\mathbf{r}) \right], \end{aligned} \tag{14}$$

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. *Under Assumptions 1-10, , we can show that, under the null hypothesis,*

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

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 $\boldsymbol{\vartheta}$, i.e., $\overline{\boldsymbol{\vartheta}}$ 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 $\{\boldsymbol{\vartheta}^{[j]}, j = 1, 2, \ldots, J\}$ be effective samples generated from $p(\boldsymbol{\vartheta}|\mathbf{y})$, consistent estimates of $\overline{\boldsymbol{\vartheta}}$ and $\mathbf{V}\left(\overline{\boldsymbol{\vartheta}}\right)$ are given by

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

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

$$\begin{aligned}
\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{\theta}_0\right) &:= \mathbf{tr}\left[\left(\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\right)^{-1}\bar{\mathbf{V}}_{\theta}\left(\boldsymbol{\theta}_0\right)\right], \\
\widehat{\mathbf{T}}\left(\mathbf{y}, \mathbf{r}\right) &:= \mathbf{tr}\left[\left(\frac{\partial R\left(\bar{\bar{\boldsymbol{\theta}}}\right)}{\partial \boldsymbol{\theta}'}\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\frac{\partial R\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)}{\partial \boldsymbol{\theta}}\right)^{-1}\bar{\mathbf{V}}_{\theta}(\mathbf{r})\right],
\end{aligned} \tag{15}$$

where

$$\bar{\mathbf{V}}_{\theta}\left(\boldsymbol{\theta}_0\right) = \frac{1}{J}\sum_{j=1}^{J}\left(\boldsymbol{\theta}^{[j]} - \boldsymbol{\theta}_0\right)\left(\boldsymbol{\theta}^{[j]} - \boldsymbol{\theta}_0\right)',$$

and

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

Remark 3.13. *Various approaches have been developed for posterior simulation. Examples include Monte Carlo (MC) integration, important 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 the random observations from $p(\boldsymbol{\vartheta}|\mathbf{y})$. From (15), the proposed statistics are by-products of posterior simulation. Furthermore, the test statistics can be applied in a variety of models.*

When $\widehat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_{\mathbf{0}})$ 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) which measures the magnitude of simulation errors. The following theorem provides formulae to calculate the NSE of $\widehat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_{\mathbf{0}})$ and $\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{r}\right)$.

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

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

where

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

and

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

Furthermore, if $R\left(\boldsymbol{\theta}\right)$ *is second-order continuously differentiable, the NSE of* $\widehat{\mathbf{T}}\left(\mathbf{y}, \mathbf{r}\right)$ *is given by*

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

where

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

$$\bar{\mathbf{v}}_3 = R\left(\frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\theta}^{[j]}\right) = R\left(\bar{\mathbf{v}}_1\right), \ \bar{\mathbf{V}}_4 = \frac{\partial R\left(\frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\theta}^{[j]}\right)}{\partial \boldsymbol{\theta}} = \frac{\partial R\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\bar{\mathbf{v}}_1},$$

$$\begin{aligned}\frac{\partial vech\left(\bar{\mathbf{V}}_4' \bar{\mathbf{V}}_2 \bar{\mathbf{V}}_4\right)}{\partial \bar{\mathbf{v}}} =& \left(\left(\bar{\mathbf{V}}_2 \bar{\mathbf{V}}_4\right)' \otimes I_m\right) \frac{\partial \bar{\mathbf{V}}_4'}{\partial \bar{\mathbf{v}}_1} \frac{\partial \bar{\mathbf{v}}_1}{\partial \bar{\mathbf{v}}} + \left(\bar{\mathbf{V}}_4 \otimes \bar{\mathbf{V}}_4'\right) \frac{\partial \bar{\mathbf{V}}_2}{\partial \bar{\mathbf{v}}} \\ & + \left(I_m \otimes \bar{\mathbf{V}}_4' \bar{\mathbf{V}}_2\right) \frac{\partial \bar{\mathbf{V}}_4}{\partial \bar{\mathbf{v}}_1} \frac{\partial \bar{\mathbf{v}}_1}{\partial \bar{\mathbf{v}}},\end{aligned}$$

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

Remark 3.14. *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 Hypothesis Testing for Latent Variable Models

Latent variable models have found a wide range of applications in microeconometrics, macroeconometrics and financial econometrics; see Stern (1997), Norets (2009), Koop and Korobilis (2009), Yu (2011). Without loss of generality, let $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)'$ denote the observed variables and $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n)'$ the latent variables. The set of parameters in the model is denoted by $\boldsymbol{\vartheta}$. Let $p(\mathbf{y}|\boldsymbol{\vartheta})$ be the likelihood function of the observed data, and $p(\mathbf{y}, \boldsymbol{z}|\boldsymbol{\vartheta})$ be the complete-data likelihood function. The relationship between these two likelihood functions is

$$p(\mathbf{y}|\boldsymbol{\vartheta}) = \int p(\mathbf{y}, \boldsymbol{z}|\boldsymbol{\vartheta})d\boldsymbol{z}. \tag{16}$$

In many latent variable models, especially dynamic latent variable models, the number of latent variables is often the same as the sample size. Hence, the integral in (16) is high-dimensional when the sample size is large. If the integral does not have an analytical expression, it will be very difficult to evaluate numerically. Consequently, statistical inferences, including estimation and hypothesis testing, are difficult to implement if they are based on the MLE.

In recent years, it has been documented that latent variable models can be efficiently analyzed using MCMC techniques; see Geweke, et al. (2011). Let $p(\boldsymbol{\vartheta})$ be the prior distribution of $\boldsymbol{\vartheta}$. To alleviate the difficulty in maximum likelihood, the data-augmentation strategy (Tanner and Wong, 1987) is often employed where the latent variables are treated as additional parameters. Then, the Gibbs sampler can be used to generate random samples from the joint posterior distribution $p(\boldsymbol{\vartheta}, \boldsymbol{z}|\mathbf{y})$, denoted by $\left\{\boldsymbol{\vartheta}^{[j]}, \boldsymbol{z}^{[j]}\right\}_{j=1}^{J}$, after a burn-in phase. The Bayesian estimates of $\boldsymbol{\vartheta}$ and the estimates of the covariance matrix can be obtained as,

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

Similarly, the proposed test can be easily computed from $\left\{\boldsymbol{\vartheta}^{[j]}\right\}_{j=1}^{J}$ and hence it is very easy to implement.

Remark 4.1. *As noted before, the test statistic of LLY in (2) requires the evaluation of the first derivative of the observed-data likelihood function. For many latent variable model, this is difficult to evaluate when the observed-data likelihood function does not have a closed-form expression. In addition, it requires estimating both the null model and the alternative model. However, the proposed test does not require evaluating the first derivative and only estimate the model under the alterative hypothesis. Clearly, the proposed test is easier to implement and faster to compute.*

5 Simulation Studies

In this section, we first design two 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 between $\mathbf{T}(\mathbf{y}, \boldsymbol{\theta}_0)$ and the Wald statistic in terms of size and power. In the second 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 1000 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

$$\boldsymbol{\mu}_0 = (0,0,0,0)', \mathbf{V}_0 = 1000 \times I_4, a = 0.0001, b = 0.0001,$$

with I_4 being the 4×4 identity matrix. In each replication, we draw 5000 i.i.d. random samples from the posterior distribution and then use the posterior samples to compute the proposed statistic. Also computed is the Wald statistic 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 the proposed test and the Wald test for a nominal size of 5%. In all cases, the size distortion for the new statistic 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 the proposed test is comparable to that of the Wald statistic. As C increases, the power of the proposed statistic approaches 100%. Similarly, as the sample size increases, the power of the proposed statistic approaches 100%.

Table 2: The size and power of the proposed test and the Wald test for different null hypothesis in a 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%

5.2 Hypothesis testing in a discrete choice model

The second 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 censor variables. Consequently, the likelihood function and the corresponding derivatives are not easy to evaluate. Consequently, Li introduced a MCMC approach to do 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,

$$\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x + \tilde{\epsilon}_i, \tilde{\epsilon}_i \sim N\left(0, \tilde{\sigma}^2\right). \tag{17}$$

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 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,

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

Table 3: The size and power of the proposed test in a discrete choice model

	Empirical Size	Empirical Power		
	$\beta_1 = 0$	$\beta_1 = 0.1$	$\beta_1 = 0.2$	$\beta_1 = 0.4$
$n = 100$	4.2%	11.0%	23.0%	75.8%
$n = 250$	5.2%	24.0%	65.0%	100%
$n = 500$	4.6%	49.4%	97.2%	100%

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,

$$\boldsymbol{\beta} = \left(\beta_0, \beta_1.\tilde{\beta}_0, \tilde{\beta}_1\right)' \sim N\left(0, 1000 \times I_4\right), \ \Sigma \sim IW\left(6, 6 \times I_2\right), \gamma_3 \sim Beta\left(1, 1\right),$$

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 (15), 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 3 for a nominal size of 5%. It is obvious that the empirical size is close 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 increase, 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 the presence of 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 labels the schooling outcome. Let $z_{hi} = 1$ if individual i dropped out of high school after completing the ninth grade, $z_{hi} = 2$ if he dropped out after completing the tenth grade, $z_{hi} = 3$ if he dropped out after completing the eleventh grade, and $z_{hi} = 4$ if he completed high school. An ordered probit is specified as

$$\begin{cases} y_{hi} = \boldsymbol{\beta}_h' \boldsymbol{x}_{hi} + \epsilon_{hi}, & \epsilon_{hi} \sim N\left(0, \sigma_h^2\right), \gamma_{z_{hi}} < y_{hi} < \gamma_{z_{hi}+1}, \\ \gamma_1 = -\infty, \gamma_2 = 0, & \gamma_2 < \gamma_3 < \gamma_4, \gamma_4 = 1, \gamma_5 = \infty, \end{cases} \tag{18}$$

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,

$$y_{ui} \begin{cases} \leq 0, & \omega_{ui} = 0, \\ = \omega_{ui}, & 0 < \omega_{ui} < 1, \\ \geq 1, & \omega_{ui} = 1. \end{cases} \tag{19}$$

Thus, the censored regression is,

$$y_{ui} = \boldsymbol{\beta}_u' \boldsymbol{x}_{ui} + \boldsymbol{s}_i' \boldsymbol{\eta} + \epsilon_{ui}, \epsilon_{ui} \sim N\left(0, \sigma_u^2\right), \tag{20}$$

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 (20), $\boldsymbol{s}_i$ 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,

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

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,

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_h', \boldsymbol{\beta}_u'\right)' \sim N\left(0_{k\times 1}, 1000 \times I_k\right),\ \Sigma \sim IW\left(6, 6 \times I_2\right),$$

$$\boldsymbol{\eta} \sim N\left(0, I_4\right),\ \gamma_3 \sim Beta\left(1, 1\right),$$

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 4, all of which are very close to those reported in Li.

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 5.[2]

According to Table 5, both $\widehat{\mathbf{T}}\left(\mathbf{y}, \boldsymbol{\theta}_0\right) - 1$ 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.

6.1.1 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 stochastic volatility model with leverage effect is defined as,

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

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

Table 4: Posterior means and posterior standard errors of parameters in a discrete choice model of Li (2006)

	$E(\cdot\|Data)$	$SE(\cdot\|Data)$
	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 5: The proposed statistic, $\widehat{\mathbf{T}}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$, $\widehat{\log BF_{10}}$, the CPU time (in seconds), and their NSEs in the discrete choice model of Li (2006).

	$\beta_4=\beta_5=0$		
	Value	NSE	CPU Time (seconds)†
$\widehat{\mathbf{T}}\left(\mathbf{y},\boldsymbol{\theta}_0\right)-1$	44.39	1.59	22.54
$\widehat{\mathbf{T}}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$	2502.00	89.57	39,096.79
$\widehat{\log BF_{10}}$	5.2019	1.03	292,886.45

† The CPU time for computing each statistic is obtained from a laptop with an Intel i5 CPU and 8 GB memory after MCMC outputs are available.

with

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

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, $\mathbf{T}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ and $\widehat{\log BF_{10}}$.[3]

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 $\widehat{\mathbf{T}}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ 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 01/10/81 to 28/06/85 with sample size 945. The series r_t is the daily mean-corrected returns. The following vague priors are used:

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

We draw 50,000 from the posterior distribution and discard the first 20,000 as build-in samples. Then we store every 5th value of the remaining samples as effective random samples. The estimation results are reported in Table 6.

Table 7 reports the proposed statistic, $\widehat{\mathbf{T}}_{LLY}\left(\mathbf{y},\boldsymbol{\theta}_0\right)$ and $\widehat{\log BF_{10}}$ and the NSEs for the first two statistics. Since the observed-data likelihood function is expensive to compute,

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

Table 6: Posterior means of parameters in the SV model with and without leverage effect for the Pound/Dollar returns.

	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 7: The proposed statistic, $\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\ \boldsymbol{\theta}_0)$, $\log\widehat{BF}_{10}$, the CPU time (in seconds), and the NSEs of the first two statistics for the Pound/Dollar returns.

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

the NSE of BF is too difficult to obtain and not report. $\log\widehat{BF}_{10}$ 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 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 8.

Table 8: Posterior means of parameters in the SV model with and without leverage effect 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 9. 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, $\widehat{\log BF_{10}}$ 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 $\widehat{\log BF_{10}}$ after MCMC outputs are available.

Table 9: The proposed statistic, $\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\ \boldsymbol{\theta}_0)$, $\widehat{\log BF_{10}}$, the CPU time (in seconds), and the NSEs of the first two statistics for the S&P500 returns.

	$\widehat{\mathbf{T}}(\mathbf{y},\boldsymbol{\theta}_0)-1$	$\widehat{\mathbf{T}}_{LLY}(\mathbf{y},\boldsymbol{\theta}_0)$	$\widehat{\log BF_{10}}$
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 test statistic is proposed to test for a point null hypothesis which 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'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 imposes nonlinear 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.

The finite sample properties of the proposed statistic is examined in a linear regression model and in a discrete choice model with latent variables. In the linear regression models, the Wald statistics 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 has little size distortion 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 apply the method to two models using real data. The first one is a discrete choice model and the second is a 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

8.1 Proof of Lemma 3.1

First, we can show that

$$\begin{aligned} & \left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta})-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\| \\ = & \left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta})-\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)+\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)-\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)+\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\| \\ \leq & \left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta})-\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\|+\left\|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)-\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\|+\left\|\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\| . \end{aligned} \tag{21}$$

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

$$P\left(\sup_{N\left(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)\right)}\left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta})-\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\|<\frac{\varepsilon}{3}\right) \rightarrow 1. \tag{22}$$

From Assumption 3 that $l_t^{(2)}(\boldsymbol{\vartheta})$ is almost surely continuous at $\boldsymbol{\vartheta}_n^0$. We also have

$$P\left(\left\|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)-\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\|<\frac{\varepsilon}{3}\right) \rightarrow 1,\ P\left(\left\|\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\|<\frac{\varepsilon}{3}\right) \rightarrow 1, \tag{23}$$

because of the uniform convergence of $l_t^{(2)}(\boldsymbol{\vartheta})$ and $\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_n^0 \xrightarrow{p} 0$ by Assumptions 1-7 (Gallant and White, 1988). Define events $A_n(\varepsilon)=\left\{\sup_{N\left(\boldsymbol{\vartheta}_n^0, \delta(\varepsilon)\right)}\left\|\bar{\mathbf{H}}_n(\boldsymbol{\vartheta})-\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\|<\frac{\varepsilon}{3}\right\}$, $B_n(\varepsilon)=\left\{\left\|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)-\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\|<\frac{\varepsilon}{3}\right\}$ and $C_n(\varepsilon)=\left\{\left\|\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\|<\frac{\varepsilon}{3}\right\}$. Then we have

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

From (22) and (23), the probability of the complementary event of $A_n(\varepsilon) \cap A_n(\varepsilon) \cap A_n(\varepsilon)$ is

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

Then

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

Hence, by (21), for any $\varepsilon > 0$

$$\begin{aligned}
& P\left(\sup_{N(\boldsymbol{\vartheta}_n^0,\delta(\varepsilon))} \left\|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}\right) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\| < \varepsilon\right) \\
\geq\ & P\left(\sup_{N(\boldsymbol{\vartheta}_n^0,\delta(\varepsilon))} \left\|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}\right) - \bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\| + \left\|\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}_n^0\right) - \mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right)\right\| + \left\|\mathbf{H}_n\left(\boldsymbol{\vartheta}_n^0\right) - \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)\right\| < \varepsilon\right) \\
\to\ & 1.
\end{aligned} \tag{24}$$

It is noted that

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

where λ_n is the smallest eigenvalue of $-\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\vartheta}}\right)$. Then from (24), for any $\varepsilon > 0$

$$P\left(\sup_{N(\boldsymbol{\vartheta}_n^0,\delta(\varepsilon)),\|\mathbf{r}_0\|=1} \left|1 - \mathbf{r}_0'\bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right)\bar{\mathbf{H}}_n\left(\boldsymbol{\vartheta}\right)\bar{\mathbf{H}}_n^{-1/2}\left(\widehat{\boldsymbol{\vartheta}}\right)\mathbf{r}_0\right| < \varepsilon\right) \to 1. \tag{25}$$

8.2 Proof of Lemma 3.2

Lemma 8.1. *Let $X_1, X_2, ..., X_q$ be independently and identically distributed, then the following inequality for the order statistic $\max_i X_i$ holds*

$$E\left[\left(\max_i |X_i|\right)^k\right] < \sqrt{2}\exp\left(\frac{5}{3}\right)\frac{q+1}{\sqrt{q}}\left[E|X_1|^{2k}\right]^{1/2},$$

under the condition that $E|X_1|^{2k} < \infty$ and $k > 0$.

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

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

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 $\rho = 1/2$, it can be shown that

$$\begin{aligned} E\left[\left|\max_i X_i\right|^k\right] &< C\left(\frac{1}{2}\right)\left\{E|X_1|^\delta g^{-1}\left(\frac{q}{q+1}\right)\right\}^{1/2} \\ &= \sqrt{2}\exp\left(\frac{5}{3}\right)\frac{q+1}{\sqrt{q}}\left[E|X_1|^{2k}\right]^{1/2}, \end{aligned} \tag{26}$$

for $q \geq 2$.

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

$$E\left[\left|\max_i X_i\right|^k\right] = E\left[|X_1|^k\right] \leq \left[E|X_1|^{2k}\right]^{1/2},$$

then

$$E\left[\left|\max_i X_i\right|^k\right] < \sqrt{2}\exp\left(\frac{5}{3}\right)\frac{1+1}{\sqrt{1}}\left[E|X_1|^{2k}\right]^{1/2}. \tag{27}$$

From (26) and (27), we can get

$$E\left[\left|\max_i X_i\right|^k\right] < \sqrt{2}\exp\left(\frac{5}{3}\right)\frac{q+1}{\sqrt{q}}\left[E|X_1|^{2k}\right]^{1/2}, \tag{28}$$

for $k > 0$ and $q \geq 1$.

Let $Y_i = |X_i|$, then it is easy to show that

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

by (28).

■

Lemma 8.2. *Suppose the posterior density of $\boldsymbol{\vartheta}$ can be written as*

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

where

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

Then

$$\lim_{n\to\infty} P\left(\int_{A_n} \|\mathbf{z}_n\|^k \left| p(\mathbf{z}_n|\mathbf{y}) - (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right| d\mathbf{z}_n > \varepsilon\right) = 0, \tag{29}$$

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* $\mathbf{z}_n$ $\left(:= \boldsymbol{\Sigma}_n^{-1/2}\left(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}}\right)\right)$, $\boldsymbol{\Sigma}_n^{-1} = -\frac{\partial^2 \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}$.

Proof. The posterior density of $\mathbf{z}_n$, $p(\mathbf{z}_n|\mathbf{y})$, can be written as

$$p(\mathbf{z}_n|\mathbf{y}) = \frac{|\boldsymbol{\Sigma}_n|^{1/2}\, p(\mathbf{y}|\boldsymbol{\vartheta})\, p(\boldsymbol{\vartheta})}{p(\mathbf{y})} = \frac{|\boldsymbol{\Sigma}_n|^{1/2}\, p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) p\left(\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)}{p(\mathbf{y})} \tag{30}$$

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

$$\begin{aligned}
& \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) \\
= \; & \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) + \frac{1}{2}\mathbf{z}_n'\boldsymbol{\Sigma}_n^{1/2}\frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\theta\partial\boldsymbol{\vartheta}'}\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n \\
= \; & \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) - \frac{1}{2}\mathbf{z}_n'\boldsymbol{\Sigma}_n^{1/2}\left[-\frac{\partial^2 \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)}{\partial\theta\partial\boldsymbol{\vartheta}'} - \frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\theta\partial\boldsymbol{\vartheta}'} + \frac{\partial^2 \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\right]\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n \\
= \; & \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) - \frac{1}{2}\mathbf{z}_n'\boldsymbol{\Sigma}_n^{1/2}\left[\boldsymbol{\Sigma}_n^{-1} - \frac{\partial^2 \log p\left(\mathbf{y}|\widetilde{\boldsymbol{\vartheta}}_1\right)}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'} - \boldsymbol{\Sigma}_n^{-1}\right]\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n \\
= \; & \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right) - \frac{1}{2}\mathbf{z}_n'\left[\mathbf{I}_q - \mathbf{R}_n(\boldsymbol{\vartheta}, \mathbf{y})\right]\mathbf{z}_n,
\end{aligned} \tag{31}$$

where $\mathbf{I}_q$ is a q-dimension identity matrix and

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

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

To prove (29), note that

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

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

and that

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

by Chen (1985) and Schervish (2012). Hence, according to (31), to verify (29), it is sufficient to show

$$P\left(\int_{A_n}\|\mathbf{z}_n\|^k\left|\frac{p\left(\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)}{p\left(\boldsymbol{\vartheta}_n^0\right)}\exp\left[-\frac{\mathbf{z}_n'[\mathbf{I}_q-\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})]\mathbf{z}_n}{2}\right]-\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|d\mathbf{z}_n<\varepsilon\right)\to 1 \tag{32}$$

Hence, to ensure (32), by assumption 9, it is enough to prove

$$P\left(\int_{A_n}\|\mathbf{z}_n\|^k\left|p\left(\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right)\exp\left[-\frac{\mathbf{z}_n'[\mathbf{I}_q-\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})]\mathbf{z}_n}{2}\right]-p\left(\boldsymbol{\vartheta}_n^0\right)\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|d\mathbf{z}_n<\varepsilon\right)\to 1 \tag{33}$$

In the following, we prove that (33) holds. Since the prior density function is continuous at $\boldsymbol{\vartheta}_n^0$, that is, 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 $\boldsymbol{\vartheta}$ satisfying $\left\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_n^0\right\|\leq\delta_1$, that is, $\boldsymbol{\vartheta}\in\mathbf{N}_0(\delta_1)=\left\{\boldsymbol{\vartheta}:\left\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_n^0\right\|\leq\delta_1\right\}$,

$$\left|p(\boldsymbol{\vartheta})-p\left(\boldsymbol{\vartheta}_n^0\right)\right|=\left|p\left(\widehat{\boldsymbol{\vartheta}}+\Sigma_n^{1/2}\mathbf{z}_n\right)-p\left(\boldsymbol{\vartheta}_n^0\right)\right|\leq\eta p\left(\boldsymbol{\vartheta}_n^0\right). \tag{34}$$

Furthermore, by Lemma 3.1, $\forall\eta>0$, $\exists\delta_2>0$,

$$\lim_{n\to\infty}P\left(\sup_{\vartheta\in N_0(\delta_2),\|\mathbf{r}_0\|=1}\left|1+\mathbf{r}_0'\boldsymbol{\Sigma}_n^{1/2}\frac{\partial^2\log p(\mathbf{y}|\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\boldsymbol{\Sigma}_n^{1/2}\mathbf{r}_0\right|<\eta\right)=1, \tag{35}$$

where $\mathbf{N}_0(\delta)=\left\{\boldsymbol{\vartheta}:\left\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_n^0\right\|\leq\delta\right\}$, see Schervish (2012).

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

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

and

$$C_n=\|\mathbf{z}_n\|^k\left|p\left(\widehat{\boldsymbol{\vartheta}}+\Sigma_n^{1/2}\mathbf{z}_n\right)\exp\left[-\frac{1}{2}\mathbf{z}_n'\left[\mathbf{I}_q-\mathbf{R}_n(\tilde{\boldsymbol{\vartheta}}_1)\right]\mathbf{z}_n\right]-p\left(\boldsymbol{\vartheta}_n^0\right)\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|. \tag{36}$$

The integration of C_n in the space A_n can be decomposed into two areas, A_{1n} and A_{2n}, i.e.,

$$J = \int_{A_n} C_n d\mathbf{z}_n = J_1 + J_2,$$

where $J_1 = \int_{A_{1n}} C_n d\mathbf{z}_n$, $J_2 = \int_{A_{2n}} C_n d\mathbf{z}_n$. In the following, we try to prove

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

For J_1, we note that

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

where

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

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

Then we have

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

where

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

It is noted that since $\delta \leq \delta_1$, from (34), we can know that $\left| p\left(\widehat{\boldsymbol{\vartheta}} + \Sigma_n^{1/2} \mathbf{z}_n\right)\right| \leq (1+\eta)\, p\left(\boldsymbol{\vartheta}_n^0\right)$. Hence, we can have

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

Let $\mathbf{r}_0 = \mathbf{z}_n / \|\mathbf{z}_n\|$, so $\|\mathbf{r}_0\| = 1$, then, we can get that

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

where $\widetilde{\boldsymbol{\vartheta}}_1$ lies between $\boldsymbol{\vartheta}$ and $\widehat{\boldsymbol{\vartheta}}$. Since $\widehat{\boldsymbol{\vartheta}} \xrightarrow{p} \boldsymbol{\vartheta}_n^0$, we can get that with probability 1, $\hat{\boldsymbol{\vartheta}} \in N_0(\delta)$, and hence, $\tilde{\boldsymbol{\vartheta}}_1 \in N_0(\delta)$ with probability 1.

Following (35), with probability 1, when $\boldsymbol{\theta} \in N_0(\delta)$, we can further get that

$$\|\mathbf{z}_n\|^k \left| \exp\left[-\frac{1}{2}\mathbf{z}_n' \left[\mathbf{I}_q - \mathbf{R}_n\left(\boldsymbol{\vartheta}, \mathbf{y}\right)\right] \mathbf{z}_n\right] - \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|$$

$$
\begin{aligned}
&= \|\mathbf{z}_n\|^k \left|\exp\left[\frac{1}{2}\mathbf{z}_n'\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})\,\mathbf{z}_n\right]-1\right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\\
&\le \|\mathbf{z}_n\|^k \exp\left[\left|\frac{1}{2}\mathbf{z}_n'\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})\,\mathbf{z}_n\right|\right] \left|\frac{1}{2}\mathbf{z}_n'\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})\,\mathbf{z}_n\right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\\
&= \|\mathbf{z}_n\|^k \exp\left[\left|\frac{1}{2}\mathbf{z}_n'\mathbf{z}_n\right| \left|\mathbf{r}_0'\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})\,\mathbf{r}_0\right|\right] \left|\frac{1}{2}\mathbf{z}_n'\mathbf{z}_n\right| \left|\mathbf{r}_0'\mathbf{R}_n(\boldsymbol{\vartheta},\mathbf{y})\,\mathbf{r}_0\right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\\
&\le \frac{\eta}{2}\|\mathbf{z}_n\|^k \exp\left[\left|\frac{\eta}{2}\mathbf{z}_n'\mathbf{z}_n\right|\right] \left|\mathbf{z}_n'\mathbf{z}_n\right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\\
&= \frac{\eta}{2}\|\mathbf{z}_n\|^{k+2} \exp\left(-\frac{(1-\eta)\mathbf{z}_n'\mathbf{z}_n}{2}\right) \qquad (37)
\end{aligned}
$$

Let

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

It follows from (37), we can get that

$$
\lim_{n\to\infty} P\left\{J_{11}^* \le \frac{\eta}{2}\int_{A_{1n}} \|\mathbf{z}_n\|^{k+2}\exp\left(-\frac{(1-\eta)\,\mathbf{z}_n'\mathbf{z}_n}{2}\right)d\mathbf{z}_n\right\} = 1. \qquad (38)
$$

It is noted that, by Lemma 8.1, we have

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

where $\mathbf{z}_{ni}$ is the ith element of $\mathbf{z}_n$ and the penultimate equation results from the fact that the central absolute moment of a scalar normal random variable X with mean μ and variance σ^2 is

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

Hence, we have,

$$\lim_{n\to\infty} P\left(\frac{J_{11}}{C_{J_1}} \le \frac{q^2\eta(1+\eta)\sqrt{(2k+1)(2k+3)}}{2(1-\eta)^{\frac{q+k+2}{2}}}\right) = 1, \tag{39}$$

where

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

In the following, we deal with J_{12}. From (34) and Lemma 8.1, we have

$$\begin{aligned}
J_{12} &\le \int_{A_{1n}} \|\mathbf{z}_n\|^k \left|p\left(\widehat{\boldsymbol{\vartheta}}+\Sigma_n^{1/2}\mathbf{z}_n\right) - p\left(\boldsymbol{\vartheta}_n^0\right)\right| \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
&\le \eta p\left(\boldsymbol{\vartheta}_n^0\right)\int_{A_{1n}} \|\mathbf{z}_n\|^k \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
&\le \eta p\left(\boldsymbol{\vartheta}_n^0\right)\int_{\mathbb{R}^q} \|\mathbf{z}_n\|^k \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
&= \eta p\left(\boldsymbol{\vartheta}_n^0\right)(2\pi)^{q/2}\int_{\mathbb{R}^q} \|\mathbf{z}_n\|^k (2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \\
&\le \eta p\left(\boldsymbol{\vartheta}_n^0\right)(2\pi)^{q/2} q^k\int_{\mathbb{R}^q} \left(\max_i |\mathbf{z}_{ni}|\right)^k (2\pi)^{-q/2}\exp\left(-\frac{1-\eta}{2}\mathbf{z}_n'\mathbf{z}_n\right) d\mathbf{z}_n \\
&\le \sqrt{2}\exp\left(\frac{5}{3}\right)\left(\frac{q+1}{\sqrt{q}}\right)\eta p\left(\boldsymbol{\vartheta}_n^0\right)(2\pi)^{q/2} q^k\left[\int_{\mathbb{R}} |t|^{2k}(2\pi)^{-1/2}\exp\left(-\frac{t^2}{2}\right) d\mathbf{t}\right]^{1/2} \\
&= \eta\sqrt{2}\exp\left(\frac{5}{3}\right) p\left(\boldsymbol{\vartheta}_n^0\right)(2\pi)^{q/2}(q+1)\,q^{k-\frac{1}{2}}2^{k/2}\left(\frac{\Gamma\left(\frac{2k+1}{2}\right)}{\sqrt{\pi}}\right)^{1/2} \\
&= \eta\exp\left(\frac{5}{3}\right) p\left(\boldsymbol{\vartheta}_n^0\right) 2^{\frac{q+k+1}{2}}\pi^{\frac{2q-1}{4}}(q+1)\,q^{k-\frac{1}{2}}\sqrt{\Gamma\left(\frac{2k+1}{2}\right)} \\
&= C_{J_1}\eta.
\end{aligned}$$

Similarly, we have

$$\lim_{n\to\infty} P\left\{\frac{J_{12}}{C_{J_1}} \le \eta\right\} = 1. \tag{40}$$

And from (39) and (40),

$$\lim_{n\to\infty} P\left\{\frac{J_{11}+J_{12}}{C_{J_1}} \le \eta\left(\frac{q^2(1+\eta)\sqrt{(2k+1)(2k+3)}}{2(1-\eta)^{\frac{q+k+2}{2}}}+1\right)\right\} = 1. \tag{41}$$

By the way how η and ε are chosen, we can get from (41) that

$$\lim_{n\to\infty} P\left\{\frac{J_1}{C_{J_1}} \le \varepsilon\right\} = 1. \tag{42}$$

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

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

Next we show that

$$J_2 \xrightarrow{p} 0. \tag{43}$$

Using (36), we can write

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

where

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

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

For J_{21}, in terms of (31), we have

$$\begin{aligned}
J_{21} &= \int_{A_{2n}} \|\mathbf{z}_n\|^k\, p\left(\widehat{\boldsymbol{\vartheta}}{+}\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) \exp\left[-\frac{1}{2}\mathbf{z}_n'\left[\mathbf{I}_q - \mathbf{R}_n\left(\boldsymbol{\vartheta},\mathbf{y}\right)\right]\mathbf{z}_n\right] d\mathbf{z}_n \\
&= \int_{A_{2n}} \|\mathbf{z}_n\|^k\, p\left(\widehat{\boldsymbol{\vartheta}}{+}\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) \exp\left[\log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) - \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)\right] d\mathbf{z}_n \\
&= \int_{A_{2n}} \|\mathbf{z}_n\|^k\, p\left(\widehat{\boldsymbol{\vartheta}}{+}\Sigma_n^{1/2}\mathbf{z}_n\right) \exp\left[\log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) - \log p\left(\mathbf{y}|\boldsymbol{\vartheta}_n^0\right)\right] d\mathbf{z}_n \\
&\quad \times \exp\left[\log p\left(\mathbf{y}|\boldsymbol{\vartheta}_n^0\right) - \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)\right].
\end{aligned} \tag{44}$$

According to Lemma 3.1 in Li et al. (2017), if $\mathbf{z}_n \in A_{2n}$, $\log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}} + \Sigma^{1/2}\mathbf{z}_n\right) - \log p\left(\mathbf{y}|\boldsymbol{\vartheta}_n^0\right) < -nK\left(\delta\right)$ with probability approaching 1. It is noted that $\exp\left[\log p\left(\mathbf{y}|\boldsymbol{\vartheta}_n^0\right) - \log p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}\right)\right] \le 1$. Hence, the integral on the right-hand side of (44) is less than

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

with probability approaching 1. Then, we can have

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

$$
\begin{aligned}
&= \exp\left[-nK(\delta)\right]\int_{\Theta\backslash\mathbf{N}_0(\delta)}\left\|\boldsymbol{\Sigma}_n^{-1/2}\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\right\|^k p(\boldsymbol{\vartheta})\left|\boldsymbol{\Sigma}_n\right|^{-1/2}d\boldsymbol{\vartheta}\\
&\leq \exp\left[-nK(\delta)\right]\int_{\Theta\backslash\mathbf{N}_0(\delta)}\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^k\left\|\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right\|^k p(\boldsymbol{\vartheta})\left|\boldsymbol{\Sigma}_n\right|^{-1/2}d\boldsymbol{\vartheta}\\
&\leq \exp\left[-nK(\delta)\right]\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^k\left|\boldsymbol{\Sigma}_n\right|^{-1/2}\int_{\boldsymbol{\Theta}}\left\|\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right\|^k p(\boldsymbol{\vartheta})\,d\boldsymbol{\vartheta}\\
&\leq \exp\left[-nK(\delta)\right]\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^k\left|\boldsymbol{\Sigma}_n\right|^{-1/2}\int_{\boldsymbol{\Theta}}\left(\|\boldsymbol{\vartheta}\|+\left\|\widehat{\boldsymbol{\vartheta}}\right\|\right)^k p(\boldsymbol{\vartheta})\,d\boldsymbol{\vartheta}\\
&\leq \exp\left[-nK(\delta)\right]\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^k\left|\boldsymbol{\Sigma}_n\right|^{-1/2}\sum_{s=0}^{k}\begin{pmatrix}k\\ s\end{pmatrix}\left\|\widehat{\boldsymbol{\vartheta}}\right\|^{k-s}\int_{\boldsymbol{\Theta}}\|\boldsymbol{\vartheta}\|^s p(\boldsymbol{\vartheta})\,d\boldsymbol{\vartheta}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\exp\left[-nK(\delta)\right]\left\|\boldsymbol{\Sigma}_n^{-1/2}\right\|^k\left|\boldsymbol{\Sigma}_n\right|^{-1/2}\\
=\;&\exp\left[-nK(\delta)\right]n^{(k+1)/2}\left\|-\bar{\mathbf{H}}_n^{-1/2}\right\|^k\left|\bar{\mathbf{H}}_n\right|^{-1/2}\xrightarrow{p}0,
\end{aligned}
$$

Furthermore, $\int_{\boldsymbol{\Theta}}\|\boldsymbol{\vartheta}\|^k p(\boldsymbol{\vartheta})\,d\boldsymbol{\vartheta}<\infty$ and $\widehat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_n^0\xrightarrow{p}0$ by the Assumptions 1-8, then we have

$$
J_{21}\xrightarrow{p}0. \tag{45}
$$

For J_{22}, we can show that

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

$$\times\left\{\int_{R^q}\left[\prod_{i=1}^q 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)\right]^2(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)d\mathbf{z}_n\right\}^{1/2}$$

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

From (28), we have

$$\int_{R^q}\left(\max_i|\mathbf{z}_{ni}|\right)^{2k}(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)d\mathbf{z}_n<\infty. \tag{46}$$

It can be shown that

$$\begin{aligned}
&\int_{R^q}\left[\prod_{i=1}^q 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)\right]^2(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)d\mathbf{z}_n\\
=\;&\int_{R^q}\prod_{i=1}^q 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)d\mathbf{z}_n\\
=\;&\prod_{i=1}^q\left[\int_R 1\left(|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta\right)(2\pi)^{-1/2}\exp\left(-\frac{\mathbf{z}_{ni}^2}{2}\right)d\mathbf{z}_{ni}\right]\\
=\;&\prod_{i=1}^q\left[\int_{|\mathbf{z}_{ni}|>\sqrt{\frac{n\lambda_n}{q+1}}\delta}(2\pi)^{-1/2}\exp\left(-\frac{\mathbf{z}_{ni}^2}{2}\right)d\mathbf{z}_{ni}\right]\\
\leq\;&\left(\sqrt{q+1}\frac{\exp\left(-n\lambda_n\delta^2/2(q+1)\right)}{\sqrt{n\lambda_n2\pi}\delta}\right)^q\\
=\;&2^{-\frac{q}{2}}(q+1)^{\frac{q}{2}}\left(\frac{1}{\sqrt{\pi}\delta}\right)^q(n\lambda_n)^{-\frac{q}{2}}\exp\left(-\frac{n\lambda_nq\delta^2}{q+1}\right)\xrightarrow{p}0,
\end{aligned}\tag{47}$$

where the last inequality results from

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

From (46) and (47), we have

$$J_{22}\xrightarrow{p}0. \tag{48}$$

From (45) and (48), we can get (43). And from (42) and (43), we have

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

■

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 $\varepsilon>0$,

$$\lim_{n\to\infty}P\left(\int_{A_n}\left\|\mathbf{z}_n\mathbf{z}_n'\right\|\left|p\left(\mathbf{z}_n|\mathbf{y}\right)-(2\pi)^{-q/2}\exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right)\right|d\mathbf{z}_n>\varepsilon\right)=0, \tag{49}$$

where $\|\cdot\|$ is the matrix norm for a matrix A defined as $\|A\| = \sup_{\|x\|=1} \|Ax\|$. It is because that by (49),

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

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

$$\int_{A_n} \mathbf{z}_n\mathbf{z}_n' p\left(\mathbf{z}_n|\mathbf{y}\right) d\mathbf{z}_n - \int_{A_n} \mathbf{z}_n\mathbf{z}_n' \left(2\pi\right)^{-q/2} \exp\left(-\frac{\mathbf{z}_n'\mathbf{z}_n}{2}\right) d\mathbf{z}_n \xrightarrow{p} \mathbf{0}_{q\times q}. \tag{50}$$

So from (30) we can get

$$\begin{aligned}
\int_{A_n} \mathbf{z}_n\mathbf{z}_n' p\left(\mathbf{z}_n|\mathbf{y}\right) d\mathbf{z}_n &= \int_{A_n} \mathbf{z}_n\mathbf{z}_n' p\left(\mathbf{y}\right)^{-1} \left|\boldsymbol{\Sigma}_n\right|^{1/2} p\left(\mathbf{y}|\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) p\left(\widehat{\boldsymbol{\vartheta}}+\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_n\right) d\mathbf{z}_n \\
&= \int_{\Theta} \boldsymbol{\Sigma}_n^{-1/2} \left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' \boldsymbol{\Sigma}_n^{-1/2} p\left(\mathbf{y}\right)^{-1} \left|\boldsymbol{\Sigma}_n\right|^{1/2} p\left(\mathbf{y}|\boldsymbol{\vartheta}\right) p\left(\boldsymbol{\vartheta}\right) \left|\boldsymbol{\Sigma}_n\right|^{-1/2} d\boldsymbol{\vartheta} \\
&= \int_{\boldsymbol{\Theta}} \boldsymbol{\Sigma}_n^{-1/2} \left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' \boldsymbol{\Sigma}_n^{-1/2} p\left(\boldsymbol{\vartheta}|\mathbf{y}\right) d\boldsymbol{\vartheta} \\
&= \boldsymbol{\Sigma}_n^{-1/2} E\left[\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)\left(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\vartheta}}\right)' |\mathbf{y}\right] \boldsymbol{\Sigma}_n^{-1/2},
\end{aligned} \tag{51}$$

by changing of variables. From (50) and (51), using Assumptions 1-9, we have

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

Hence, we can have that

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

Since $\left\|\mathbf{z}_n\mathbf{z}_n'\right\| \leq \left\|\mathbf{z}_n\right\|^2$, when $k=2$, the formula (29) holds so that (49) is also held. Similarly, from this Lemma with $k=1$, it is also easy to derive that $\sqrt{n}\left(\bar{\boldsymbol{\vartheta}}-\widehat{\boldsymbol{\vartheta}}\right) \xrightarrow{p} 0$.

8.3 Proof of Theorem 3.1

According to Lemma 3.2, we have

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

Hence, based on Lemma 3.2, we have

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

According to the maximum likelihood theory (White, 1996), $\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0=O_p(n^{-1/2})$ under the null hypothesis. Thus, we can show that

$$\begin{aligned}
&(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)'\left[\mathbf{V}_{\theta\theta}(\bar{\boldsymbol{\vartheta}})\right]^{-1}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[-n^{-1}\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}}_n)+o_p(n^{-1})\right]^{-1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
=\;& \sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}})+o_p(1)\right]^{-1}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
=\;& \sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}})\right]^{-1}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(1)\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \\
=\;& \sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}})\right]^{-1}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(1)\sqrt{n}O_p(n^{-1/2})\sqrt{n}O_p(n^{-1/2}) \\
=\;& \sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\left[-\bar{\mathbf{H}}_{n,\theta\theta}^{-1}(\widehat{\boldsymbol{\vartheta}})\right]^{-1}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(1) \\
=\;& \text{Wald}+o_p(1). \qquad (52)
\end{aligned}$$

Furthermore, we can simply derive that

$$\begin{aligned}
&V\left(\boldsymbol{\vartheta}_0\right)=E\left[\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)'|\mathbf{y}\right] \\
=\;& E\left[\left(\boldsymbol{\vartheta}-\bar{\boldsymbol{\vartheta}}+\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\boldsymbol{\vartheta}-\bar{\boldsymbol{\vartheta}}+\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)'|\mathbf{y}\right] \\
=\;& V\left(\bar{\boldsymbol{\vartheta}}\right)+2\left(\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\bar{\boldsymbol{\vartheta}}-\bar{\boldsymbol{\vartheta}}\right)'+\left(\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)' \\
=\;& V\left(\bar{\boldsymbol{\vartheta}}\right)+\left(\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\left(\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)', \qquad (53)
\end{aligned}$$

under the null hypothesis.

Hence, we can further prove that

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

$$
\begin{aligned}
&= \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+\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)+o_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) \\
&= q_\theta+\text{Wald}+o_p(1),
\end{aligned}
$$

form (52) and (53).

From the above derivation, it is easy to show that

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

under the null hypothesis.

8.4 Proof of Theorem 3.3

Note that

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

We have

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

$$
\begin{aligned}
&= E\left[n\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right)' \middle| \mathbf{y}, H_1\right] \\
&\quad +2E\left[n\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R\left(\widehat{\boldsymbol{\theta}}\right) - \boldsymbol{r}\right)' \middle| \mathbf{y}, H_1\right] \\
&\quad +n\left(R\left(\widehat{\boldsymbol{\theta}}\right) - \mathbf{r}\right)\left(R\left(\widehat{\boldsymbol{\theta}}\right) - \mathbf{r}\right)'. \qquad (54)
\end{aligned}
$$

By the Taylor expansion, we can show that

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

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 Θ is compact. Thus, we have

$$
\left\| \frac{\partial^2 R\left(\widetilde{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \right\| \leq M', \qquad (55)
$$

for some $0 < M' < \infty$. Furthermore, by Bayesian large-sample theory, $\sqrt{n}(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}) = O_p(1)$. Hence, from (55), we can further derive that

$$
\begin{aligned}
&\left[\sqrt{n}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)' \otimes \mathbf{I}_m\right]\frac{\partial R^2\left(\widetilde{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) \\
&= O_p(1)O(1)\frac{1}{\sqrt{n}} = O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1), \qquad (56)
\end{aligned}
$$

Since $\int \sqrt{n}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) p(\boldsymbol{\vartheta}|\mathbf{y})\, \mathrm{d}\boldsymbol{\vartheta} = \sqrt{n}(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = o_p(1)$ according to Lemma 3.2 and also $\sqrt{n}\left(R\left(\widehat{\boldsymbol{\theta}}\right) - R(\boldsymbol{\theta}_0)\right) = O_p(1)$ by using Delta method and the consistency property of MLE, the second term of (54) is

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

For the first term of (54), after integrating out the nuisance parameters, we have,

$$
E\left[n\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R(\boldsymbol{\theta}) - R\left(\widehat{\boldsymbol{\theta}}\right)\right)' \middle| \mathbf{y}, H_1\right]
$$

$$
\begin{aligned}
&= \int_{\Theta_{\theta}} n\left(R(\boldsymbol{\theta})-R\left(\widehat{\boldsymbol{\theta}}\right)\right)\left(R(\boldsymbol{\theta})-R\left(\widehat{\boldsymbol{\theta}}\right)\right)' p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta} \\
&= \frac{\partial R\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}'} \int_{\Theta_{\theta}} n\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}\right)' p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta} \frac{\partial R\left(\widehat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}} + o_p(1),
\end{aligned}
$$

by the Taylor expansion. Based on Lemma 3.2,

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

Therefore, we have

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

Since $V_{\theta\theta}\left(\bar{\boldsymbol{\vartheta}}\right) = -\frac{1}{n}\bar{\mathbf{H}}_{n,\theta\theta}^{-1}\left(\widehat{\boldsymbol{\vartheta}}\right) + o_p\left(n^{-1}\right)$, $\overline{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} + o_p\left(n^{-1/2}\right) = \boldsymbol{\theta}_0 + O_p\left(n^{-1/2}\right)$, by (56), we have

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

Finally, the third term of (54) can be expressed as

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

Therefore, under the null hypotheses, we have

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

8.5 Proof of Theorem 3.4

Let $\{\boldsymbol{\vartheta}^{[j]}, j = 1, 2, \cdots, J\}$ be the efficient random draws from $p(\boldsymbol{\vartheta}|\mathbf{y})$. Then, we have

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

$$\bar{\mathbf{v}}_1 = \bar{\bar{\boldsymbol{\theta}}} = \frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\theta}^{[j]},$$

Hence, $\widehat{\mathbf{T}}(\mathbf{y}, \boldsymbol{\theta}_0)$ in (15) can be rewritten as

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

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

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

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

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

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

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

where

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

By the Delta method,

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

The expression of the NSE for $\widehat{\mathbf{T}}\left(\mathbf{y},\mathbf{r}\right)$ can also be obtained in the similar way.

$$\begin{aligned}\widehat{\mathbf{T}}\left(\mathbf{y},\mathbf{r}\right)&=\mathbf{tr}\left[\left(\frac{\partial R\left(\bar{\bar{\boldsymbol{\theta}}}\right)}{\partial\boldsymbol{\theta}'}\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\frac{\partial R\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)}{\partial\boldsymbol{\theta}}\right)^{-1}\bar{\mathbf{V}}_{\theta}(\mathbf{r})\right]\\&=m+tr\left\{\left(R\left(\bar{\bar{\boldsymbol{\theta}}}\right)-\mathbf{r}\right)\left(R\left(\bar{\bar{\boldsymbol{\theta}}}\right)-\mathbf{r}\right)'\left(\frac{\partial R\left(\bar{\bar{\boldsymbol{\theta}}}\right)}{\partial\boldsymbol{\theta}'}\bar{\mathbf{V}}_{\theta\theta}\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)\frac{\partial R\left(\bar{\bar{\boldsymbol{\vartheta}}}\right)}{\partial\boldsymbol{\theta}}\right)^{-1}\right\}\\&=m+tr\left[\left(\bar{\mathbf{v}}_3-\mathbf{r}\right)\left(\bar{\mathbf{v}}_3-\mathbf{r}\right)'\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\right],\end{aligned}$$

where similarly,

$$\bar{\mathbf{v}}_3=R\left(\frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\theta}^{[j]}\right)=R\left(\bar{\mathbf{v}}_1\right),\ \bar{\mathbf{V}}_4=\frac{\partial R\left(\frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\theta}^{[j]}\right)}{\partial\boldsymbol{\theta}}=\frac{\partial R\left(\bar{\mathbf{v}}_1\right)}{\partial\boldsymbol{\theta}},\ \bar{\mathbf{v}}=\left(\bar{\mathbf{v}}_1',\bar{\mathbf{v}}_2'\right)'.$$

So that,

$$\begin{aligned}\frac{\partial\widehat{\mathbf{T}}\left(\mathbf{y},\mathbf{r}\right)}{\partial\bar{\mathbf{v}}}=&vec\left(I_m\right)'\left\{\left[\left(\left(\bar{\mathbf{v}}_3-\mathbf{r}\right)'\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\right)'\otimes I_m\right]\frac{\partial\bar{\mathbf{v}}_3}{\partial\bar{\mathbf{v}}_1}\frac{\partial\bar{\mathbf{v}}_1}{\partial\bar{\mathbf{v}}}\right.\\&+\left[\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\otimes\left(\bar{\mathbf{v}}_3-\mathbf{r}\right)\right]\frac{\partial\bar{\mathbf{v}}_3'}{\partial\bar{\mathbf{v}}_1}\frac{\partial\bar{\mathbf{v}}_1}{\partial\bar{\mathbf{v}}}\\&+\left[I_m\otimes\left(\bar{\mathbf{v}}_3-\mathbf{r}\right)\left(\bar{\mathbf{v}}_3-\mathbf{r}\right)'\right]\left[\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\otimes\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)^{-1}\right]\\&\left.\times\frac{\partial vec\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)}{\partial\bar{\mathbf{v}}}\right\},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial vec\left(\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)}{\partial\bar{\mathbf{v}}}=&\left(\left(\bar{\mathbf{V}}_2\bar{\mathbf{V}}_4\right)'\otimes I_m\right)\frac{\partial\bar{\mathbf{V}}_4'}{\partial\bar{\mathbf{v}}_1}\frac{\partial\bar{\mathbf{v}}_1}{\partial\bar{\mathbf{v}}}+\left(\bar{\mathbf{V}}_4\otimes\bar{\mathbf{V}}_4'\right)\frac{\partial\bar{\mathbf{V}}_2}{\partial\bar{\mathbf{v}}}\\&+\left(I_m\otimes\bar{\mathbf{V}}_4'\bar{\mathbf{V}}_2\right)\frac{\partial\bar{\mathbf{V}}_4}{\partial\bar{\mathbf{v}}_1}\frac{\partial\bar{\mathbf{v}}_1}{\partial\bar{\mathbf{v}}},\end{aligned}$$

where the derivatives of $\bar{\mathbf{V}}_4$ and $\bar{\mathbf{v}}_3$ depend on the form of the function $R\left(\boldsymbol{\theta}\right)$. By the Delta method, we have

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

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*, forthcoming.

Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica*, **55(6)**, 1465-1471.

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

Berger, J. O., and Bernardo, J. M. (1992). On the development of the reference prior method. *Bayesian statistics*, **4**, 35-60.

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

Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.

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. (2016). MCMC Confidence Sets for Identified Sets. Working paper. Cowles Foundation for Economic Research.

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

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

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, **96(453)**, 270-281.

DiCiccio, T. J., Kass, R. E., Raftery, A., and Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, **92(439)**, 903-915.

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

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

Ghosh, J. and Ramamoorthi, R. (2003). *Bayesian Nonparametrics*, Springer-Verlag, New York.

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.

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

Heyde, C. C., and Johnstone, I. M. (1979). On asymptotic posterior normality for stochastic processes. *Journal of the Royal Statistical Society Series B (Methodological)*, 184-189.

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

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(2)**, 329-366.

Koop, G. and Korobilis, D. (2009). Bayesian Multivariate Time Series, Methods for Empirical Macroeconomics, *Foundations and Trends in Econometrics*, 3(4), 267-358.

Liao, Y., and Simoni, A. (2015). Posterior properties of the support function for set inference. Working paper, University of Maryland and CREST.

Li, M. (2006). High school completion and future youth unemployment: New evidence from High School and Beyond. *Journal of Applied Econometrics*, **21(1)**, 23-53.

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

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

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

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. (2018). Integrated deviance information criterion for latent variable models. Working paper. Singapore Management University.

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

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

Norets, A. (2009). Inference in dynamic discrete choice models with serially orrelated unobserved state variables. *Econometrica*, **77(5)**, 1665-1682.

Norets, A., and Tang, X. (2013). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies*, **81(3)**, 1229-1262.

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

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

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

Robert, C. (1993). A note on Jeffreys-Lindley paradox. *Statistica Sinica*, **3**, 601-608.

Robert, C. (2001). The Bayesian choice: from decision-theoretic foundations to computational implementation, 2nd ed. Springer-Verlag New York.

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

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

Stern, S. (1997). Simulation-based estimation. *Journal of Economic Literature*, **35**, 2006- 2039.

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

White, H. (1996). *Estimation, inference and specification analysis.* Cambridge university press.

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

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

Yu, J. (2011). Simulation-based Estimation Methods for Financial Time Series Models, *Handbook of Computational Finance*, edt by Duan, J., Härdle, W., Gentle, J., 427-465, Springer.