

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Economics

School of Economics

7-2019

Improved marginal likelihood estimation via power posteriors and importance sampling

Yong LI

Nianling WANG

Jun YU

Singapore Management University, yujun@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/soe_research



Part of the [Econometrics Commons](#)

Citation

LI, Yong; WANG, Nianling; and Jun YU. Improved marginal likelihood estimation via power posteriors and importance sampling. (2019). 1-37.

Available at: https://ink.library.smu.edu.sg/soe_research/2287

This Working Paper is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylids@smu.edu.sg.

**Improved Marginal Likelihood Estimation via
Power Posteriors and Importance Sampling**

Yong Li, Nianling Wang, Jun Yu

July 2019

Paper No. 16-2019

Improved Marginal Likelihood Estimation via Power Posteriors and Importance Sampling*

Yong Li, Nianling Wang
Renmin University of China

Jun Yu
Singapore Management University

July 22, 2019

Abstract

The power-posterior method of Friel and Pettitt (2008) has been used to estimate the marginal likelihoods of competing Bayesian models. In this paper it is shown that the Bernstein-von Mises (BvM) theorem holds for the power posteriors under regularity conditions. Due to the BvM theorem, the power posteriors, when adjusted by the square root of the corresponding grid points, converge to the same normal distribution as the original posterior distribution, facilitating the implementation of importance sampling for the purpose of estimating the marginal likelihood. Unlike the power-posterior method that requires repeated posterior sampling from the power posteriors, the new method only requires the posterior output from the original posterior. Hence, it is computationally more efficient to implement. Moreover, it completely avoids the coding efforts associated with drawing samples from the power posteriors. Numerical efficiency of the proposed method is illustrated using two models in economics and finance.

JEL classification: C11, C12

Keywords: Bayes factor; Marginal likelihood; Markov Chain Monte Carlo; Model choice; Power posteriors; Importance sampling.

1 Introduction

A highly important statistical decision faced by practitioners is model comparison. In the Bayesian paradigm, the Bayes factor (BF) is arguably the most widely used Bayesian statistic for comparing models (Kass and Raftery (1995), Young and Pettit (1996)). Calculation of BFs generally requires the marginal likelihood of the data for a given model, which conducts integrations over the entire parameter space (Chan and Eisenstat, 2015).

*We would like to thank for Nial Friel for helpful discussions. Yong Li, school of Economics, Renmin University of China, Beijing, 1000872, P.R. China. Email for Yong Li: gibbsli@ruc.edu.cn. Li gratefully acknowledges the financial support of the Chinese Natural Science Fund (No.71773130). Wang gratefully acknowledges the hospitality during her research visits to Singapore Management University. Yu's research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund.

When the parameter space is of high dimension, the integrations can impose serious computational challenges.

In the literature, various MCMC-based approaches have been proposed to compute the marginal likelihood. Some excellent reviews are found in DiCiccio et al (1997) and Han and Carlin (2001). When the dimension of the parameter space is large, as is typical in latent variable models, several interesting methods have been proposed in the literature for computing BF's from the MCMC output; see, for example, Chib (1995) and Chib and Jeliazkov (2001).

In this paper, we plan to improve a method developed by Friel and Pettitt (2008) which is based on random samples from distributions proportional to the likelihood raised to a power (the so-called power posteriors). Compared with other approaches, the power-posterior approach requires very little tuning, is easy to implement, and leads to small Monte Carlo errors.

To fix ideas, given a constant $b \in [0, 1]$, Friel and Pettitt (2008) introduced the power posterior, denoted by $p(\boldsymbol{\theta}|\mathbf{y}, b)$, as

$$p(\boldsymbol{\theta}|\mathbf{y}, b) = \frac{p(\mathbf{y}|\boldsymbol{\theta})^b p(\boldsymbol{\theta})}{p(\mathbf{y}|b)}, \quad p(\mathbf{y}|b) = \int_{\boldsymbol{\Theta}} p(\mathbf{y}|\boldsymbol{\theta})^b p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (1)$$

where \mathbf{y} is data, $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function, and $p(\boldsymbol{\theta})$ is the prior distribution. With the power posterior, the (log-) marginal likelihood can be expressed as a one-dimensional integral with respect to b from 0 to 1, i.e.,

$$\ln p(\mathbf{y}) = \int_0^1 E_{\boldsymbol{\theta}|\mathbf{y}, b} \ln p(\mathbf{y}|\boldsymbol{\theta}) db = \int_0^1 \left[\int_{\boldsymbol{\Theta}} \ln p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}, b) d\boldsymbol{\theta} \right] db.$$

In practice, this integral with respect to b does not normally have a closed-form expression and numerical integrations are needed.

Friel and Pettitt (2008) suggest approximating the integral using the trapezoidal rule,

$$\ln p(\mathbf{y}) \approx \sum_{s=0}^{S-1} (b_{s+1} - b_s) \frac{E_{\boldsymbol{\theta}|\mathbf{y}, b_{s+1}} \ln p(\mathbf{y}|\boldsymbol{\theta}) + E_{\boldsymbol{\theta}|\mathbf{y}, b_s} \ln p(\mathbf{y}|\boldsymbol{\theta})}{2},$$

where $\{b_s = (s/S)^c\}_{s=0}^S$ with $c > 1$ is a grid from $[0, 1]$. Furthermore, when $E_{\boldsymbol{\theta}|\mathbf{y}, b_s} \ln p(\mathbf{y}|\boldsymbol{\theta})$ is not available analytically, we can approximate it by

$$E_{\boldsymbol{\theta}|\mathbf{y}, b_s} \ln p(\mathbf{y}|\boldsymbol{\theta}) \approx \frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y}|\boldsymbol{\theta}^{(j)}(b_s)),$$

where $\{\boldsymbol{\theta}^{(j)}(b_s)\}_{j=1}^J$ is a sequence of effective posterior draws from the power posterior $p(\boldsymbol{\theta}|\mathbf{y}, b_s)$ after discarding some burn-in samples. It can be shown that, when $S \rightarrow +\infty$

and $J \rightarrow +\infty$,

$$\sum_{s=0}^{S-1} (b_{s+1} - b_s) \frac{\frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y} | \boldsymbol{\theta}^{(j)}(b_{s+1})) + \frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y} | \boldsymbol{\theta}^{(j)}(b_s))}{2} \xrightarrow{p} \ln p(\mathbf{y}). \quad (2)$$

There are many good features in the power-posterior approach as mentioned earlier. However, there are several drawbacks in the power-posterior approach. First and foremost, sampling from the power posterior at for each grid point b_s is required. Hence, such sampling has to be repeated for S times, greatly increasing the computational cost when S is moderate or large. It is well-known that drawing MCMC samples once is often time-consuming. Repeating the MCMC drawing for a large number of times is even more time-consuming. Second, to calculate the Monte Carlo standard error of the marginal likelihood estimate, independent MCMC chains, at all grid points, have to be obtained. As a result, the computational cost would inevitably increase sharply. Third, for many standard models with regular distributions, the power posteriors may lead to non-standard distributions so that standard Bayesian software such as WinBUGS (Spiegelhalter et al, 2003) is difficult to use.

To overcome these disadvantages in the power posteriors, in the present paper, we propose a novel approach to estimate the marginal likelihood by extending the idea of the power posteriors. The theoretical underpinning of the proposed approach is the Bernstein-von Mises (BvM) theorem that we manage to develop for the power posteriors. Due to the BvM theorem, we show that the power posteriors, when adjusted by the square root of the grid points, have the same asymptotic normal distribution as the original posterior distribution. This property suggests that we can use the original posterior distribution, adjusted by a simple linear transformation, to design a proposal distribution for importance sampling. After that, via the self-normalized importance sampling technique, an estimate of the marginal likelihood is obtained. The new method avoids the need to make random draws from the power posterior at any grid point. Moreover, the new method completely avoids coding efforts to draw random samples from the power posteriors.

The rest of this paper is organized as follows. Section 2 reviews the power-posterior approach of Friel and Pettitt (2008). Section 3 establishes the BvM theorem for power posteriors and introduces the new approach to estimate the marginal likelihood. In Section 4, we compare the proposed method with the power-posterior method in terms of accuracy and computational efficiency using two examples. Section 5 concludes the paper. Appendix collects the proof of theoretic results in the paper.

2 A Review of Power Posteriors

In this section, we review the idea of the power posteriors of Friel and Pettitt (2008). Assume $\mathbf{y} = (y_1, \dots, y_n)$ is our data with n observations. Let $p(\mathbf{y}|\boldsymbol{\theta})$ be the likelihood function of a parametric model. Let $\boldsymbol{\theta} \in \Theta$ where Θ is the parameter space. Let $p(\boldsymbol{\theta})$ be an informative prior distribution for $\boldsymbol{\theta}$.

According to Friel and Pettitt (2008), for a prior $p(\boldsymbol{\theta})$ and any $b \in [0, 1]$, the power posterior and corresponding marginal likelihood can be defined as in (1). It is easy to see

$$p(\mathbf{y}|1) = \int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} = p(\mathbf{y}), \quad p(\mathbf{y}|0) = \int p(\boldsymbol{\theta})d\boldsymbol{\theta} = 1. \quad (3)$$

It can be shown that the first derivative of the power marginal likelihood $p(\mathbf{y}|b)$ (denoted by $\mathcal{U}(b)$) is

$$\begin{aligned} \mathcal{U}(b) &:= \frac{\partial \ln p(\mathbf{y}|b)}{\partial b} = \frac{1}{p(\mathbf{y}|b)} \frac{\partial p(\mathbf{y}|b)}{\partial b} = \int_{\Theta} \frac{\partial \ln p^b(\mathbf{y}|\boldsymbol{\theta})}{\partial b} \frac{p(\mathbf{y}|\boldsymbol{\theta})^b p(\boldsymbol{\theta})}{p(\mathbf{y}|b)} d\boldsymbol{\theta} \\ &= \int_{\Theta} \ln p(\mathbf{y}|\boldsymbol{\theta}) \frac{p(\mathbf{y}|\boldsymbol{\theta})^b p(\boldsymbol{\theta})}{p(\mathbf{y}|b)} d\boldsymbol{\theta} = \int_{\Theta} \ln p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}, b) d\boldsymbol{\theta} = E_{\boldsymbol{\theta}|\mathbf{y}, b} \ln p(\mathbf{y}|\boldsymbol{\theta}). \end{aligned} \quad (4)$$

Based on (3) and (4), we can recover the integral from the first-order derivative as

$$\ln p(\mathbf{y}) = \ln p(\mathbf{y}) - 0 = \int_0^1 \mathcal{U}(b)db = \int_0^1 E_{\boldsymbol{\theta}|\mathbf{y}, b} \ln p(\mathbf{y}|\boldsymbol{\theta})db. \quad (5)$$

Equation (5) suggests a powerful approach to estimating the marginal likelihood via the power posteriors as shown in Friel and Pettitt (2008).

In many cases, the integral $\int_0^1 \mathcal{U}(b)db$ does not have an analytical solution. Friel and Pettitt (2008) proposed to numerically approximate it using the trapezoidal rule. In particular, based on the grid $\{b_s = (s/S)^c\}_{s=0}^S$ with $c > 1$, $\ln p(\mathbf{y})$ is approximated by

$$\ln p(\mathbf{y}) \approx \sum_{s=0}^{S-1} (b_{s+1} - b_s) \frac{\mathcal{U}(b_{s+1}) + \mathcal{U}(b_s)}{2}. \quad (6)$$

Clearly, as $S \rightarrow +\infty$,

$$\sum_{s=0}^{S-1} (b_{s+1} - b_s) \frac{\mathcal{U}(b_{s+1}) + \mathcal{U}(b_s)}{2} \rightarrow \ln p(\mathbf{y}). \quad (7)$$

Furthermore, since $\mathcal{U}(b_s) = E_{\boldsymbol{\theta}|\mathbf{y}, b_s} [\ln p(\mathbf{y}|\boldsymbol{\theta})]$ does not have an analytical expression in most cases, we can approximate it via

$$\mathcal{U}(b_s) = E_{\boldsymbol{\theta}|\mathbf{y}, b_s} \ln p(\mathbf{y}|\boldsymbol{\theta}) \approx \frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y}|\boldsymbol{\theta}^{(j)}(b_s)), \quad (8)$$

where $\{\boldsymbol{\theta}^{(j)}(b_s)\}_{j=1}^J$ are effective random draws from the power posterior $p(\boldsymbol{\theta}|\mathbf{y}, b_s)$. By the law of large numbers for ergodic sequences, for any b_s , as $J \rightarrow +\infty$,

$$\frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y}|\boldsymbol{\theta}^{(j)}(b_s)) \xrightarrow{p} E_{\boldsymbol{\theta}|\mathbf{y}, b_s} \ln p(\mathbf{y}|\boldsymbol{\theta}). \quad (9)$$

Combining (7) and (9), as $S \rightarrow +\infty$ and $J \rightarrow +\infty$, we get (2).

In the present paper, we name the marginal likelihood estimation approach mentioned above as the FP algorithm and it can be summarized as follows:

FP Algorithm

1. Choose a grid $\{b_s = (s/S)^c\}_{s=0}^S$ with $c > 1$.
2. For each b_s , draw J random samples $\{\boldsymbol{\theta}^{(j)}(b_s)\}_{j=1}^J$ (such as MCMC samples) from the power posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}, b_s)$.
3. For each b_s , evaluate the integration $\mathcal{U}(b_s)$ based on Equation (8).
4. Using the trapezoidal rule, the marginal likelihood is estimated by Equation (6).

To check reliability of the marginal likelihood estimate, one can compute the Monte Carlo standard error (MCSE). To do so, Friel and Pettitt (2008) suggested running the FP algorithm independently for at least 100 times. To estimate the log-marginal likelihood (denoted by LML), let

$$\widehat{\text{LML}}_r = \sum_{s=0}^{S-1} (b_{s+1} - b_s) \frac{\frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y}|\boldsymbol{\theta}^{(r,j)}(b_{s+1})) + \frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y}|\boldsymbol{\theta}^{(r,j)}(b_s))}{2}, \quad (10)$$

where $\{\boldsymbol{\theta}^{(r,j)}(b_s)\}_{j=1}^J$ are effective random draws from the power posterior $p(\boldsymbol{\theta}|\mathbf{y}, b_s)$ independently across r , where $R \geq 100$. Then, we can calculate the MCSE as

$$\text{MCSE}_{FP} = \frac{1}{R} \sum_{r=1}^R \left(\widehat{\text{LML}}_r - \overline{\text{LML}} \right)^2, \text{ where } \overline{\text{LML}} = \frac{1}{R} \sum_{r=1}^R \widehat{\text{LML}}_r. \quad (11)$$

Remark 2.1 *The power-posterior method requires repeated sampling from the power posteriors corresponding to all grid points $\{b_s\}_{s=0}^S = (s/S)^c$. It is well-known that MCMC sampling is time-consuming. Obtaining MCMC samples $S + 1$ times makes it time-consuming to estimate the marginal likelihood. From Equation (11), it is easy to see that MCSE is even more time-consuming to obtain. Parallel computing can be helpful, as shown in Hoehna et al (2017).*

Remark 2.2 *The power posteriors of most models lead to non-standard distributions. As a result, it may be impossible to use standard distributions in software such as WinBUGS to obtain MCMC samples. In this case, researchers must first define model-specific new distributions and then draw MCMC samples.*

3 New Approach

3.1 The BvM theorem for power posteriors

Before introducing the new method, we first establish the BvM theorem for the power posteriors. Under some regularity condition, according to the standard BvM theorem, the posterior distribution is asymptotically independent of the prior distribution and converges to a normal distribution,

$$\sqrt{n} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \right) | \mathbf{y} \xrightarrow{d} N(0, n \Sigma_n),$$

where $\Sigma_n = \left(-\frac{\partial^2 \ln p(\mathbf{y} | \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1}$ and $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$; see Gelman et al (2004) and Schervish (2012) for details about the BvM theorem.

In this section, we extend the BvM theorem to cover the power posteriors. To do so, we need to impose some regularity conditions, similar to those in Schervish (2012). Let

$$\mathcal{L}(\boldsymbol{\theta}) = \ln p(\mathbf{y} | \boldsymbol{\theta}), \dot{\mathcal{L}}(\boldsymbol{\theta}) = \frac{\partial \ln p(\mathbf{y} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \ddot{\mathcal{L}}(\boldsymbol{\theta}) = \frac{\partial^2 \ln p(\mathbf{y} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

Assumption 1: $\boldsymbol{\Theta} \subseteq \mathbb{R}^q$ for some finite q .

Assumption 2: Let $\boldsymbol{\theta}_n^0$ be the quasi-true value that minimizes the Kullback-Leibler (KL) loss between the DGP and the candidate model

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

where $\{\boldsymbol{\theta}_n^0\}$ is the sequence of minimizers interior to $\boldsymbol{\Theta}$ uniformly in n .

Assumption 3: The prior distribution, $p(\boldsymbol{\theta})$, is positive and continuous at $\boldsymbol{\theta}_n^0$. Furthermore, it is second-times continuously differentiable, $p(\boldsymbol{\theta}_n^0) > 0$ and $\int \|\boldsymbol{\vartheta}\|^2 p(\boldsymbol{\vartheta}) d\boldsymbol{\vartheta} < \infty$.

Assumption 4: There exists a neighborhood $N_0 \subseteq \boldsymbol{\Theta}$ of $\boldsymbol{\theta}_n^0$ and $\mathcal{L}(\boldsymbol{\theta})$ is twice continuously differentiable with respect to all coordinates of $\boldsymbol{\theta}$ in this neighborhood and $\ddot{\mathcal{L}}(\boldsymbol{\theta}_n^0) = O_p(1)$.

Assumption 5: The largest eigenvalues of Σ_n converges to zero with probability approaching one.

Assumption 6: For $\delta > 0$ and $N_0(\delta) \subseteq \boldsymbol{\Theta}$, there exists $K(\delta) > 0$ such that

$$\lim_{n \rightarrow \infty} P_{\boldsymbol{\theta}_n^0} \left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus N_0(\delta)} \lambda_n [\mathcal{L}(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta}_n^0)] < -K(\delta) \right) = 1,$$

where $N_0(\delta)$ is an open ball of radius δ around $\boldsymbol{\theta}_n^0$ and λ_n is the smallest eigenvalues of Σ_n .

Assumption 7: For each $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that

$$\lim_{n \rightarrow \infty} P_{\boldsymbol{\theta}_n^0} \left(\sup_{\boldsymbol{\theta} \in N_0(\delta(\epsilon)), \|\boldsymbol{\gamma}\|=1} \left| 1 + \boldsymbol{\gamma}' \Sigma_n^{\frac{1}{2}} \ddot{\mathcal{L}}(\boldsymbol{\theta}) \Sigma_n^{\frac{1}{2}} \boldsymbol{\gamma} \right| < \epsilon \right) = 1.$$

Remark 3.1 Under these assumptions, Schervish (2012) established the BvM theorem to show that the posterior distribution converges to a normal distribution with the MLE as its mean and Σ_n as its covariance. Note that we have changed Assumption 2 of Schervish (2012) by allowing the quasi-true value $\boldsymbol{\theta}_n^0$ to be dependent on the sample size. This change makes the assumption more reasonable for dependent data. Other regularity conditions for establishing the BvM theorem are possible; see, for example, Chen (1985), Ghosh and Ramamoorthi (2003). Assumption 3 is to ensure that the first and second moments of the posterior distribution exist. More details about these regularity conditions, one can see Schervish (2012) and Li et al (2018).

Let $\boldsymbol{\theta}_b$ be the parameter in the power posterior distribution $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$ so that we distinguish it from the parameter $\boldsymbol{\theta}$ in the original posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. That is,

$$p(\boldsymbol{\theta}_b|\mathbf{y}, b) = \frac{p(\mathbf{y}|\boldsymbol{\theta}_b)^b p(\boldsymbol{\theta}_b)}{p(\mathbf{y}|b)}, p(\mathbf{y}|b) = \int p(\mathbf{y}|\boldsymbol{\theta}_b)^b p(\boldsymbol{\theta}_b) d\boldsymbol{\theta}_b.$$

The BvM theorem for the power posteriors is given in the following theorem.

Theorem 3.1 For any constant $b \in (0, 1]$, let $p(\mathbf{y}|\boldsymbol{\theta}_b)$ be the statistical model corresponding to the power posterior $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$. Under Assumptions 1-7, we have, as $n \rightarrow +\infty$,

$$\sqrt{n}\sqrt{b} \left(\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b \right) | \mathbf{y}, b \xrightarrow{d} N(0, n\Sigma_n),$$

where $\hat{\boldsymbol{\theta}}_b$ is the MLE of $\boldsymbol{\theta}_b$ and, hence, $\hat{\boldsymbol{\theta}}_b = \hat{\boldsymbol{\theta}}$.

Remark 3.2 Let $\bar{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$ be the posterior mean of $\boldsymbol{\theta}$. Under Assumptions 1-7, based on Li et al (2018), we have

$$\begin{aligned} \bar{\boldsymbol{\theta}} &= E[\boldsymbol{\theta}|\mathbf{y}] = \hat{\boldsymbol{\theta}} + o_p(n^{-1/2}), \\ \mathbf{V}(\hat{\boldsymbol{\theta}}) &= E[(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' | \mathbf{y}] = - \left[\frac{\partial^2 \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} + o_p(n^{-1}). \end{aligned} \quad (12)$$

Hence, it is easy to show that, given \mathbf{y} and $b \in (0, 1]$, we have

$$\begin{aligned} \sqrt{n}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) &= \sqrt{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) = \sqrt{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + o_p(1) \xrightarrow{d} N(0, n\Sigma_n), \\ \sqrt{n}\sqrt{b}(\boldsymbol{\theta}_b - \bar{\boldsymbol{\theta}}) &= \sqrt{n}\sqrt{b}(\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b) + \sqrt{n}\sqrt{b}(\hat{\boldsymbol{\theta}}_b - \bar{\boldsymbol{\theta}}) \xrightarrow{d} N(0, n\Sigma_n). \end{aligned}$$

Remark 3.3 *The BvM theorem suggests that the power posterior converges to the normal distribution and, when adjusted by square root of corresponding grid point, converges to the same normal distribution as the original posterior distribution. According to Remark 3.2, $\sqrt{n}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})$ and $\sqrt{n}\sqrt{b}(\boldsymbol{\theta}_b - \bar{\boldsymbol{\theta}})$ share the same asymptotic distribution, $N(0, n\boldsymbol{\Sigma}_n)$. Hence, a natural idea to approximate the power posterior $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$ is to make a linear transformation of $\boldsymbol{\theta}|\mathbf{y}$. In the next subsection, based on the importance sampling technique, a new approach that explores this relationship to estimate the marginal likelihood will be introduced.*

3.2 The new approach

For any $b \in (0, 1]$, we propose the following linear transformation

$$\boldsymbol{\theta}_b = \frac{1}{\sqrt{b}}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) + \bar{\boldsymbol{\theta}}. \quad (13)$$

That is, $\sqrt{b}(\boldsymbol{\theta}_b - \bar{\boldsymbol{\theta}}) + \bar{\boldsymbol{\theta}} = \boldsymbol{\theta}$. Based on this linear transformation, the probability density function of $\boldsymbol{\theta}_b$ conditional on \mathbf{y} and b , denoted by $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$, can be expressed as

$$p_A(\boldsymbol{\theta}_b|\mathbf{y}, b) = p(\boldsymbol{\theta}|\mathbf{y})\sqrt{b} = \sqrt{b} \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} = \frac{\sqrt{b}p(\mathbf{y}|\sqrt{b}(\boldsymbol{\theta}_b - \bar{\boldsymbol{\theta}}) + \bar{\boldsymbol{\theta}})p(\sqrt{b}(\boldsymbol{\theta}_b - \bar{\boldsymbol{\theta}}) + \bar{\boldsymbol{\theta}})}{p(\mathbf{y})}.$$

According to the BvM theorem, $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$ converges to the same normal distribution as $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$. Hence, $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$ provides a good approximation to $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$ when n is large. Random samples from $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$ can serve as a good approximation to random samples from $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$.

To ensure the transformed parameter $\boldsymbol{\theta}_b$ is in the same parameter space as original parameter $\boldsymbol{\theta}$, we first impose the following assumption and later we relax it.

Assumption 8: For any positive constant c_1 and q -dimensional vector C_1 , we assume $c_1\boldsymbol{\theta} + C_1 \in \boldsymbol{\Theta}$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Based on the importance sampling technique (Geweke, 1989), we can get

$$\begin{aligned} E_{\boldsymbol{\theta}_b|\mathbf{y}, b} \ln p(\mathbf{y}|\boldsymbol{\theta}_b) &= \int_{\boldsymbol{\Theta}} \ln p(\mathbf{y}|\boldsymbol{\theta}_b) p(\boldsymbol{\theta}_b|\mathbf{y}, b) d\boldsymbol{\theta}_b \\ &= \int_{\boldsymbol{\Theta}} \ln p(\mathbf{y}|\boldsymbol{\theta}_b) \frac{p(\boldsymbol{\theta}_b|\mathbf{y}, b)}{p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)} p_A(\boldsymbol{\theta}_b|\mathbf{y}, b) d\boldsymbol{\theta}_b \\ &= \int_{\boldsymbol{\Theta}} \ln p(\mathbf{y}|\boldsymbol{\theta}_b) w(\boldsymbol{\theta}_b) p_A(\boldsymbol{\theta}_b|\mathbf{y}, b) d\boldsymbol{\theta}_b, \end{aligned} \quad (14)$$

where

$$w(\boldsymbol{\theta}_b) = \frac{p(\boldsymbol{\theta}_b|\mathbf{y}, b)}{p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)} = \frac{\frac{p(\mathbf{y}|\boldsymbol{\theta}_b)^b p(\boldsymbol{\theta}_b)}{p(\mathbf{y}|b)}}{\frac{\sqrt{b}p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}} = \frac{p(\mathbf{y}|\boldsymbol{\theta}_b)^b p(\boldsymbol{\theta}_b)}{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})} \frac{p(\mathbf{y})}{\sqrt{b}p(\mathbf{y}|b)}.$$

Let $\{\boldsymbol{\theta}^{(j)}\}_{j=1}^J$ be some effective random samples generated from this posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. To generate random samples from $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$, we can simply use the relation $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)/\sqrt{b} = p(\boldsymbol{\theta}|\mathbf{y})$, namely,

$$\boldsymbol{\theta}_b^{(j)} = \frac{1}{\sqrt{b}} \left(\boldsymbol{\theta}^{(j)} - \bar{\boldsymbol{\theta}} \right) + \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}} \approx \frac{1}{J} \sum_{j=1}^J \boldsymbol{\theta}^{(j)}.$$

Clearly $\{\boldsymbol{\theta}_b^{(j)}\}_{j=1}^J$ are effective random samples from $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$ and can be regarded as effective random samples from $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$ when n is moderate or large.

Based on Theorem 3.1 and in the spirit of importance sampling, we can estimate $\mathcal{U}(b)$ as:

$$\mathcal{U}(b) = \int_{\boldsymbol{\Theta}} \ln p(\mathbf{y}|\boldsymbol{\theta}_b) w(\boldsymbol{\theta}_b) p_A(\boldsymbol{\theta}_b|\mathbf{y}, b) d\boldsymbol{\theta}_b \approx \frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y}|\boldsymbol{\theta}_b^{(j)}) w(\boldsymbol{\theta}_b^{(j)}). \quad (15)$$

Since $w(\boldsymbol{\theta}_b^{(j)})$ involves some unknown constants, based on the self-normalized importance sampling technique, we can have

$$\begin{aligned} \hat{W}(\boldsymbol{\theta}_b^{(j)}) &= \frac{w(\boldsymbol{\theta}_b^{(j)})}{\sum_{j=1}^J w(\boldsymbol{\theta}_b^{(j)})} = \frac{\frac{p(\mathbf{y}|\boldsymbol{\theta}_b^{(j)})^b p(\boldsymbol{\theta}_b^{(j)})}{p(\mathbf{y}|\boldsymbol{\theta}^{(j)}) p(\boldsymbol{\theta}^{(j)})}}{\sum_{j=1}^J \frac{p(\mathbf{y}|\boldsymbol{\theta}_b^{(j)})^b p(\boldsymbol{\theta}_b^{(j)})}{p(\mathbf{y}|\boldsymbol{\theta}^{(j)}) p(\boldsymbol{\theta}^{(j)})}} \\ &= \frac{\exp \left\{ b \ln p(\mathbf{y}|\boldsymbol{\theta}_b^{(j)}) - \ln p(\mathbf{y}|\boldsymbol{\theta}^{(j)}) + \ln p(\boldsymbol{\theta}_b^{(j)}) - \ln p(\boldsymbol{\theta}^{(j)}) \right\}}{\sum_{j=1}^J \exp \left\{ b \ln p(\mathbf{y}|\boldsymbol{\theta}_b^{(j)}) - \ln p(\mathbf{y}|\boldsymbol{\theta}^{(j)}) + \ln p(\boldsymbol{\theta}_b^{(j)}) - \ln p(\boldsymbol{\theta}^{(j)}) \right\}}. \end{aligned} \quad (16)$$

Then, we can get that a consistent estimate of $\mathcal{U}(b)$, denoted as $\hat{\mathcal{U}}_{LWY}(b)$, as

$$\hat{\mathcal{U}}_{LWY}(b) = \sum_{j=1}^J \ln p(\mathbf{y}|\boldsymbol{\theta}_b^{(j)}) \hat{W}(\boldsymbol{\theta}_b^{(j)}). \quad (17)$$

Remark 3.4 Due to the BvM theorem, $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$ provides a good approximation to the power posterior $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$ for any b . Instead of using $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$ to replace $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$ directly, we use $p_A(\boldsymbol{\theta}_b|\mathbf{y}, b)$ as a proposal distribution for importance sampling. Hence, our proposed approach does not require $n \rightarrow +\infty$.

Remark 3.5 Based on the simple linear transformation given in (13), the proposed approach only requires $\{\boldsymbol{\theta}^{(j)}\}_{j=1}^J$ which are effective random samples generated from the original posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. There is no need to draw MCMC samples from the power posteriors. Hence, our method reduces computational cost. Moreover, coding efforts associated with drawing MCMC samples from the power posterior is completely avoided.

Remark 3.6 *The BvM theorem is a large sample theory and the Gaussian approximation works better when $n \times b$ takes a larger value. When the grid point b is very small, for example $b \leq 1/n$, we should not use the Gaussian approximation. Since such b is close enough to zero, the prior distribution provides a good approximation to the power posterior $p(\boldsymbol{\theta}|\mathbf{y}, b)$. In this case, the prior distribution can be used as the proposal distribution. In particular, we can draw J samples $\{\boldsymbol{\theta}^{(0,1)}, \boldsymbol{\theta}^{(0,2)}, \dots, \boldsymbol{\theta}^{(0,J)}\}$ from the prior distribution $p(\boldsymbol{\theta})$. In this case, using the self-normalized importance sampling technique, we get*

$$\mathcal{U}(b) = \int_{\boldsymbol{\Theta}} \ln p(\mathbf{y}|\boldsymbol{\theta}_b) w_0(\boldsymbol{\theta}_b) p(\boldsymbol{\theta}_b) d\boldsymbol{\theta}_b \approx \sum_{j=1}^J \ln p(\mathbf{y}|\boldsymbol{\theta}_b^{(j)}) \hat{W}_0(\boldsymbol{\theta}_b^{(0,j)}), \quad (18)$$

where

$$w_0(\boldsymbol{\theta}_b) = \frac{p(\boldsymbol{\theta}_b|\mathbf{y}, b)}{p(\boldsymbol{\theta}_b)} = \frac{\frac{p(\mathbf{y}|\boldsymbol{\theta}_b)^b p(\boldsymbol{\theta}_b)}{p(\mathbf{y}|b)}}{p(\boldsymbol{\theta}_b)} = \frac{p(\mathbf{y}|\boldsymbol{\theta}_b)^b}{p(\mathbf{y}|b)},$$

$$\hat{W}_0(\boldsymbol{\theta}_b^{(0,j)}) = \frac{w_0(\boldsymbol{\theta}_b^{(0,j)})}{\sum_{j=1}^J w_0(\boldsymbol{\theta}_b^{(0,j)})} = \frac{p(\mathbf{y}|\boldsymbol{\theta}_b^{(0,j)})^b}{\sum_{j=1}^J p(\mathbf{y}|\boldsymbol{\theta}_b^{(0,j)})^b} = \frac{\exp\{b \ln p(\mathbf{y}|\boldsymbol{\theta}_b^{(0,j)})\}}{\sum_{j=1}^J \exp\{b \ln p(\mathbf{y}|\boldsymbol{\theta}_b^{(0,j)})\}}.$$

In practice, when the sample size is moderate or large, $1/n$ is small and there are very few grid points such that $n \times b \leq 1$. For example, in Example 2 of Section 4 where $n = 5823$, if c is set to 3 and $S = 20$, there is only one grid point less than $1/n$; if c is set to 3 and $S = 40$, there are two grid points less than $1/n$; if c is set to 3 and $S = 100$, there are only five grid points less than $1/n$.

Remark 3.7 *An important difference between the FP method and the proposed method is that the proposed method is based on the importance sampling approach where the proposal distribution is developed based on the BvM theorem. When the sample size is very small such that the posterior distribution is far away from the Gaussian distribution, the FP method is still a good choice to estimate the marginal likelihood.*

In the present paper, we name the proposed marginal likelihood estimation approach as the LWY algorithm and it can be summarized as follows:

LWY Algorithm

1. Choose a grid $\{b_s = (s/S)^c\}_{s=0}^S$ with $c > 1$.
2. Draw J samples $\{\boldsymbol{\theta}^{(0,1)}, \boldsymbol{\theta}^{(0,2)}, \dots, \boldsymbol{\theta}^{(0,J)}\}$ from the prior distribution $p(\boldsymbol{\theta})$.
3. When $b_s \leq 1/n$, estimate $\mathcal{U}(b_s)$ by Equation (18).

4. Draw J samples $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(J)}\}$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$.
5. When $b_s > 1/n$, by the proposed linear transformation, we get

$$\boldsymbol{\theta}_{b_s}^{(j)} = \frac{1}{\sqrt{b_s}}(\boldsymbol{\theta}^{(j)} - \bar{\boldsymbol{\theta}}) + \bar{\boldsymbol{\theta}}, \quad \bar{\boldsymbol{\theta}} \approx \frac{1}{J} \sum_{j=1}^J \boldsymbol{\theta}^{(j)}.$$

For each b_s , we evaluate $\mathcal{U}(b_s)$ by Equation (16) and (17)

6. Obtain the estimate of the marginal likelihood by Equation (6).

In practice, Assumption 8 is often violated. A simple example where Assumption 8 is too strong is the correlation coefficient parameter whose parameter space is $[-1, 1]$. If a grid point b_s is close to zero such that $1 - \frac{1}{\sqrt{b_s}} < 0$, then $\boldsymbol{\theta}_{b_s}^{(j)} = \frac{1}{\sqrt{b_s}}\boldsymbol{\theta}^{(j)} + \left(1 - \frac{1}{\sqrt{b_s}}\right)\bar{\boldsymbol{\theta}}$ may take a value outside of $[-1, 1]$. Hence, the importance sampling technique as in Equations (14) and (15) does not work any more. To deal with this difficulty, we propose to perform a parameter transformation and need the following assumption to replace Assumption 8.

Assumption 8*: Assume that there is a one-to-one monotonic transformation between the original parameter $\boldsymbol{\theta}$ and the new parameter $\boldsymbol{\phi}$ such that $\boldsymbol{\theta} = g(\boldsymbol{\phi})$. Let the inverse of the transformation be $\boldsymbol{\phi} = g^{-1}(\boldsymbol{\theta})$. Denote the parameter space of $\boldsymbol{\phi}$ by $\boldsymbol{\Phi}$. For any positive constant c_1 and q -dimensional vector C_1 , we assume $c_1\boldsymbol{\phi} + C_1 \in \boldsymbol{\Phi}$ for any $\boldsymbol{\phi} \in \boldsymbol{\Phi}$.

Under Assumption 8*, we have

$$\begin{aligned} \ln p(\mathbf{y}) &= \int_0^1 \mathcal{U}(b)db = \int_0^1 E_{\boldsymbol{\theta}|\mathbf{y},b} \ln p(\mathbf{y}|\boldsymbol{\theta})db = \int_0^1 \int_{\boldsymbol{\Theta}} \ln p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y},b)d\boldsymbol{\theta} \\ &= \int_0^1 \int_{\boldsymbol{\Phi}} \ln p_{\boldsymbol{\phi}}(\mathbf{y}|\boldsymbol{\phi})p_{\boldsymbol{\phi}}(\boldsymbol{\phi}|\mathbf{y},b)d\boldsymbol{\phi}, \end{aligned} \quad (19)$$

where $p_{\boldsymbol{\phi}}(\mathbf{y}|\boldsymbol{\phi}) := p(\mathbf{y}|g(\boldsymbol{\phi}))$ is the likelihood function of the model (expressed as a function of $\boldsymbol{\phi}$) and $p_{\boldsymbol{\phi}}(\boldsymbol{\phi}|\mathbf{y},b)$ is the power posterior of $\boldsymbol{\phi}$ which is given by

$$p_{\boldsymbol{\phi}}(\boldsymbol{\phi}|\mathbf{y},b) = \frac{p_{\boldsymbol{\phi}}(\mathbf{y}|\boldsymbol{\phi})^b p_{\boldsymbol{\phi}}(\boldsymbol{\phi})}{p(\mathbf{y})},$$

where $p_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ is the prior density of $\boldsymbol{\phi}$.

For any $b \in (0, 1]$, we can do the same simple linear transformation for $\boldsymbol{\phi}$ as for $\boldsymbol{\theta}$, i.e.,

$$\boldsymbol{\phi}_b = \frac{1}{\sqrt{b}}(\boldsymbol{\phi} - \bar{\boldsymbol{\phi}}) + \bar{\boldsymbol{\phi}}, \quad \boldsymbol{\phi} = \sqrt{b}(\boldsymbol{\phi}_b - \bar{\boldsymbol{\phi}}) + \bar{\boldsymbol{\phi}}.$$

Based on this linear transformation, the probability density function of $\boldsymbol{\phi}_b$ conditional on \mathbf{y} and b , denoted by $p_{A\boldsymbol{\phi}}(\boldsymbol{\phi}_b|\mathbf{y},b)$, can be expressed as

$$p_{A\boldsymbol{\phi}}(\boldsymbol{\phi}_b|\mathbf{y},b) = p_{\boldsymbol{\phi}}(\boldsymbol{\phi}|\mathbf{y})\sqrt{b} = \sqrt{b} \frac{p_{\boldsymbol{\phi}}(\mathbf{y}|\boldsymbol{\phi})p_{\boldsymbol{\phi}}(\boldsymbol{\phi})}{p(\mathbf{y})} = \frac{\sqrt{b}p_{\boldsymbol{\phi}}(\mathbf{y}|\sqrt{b}(\boldsymbol{\phi}_b - \bar{\boldsymbol{\phi}}) + \bar{\boldsymbol{\phi}})p_{\boldsymbol{\phi}}(\sqrt{b}(\boldsymbol{\phi}_b - \bar{\boldsymbol{\phi}}) + \bar{\boldsymbol{\phi}})}{p(\mathbf{y})}.$$

The BvM theorem continues to hold, implying that $p_{A\phi}(\phi_b|\mathbf{y}, b)$ converges to the same normal distribution as $p_\phi(\phi_b|\mathbf{y}, b)$. Hence, $p_{A\phi}(\phi_b|\mathbf{y}, b)$ provides a good approximation to $p_\phi(\phi_b|\mathbf{y}, b)$ when n is moderate or large. Random samples from $p_{A\phi}(\phi_b|\mathbf{y}, b)$ can serve as a good approximation to random samples from $p_\phi(\phi_b|\mathbf{y}, b)$. An estimate of the marginal likelihood can be obtained using the self-normalized importance sampling technique. The result is given in the following theorem.

Theorem 3.2 *Let $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(J)}\}$ be the random samples from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. Let $\boldsymbol{\phi}^{(j)} = g^{-1}(\boldsymbol{\theta}^{(j)})$ for $j = 1, \dots, J$. For a constant $b \in (0, 1]$, denote*

$$\boldsymbol{\phi}_b^{(j)} = \frac{1}{\sqrt{b}}(\boldsymbol{\phi}^{(j)} - \bar{\boldsymbol{\phi}}) + \bar{\boldsymbol{\phi}}, \quad \bar{\boldsymbol{\phi}} \approx \frac{1}{J} \sum_{j=1}^J \boldsymbol{\phi}^{(j)}.$$

Let

$$\hat{W}(\boldsymbol{\phi}_b^{(j)}) = \frac{\exp \left\{ b \ln p_\phi(\mathbf{y}|\boldsymbol{\phi}_b^{(j)}) - \ln p_\phi(\mathbf{y}|\boldsymbol{\phi}^{(j)}) + \ln p_\phi(\boldsymbol{\phi}_b^{(j)}) - \ln p_\phi(\boldsymbol{\phi}^{(j)}) \right\}}{\sum_{j=1}^J \exp \left\{ b \ln p_\phi(\mathbf{y}|\boldsymbol{\phi}_b^{(j)}) - \ln p_\phi(\mathbf{y}|\boldsymbol{\phi}^{(j)}) + \ln p_\phi(\boldsymbol{\phi}_b^{(j)}) - \ln p_\phi(\boldsymbol{\phi}^{(j)}) \right\}}, \quad (20)$$

where $p_\phi(\boldsymbol{\phi}) = p(\boldsymbol{\theta}) \left| \frac{\partial g(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} \right|$. Then, under Assumptions 1-7 and Assumption δ^* , we can get a consistent estimate of $\mathcal{U}(b)$ by $\hat{\mathcal{U}}_{LWY}^*(b)$ which is defined as,

$$\hat{\mathcal{U}}_{LWY}^*(b) = \sum_{j=1}^J \ln p(\mathbf{y}|g(\boldsymbol{\phi}_b^{(j)})) \hat{W}(\boldsymbol{\phi}_b^{(j)}). \quad (21)$$

Hence, under the parameter transformation, based on Theorem 3.2, the LWY algorithm can be revised as:

LWY* Algorithm

1. Choose a grid $\{b_s = (s/S)^c\}_{s=0}^S$ with $c > 1$.
2. Draw J samples $\{\boldsymbol{\theta}^{(0,1)}, \boldsymbol{\theta}^{(0,2)}, \dots, \boldsymbol{\theta}^{(0,J)}\}$ from the prior distribution $p(\boldsymbol{\theta})$.
3. When $b_s \leq 1/n$, estimate $\mathcal{U}(b_s)$ by Equation (18).
4. Draw J samples $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(J)}\}$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$.
5. Based on the transformation $\boldsymbol{\phi}^{(j)} = g^{-1}(\boldsymbol{\theta}^{(j)})$, we get J samples $\{\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}, \dots, \boldsymbol{\phi}^{(J)}\}$ from the posterior distribution $p_\phi(\boldsymbol{\phi}|\mathbf{y})$.

6. When $b_s > 1/n$, by the linear transformation of parameters, we get

$$\phi_{b_s}^{(j)} = \frac{1}{\sqrt{b_s}}(\phi^{(j)} - \bar{\phi}) + \bar{\phi}, \quad \bar{\phi} \approx \frac{1}{J} \sum_{j=1}^J \phi^{(j)}.$$

For each b_s , evaluate $\mathcal{U}(b_s)$ by Equations (20) and (21).

7. Obtain the estimate of the marginal likelihood by

$$\sum_{s=0}^{S-1} (b_{s+1} - b_s) \frac{\frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y} | g(\phi^{(j)}(b_{s+1}))) + \frac{1}{J} \sum_{j=1}^J \ln p(\mathbf{y} | g(\phi^{(j)}(b_s)))}{2}. \quad (22)$$

Remark 3.8 *In practice, it is fairly easy to find a transformation that satisfies Assumption 8*. For example, for the degrees of freedom parameter v in the Student t distribution which is constrained to be larger than 2, we can use the transformation $\phi = \ln(v - 2) \in \mathcal{R}$. In this case, $g(\phi) = \exp(\phi) + 2$ and $g^{-1}(v) = \ln(v - 2)$. For another example, for the correlation coefficient δ which is constrained to be in the interval of $[-1, 1]$, we can use the transformation $\phi = \tan(\frac{\pi}{2}\delta) \in \mathcal{R}$. In this case, $g(\phi) = \frac{2}{\pi} \arctan(\phi)$ and $g^{-1}(\delta) = \tan(\frac{\pi}{2}\delta)$.*

Remark 3.9 *Hoehna et al (2017) explained how to use the parallel computing technique for fast computation of the marginal likelihoods using the FP algorithm. One can also use the parallel computing technique to implement the LWY algorithm. This is because, while our method does not require MCMC sampling from the power posterior for each grid point, we need to evaluate the likelihood function and obtain the importance weights at each grid point. Obviously, these calculations can be parallelized too. In the examples discussed below, we only report the CPU time without resorting the parallel computing technique. If the parallel computing technique is used, the computing time of both algorithms can be reduced but the relative computational cost will be the same.*

4 Examples

In this section, we use two examples to evaluate and compare the performance of the proposed algorithm and the FP algorithm by calculating the mean (or bias) and the MCSE of the marginal likelihood. We also compare the computational efficiency of both algorithms. Computational efficiency is measured based on the CPU time on a common desktop with Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz 3.40GHz.

In the first example, we consider a multivariate linear regression model with Gaussian errors for which the marginal likelihood is available in closed-form. Based on the closed-form expression, we can accurately evaluate and compare the performance of the

two algorithms. We repeat both estimation procedures for 100 times and then use Equation (11) to calculate the mean (or bias) and the MCSE of the marginal likelihood. To further illustrate the computational advantage of the newly proposed algorithm, we also investigate the multivariate linear regression model with Student t errors. The t distribution complicates the likelihood function as well as the power posterior sampling. It allows us to highlight the computational efficiency of the proposed algorithm relative to the FP algorithm. To illustrate potential extra coding efforts required by the FP algorithm, we use WinBUGS to draw MCMC samples in this example. The LWY algorithm is easily implementable in WinBUGS without extra coding efforts or sampling efforts from the power posterior. Whereas, the FP algorithm requires users to code the likelihood density corresponding to the power posterior using the “zeros trick” technique (Chapter 9 in Spiegelhalter et al 2003) in WinBUGS. Compared with using existing distributions in WinBUGS, the “zeros trick” technique greatly slows down the sampling speed.

In the second example, we consider several copula models. Most copula models do not lead to standard distributions, making it difficult to use WinBUGS to obtain MCMC samples. In this paper, we use the “mcmc” package in R to obtain MCMC sample when implementing the FP and the LWY algorithms. To use this package, one only needs to specify the posterior density directly. As a result, no extra coding effort is needed to implement the FP algorithm one can conveniently raise the original likelihood to any power $b_s \in (0, 1]$. However, as will be reported below, the FP algorithm is much slower to implement than the LWY algorithm.

4.1 Linear regression models

In the first example, we use a linear regression model with multiple explanatory variables to illustrate the effectiveness of the proposed approach. The data contains sale price of 546 houses sold in Windsor, Canada in 1987. For more details about the data, one can refer to Koop (2003). We are interested in factors that can influence house prices. There are four explanatory variables, including the size, the number of bedrooms, the number of bath rooms, and the number of storeys. The following two linear models are considered:

$$\mathbf{M}_1 : y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, \dots, n.$$

$$\mathbf{M}_2 : y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i, \quad \varepsilon_i \sim t(0, \sigma^2, v), \quad i = 1, 2, \dots, n.$$

For \mathbf{M}_1 , we use the same priors as in Koop (2003), that is,

$$\beta \sim N(\beta_0, h^{-1}V_0), \quad h := \frac{1}{\sigma^2} \sim \Gamma(s, r),$$

Table 1: Bias and MCSE (in parenthesis) of the log-marginal likelihood estimates in \mathbf{M}_1

	$c = 1$		$c = 3$	
	FP	LWY	FP	LWY
$S = 20$	-494.95(4.12)	-494.95(4.14)	-1.85(0.03)	-1.84(0.17)
$S = 40$	-243.74(2.06)	-243.74(2.09)	-0.29(0.01)	-0.28(0.22)
$S = 100$	-94.07(0.82)	-94.07(0.86)	0.22(0.01)	0.23(0.17)

where β_0, V_0 are the prior mean and the prior variance of β , h is inverse of the error variance with s, r being the scale parameter and the rate parameter of a Gamma distribution. Furthermore, following Koop (2003), we set $\beta_0 = [0, 10, 5000, 10^4, 10^4]'$, $V_0 = \text{diag}(2.4, 6 \times 10^{-7}, 0.15, 0.6, 0.6)$, $s = 2.5$, $r = 6.25 \times 10^7$. In both models, h is the precision parameter which has to be greater than zero. To ensure Assumption 8, we use the transformation $\phi(h) = \ln(h) \in \mathcal{R}$.

Note that for \mathbf{M}_1 the marginal likelihood is available analytically and hence can be calculated without any error. The true marginal likelihood value is -6151. Moreover, the power posterior has a closed-form expression which is always the normal-gamma distribution for any grid point. Therefore, it is easy to directly draw from the power posteriors.

In \mathbf{M}_2 , v is the degrees of freedom parameter with $v > 2$ in the t distribution. To ensure Assumption 8, we use the transformation $\phi(v) = \ln(v - 2) \in \mathcal{R}$. We assign the same prior distribution for h as in \mathbf{M}_1 . We choose the prior distribution for $v - 2$ to be an exponential distribution, i.e., $v - 2 \sim \text{Exp}(0.05)$. The power posteriors do not have a closed-form expression for \mathbf{M}_2 .

We generate the MCMC output from the original posterior for model \mathbf{M}_2 using WinBUGS. We also generate the MCMC output from the power posterior for \mathbf{M}_2 by using the “zeros trick” technique and defining the power posterior distribution corresponding to each grid point as a new distribution in WinBUGS. For each chain, we draw 100,000 samples in total with the first 40,000 samples being discarded, and next 60,000 being kept as effective samples. We take one sample from every three samples to reduce the dependence of the chain so that $J = 20,000$. We then estimate the marginal likelihood using the FP and LWY algorithms.

For the choice of other tuning parameters, say c and S , we follow Friel and Pettitt (2008). They suggested choosing $c = 3$ or 5 and S between 20 and 100 . As $\mathcal{U}(b_s)$ involves a higher level of non-linearity as b_s is closer to zero, to calculate $\int_0^1 \mathcal{U}(b)db$, a fine grid is needed near zero. With $c > 1$, more grid points are assigned in the region near zero.

Table 1 reports the bias and the MCSEs of the (log-) marginal likelihood estimates from the two algorithms when $c = 1, 3$, and $S = 20, 40, 100$ in \mathbf{M}_1 . Table 2 reports the (log-) marginal likelihood estimates from the two algorithms (denoted by LML_{FP} and

Table 2: Log-marginal likelihood (LML) estimates for \mathbf{M}_2 with $c = 3$

	$S = 20$	$S = 40$	$S = 100$
LML_{FP}	-6514	-6513	-6513
LML_{LWY}	-6518	-6518	-6517

Table 3: CPU time (in minutes or in hours) of linear regression models

	\mathbf{M}_1		\mathbf{M}_2	
	FP	LWY	FP	LWY
$S = 20$	19.71 min	20.31 min	3.81 h	0.35 h
$S = 40$	40.25 min	39.21 min	9.12 h	0.84 h
$S = 100$	108.96 min	93.49 min	22.80 h	1.91 h

LML_{LWY}) when $c = 3$, and $S = 20, 40, 100$ in \mathbf{M}_2 . We cannot obtain the bias in \mathbf{M}_2 as the true value of the marginal likelihood is unknown in \mathbf{M}_2 .

We can see from Table 1 that both FP and LWY provide good approximations to the true value when S is moderate and $c = 3$. The MCSEs always take small values, reinforcing the finding in Friel and Pettitt (2008). When $c = 1$, the quality of the approximations is much worse, confirming the suggestion that a fine grid should be used in the regions near zero. This is the reason why we only choose $c = 3$ in the rest of the paper. In this case, for all three values of S , the two algorithms provide very similar estimates. Based on the marginal likelihood values of \mathbf{M}_1 and those of \mathbf{M}_2 , one can obtain the BF. It is evident that \mathbf{M}_1 fits the data better than \mathbf{M}_2 .

In Table 3 we report the CPU time for estimating the marginal likelihood once. Since \mathbf{M}_1 has a closed-form expression for the power posteriors, not surprisingly, there is not much computational gain in using the LWY algorithm relative to the FP algorithm as drawing from the power posteriors is easy. However, there is a substantial gain in the LWY algorithm in terms of the computational cost relative to the FP algorithm in \mathbf{M}_2 . While not reported, the LWY algorithm saves more of the CPU time if both methods are used to compute MCSEs in \mathbf{M}_2 .

4.2 Copula models

In this subsection, following Hurn et al (2019), we consider several copula models for stock returns. Unlike Hurn et al (2019) where the copula models are estimated using maximum likelihood, we estimate competing models using MCMC. For each competing model, we use the FP and LWY algorithms to estimate the marginal likelihood and then obtain the BFs to make a pair-wise comparison of nested and nonnested models.

Let r_{1t} and r_{2t} be the daily log returns at time t . Assume

$$\begin{aligned} r_{1t} &= \mu_1 + \sigma_1 z_{1t}, \\ r_{2t} &= \mu_2 + \sigma_2 z_{2t}, \end{aligned}$$

where μ_i, σ_i are the mean and the standard deviation of r_{it} for $i = 1, 2$. The joint distribution of returns is modeled by a copula function, i.e.,

$$F(r_{1t}, r_{2t}) = C(F_1(r_{1t}), F_2(r_{2t}); \delta),$$

where $F_i(\cdot)$ is the marginal distribution for r_{it} and $C(\cdot; \delta)$ is the copula function with parameter δ . Different assumptions about the marginal distribution of z_{it} and the copula function are made below, leading to different models. All competing models are fit to daily log returns on the S&P 100 and S&P 600 Indices for the period 17 August 1995 to 28 December 2018.¹

We use the “mcmc” package in R to obtain the MCMC output. It requires users to provide the kernel of the likelihood function and the prior. Recall that

$$p(\boldsymbol{\theta}|\mathbf{y}, b) \propto p(\mathbf{y}|\boldsymbol{\theta})^b p(\boldsymbol{\theta}).$$

The kernel of the target functions for the model corresponding to the power posterior, in the log form, is:

$$b \ln p(\mathbf{y}|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta}).$$

For each b , we iterate 100,000 times in total, and the first half of the chain is discarded as burn-in. For the remaining 50,000 samples, we keep one out of every five samples to reduce the dependence of the chain so that $J = 10,000$.

4.2.1 Gaussian copula normal marginals

In this model we assume $z_{1t}, z_{2t} \sim N(0, 1)$ and $C(\cdot; \delta)$ to be the Gaussian copula function. This is equivalent to assuming $(r_{1t}, r_{2t})'$ follows a bivariate normal distribution with the correlation coefficient $\delta \in [-1, 1]$. The log likelihood function at time t is:

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

where $h_i = 1/\sigma_i^2$ is the precision parameter, and $z_{it} = (r_{it} - \mu_i)h_i^{1/2}$ for $i = 1, 2$. The parameters of interest are $\boldsymbol{\theta} = (\mu_1, h_1, \mu_2, h_2, \delta)'$.

¹We have extended the sample period of the same returns from 17 August 1995 – 20 May 2011, as used in Hurn et al (2019), to 17 August 1995 – 28 December 2018.

Table 4: Posterior means and posterior standard errors of parameters for the Gaussian copula normal marginals model

Parameters	μ_1	h_1	μ_2	h_2	δ
posterior mean	0.0265	0.7058	0.0367	0.5478	0.8422
posterior sd	0.0163	0.0132	0.0186	0.0102	0.0038

Table 5: Log-marginal likelihood estimates for the Gaussian copula normal marginals model with $c = 3$

	$S = 20$	$S = 40$	$S = 100$
LML_{FP}	-15726	-15720	-15717
LML_{LWY}	-15729	-15721	-15718

To do Bayesian analysis, we assign the following prior distributions on parameters,

$$\mu_i \sim N(0, 25), h_i \sim \Gamma(0.1, 1), i = 1, 2, \text{ and } \delta \sim U[-1, 1].$$

To validate Assumption 8, we use the transformation $\phi(h_i) = \ln(h_i) \in \mathcal{R}$ and $\phi(\delta) = \tan(\frac{\pi}{2}\delta) \in \mathcal{R}$ when implementing the LWY algorithm.

The posterior means and posterior standard errors of these parameters are reported in Table 4. These estimates are reasonable. For example, the posterior mean of δ is 0.8422, suggesting that there is a strong linear relationship between the two daily returns. However, the Gaussian copula implies that there is no tail dependence between the two daily returns. The estimates of the marginal likelihood by the FP and LWY algorithms are reported in Table 5 while the CPU time for the two algorithms is reported in Table 6. It is clear that both methods provide reliable estimates. However, our method is much cheaper to be implemented computationally than the FP method, using only 10% of the CPU time.

Table 6: CPU time for the two algorithms for the Gaussian copula normal marginals model

	FP	LWY
$S = 20$	3.80 min	0.54 min
$S = 40$	8.95 min	0.89 min
$S = 100$	19.11 min	2.15 min

Table 7: Posterior means and posterior standard errors of parameters for the Gaussian copula t marginals model

Parameters	μ_1	h_1	μ_2	h_2	δ	v
Posterior mean	0.0490	1.5492	0.0618	1.0860	0.8236	3.8831
Posterior sd	0.0116	0.0401	0.0139	0.0271	0.0042	0.0957

Table 8: Log-marginal likelihood estimates for the Gaussian copula t marginals model with $c = 3$

	$S = 20$	$S = 40$	$S = 100$
LML_{FP}	-14879	-14879	-14880
LML_{LWY}	-14887	-14881	-14879

4.2.2 Gaussian copula t marginals

In this model we assume $z_{1t}, z_{2t} \sim t(0, 1, v)$ and $C(\cdot; \delta)$ to be the Gaussian copula function. The log likelihood function at time t is:

$$\ln L_t = -\frac{1}{2} \ln(1 - \delta^2) - \frac{q_{1t}^2 + q_{2t}^2 - 2\delta q_{1t}q_{2t}}{2(1 - \delta^2)} + \frac{1}{2}(q_{1t}^2 + q_{2t}^2) + \ln \left(h_1^{1/2} f(z_{1t}; v) \right) + \ln \left(h_2^{1/2} f(z_{2t}; v) \right),$$

where $\delta \in [-1, 1]$, $q_{it} = \Phi^{-1}(F(z_{it}; v))$, $z_{it} = (r_{it} - \mu_i)h_i^{1/2}$, $F(z_{it}; v)$, $f(z_{it})$ are the CDF and PDF of the t distribution with v degrees of freedom ($v > 2$), and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. The Gaussian copula t marginals model nests the Gaussian copula normal marginals model. If $v \rightarrow \infty$, the two models are the same. To validate Assumption 8, we use the transformations $\phi(h_i) = \ln(h_i) \in \mathcal{R}$, $\phi(v) = \ln(v - 2) \in \mathcal{R}$, $\phi(\delta) = \tan\left(\frac{\pi}{2}\delta\right) \in \mathcal{R}$. Since v should be larger than 2 and $\delta \in [-1, 1]$, we use an exponential prior distribution for $v - 2$ and a uniform prior for δ , i.e., $v - 2 \sim \text{Exp}(1)$, $\delta \sim U[-1, 1]$.

The likelihood function of this model is complicated than the Gaussian copula normal marginals model. It requires a longer CPU time to do the posterior sampling. For example, to sample from the posterior distribution, for the Gaussian copula normal marginals model it only takes 10 seconds, whereas for the Gaussian copula t marginals model it takes about 17 minutes. Consequently, the FP algorithm requires more CPU time to estimate the marginal likelihood of the Gaussian copula t marginals model.

The posterior means and posterior standard errors of the parameters are reported in Table 7. Again, these estimates are reasonable. For example, the posterior mean of v is 3.8831, suggesting the evidence of very heavy tails in the daily returns. The estimates of the marginal likelihood by the FP and LWY algorithms are reported in Table 8 while the CPU time for the two algorithms is reported in Table 9. Both methods provide reliable

Table 9: CPU time for the two algorithms for the Gaussian copula t marginals model

	FP	LWY
$S = 20$	6.73 h	0.91 h
$S = 40$	12.88 h	1.54 h
$S = 100$	32.75 h	3.33 h

Table 10: Posterior means and posterior standard errors of parameters for the t copula t marginals model

Parameters	μ_1	h_1	μ_2	h_2	δ	v	η
Posterior mean	0.0558	1.7318	0.0704	1.2037	0.8168	3.3382	3.6102
Posterior sd	0.0115	0.0557	0.0139	0.0364	0.0051	0.1334	0.1413

estimates. Comparing the marginal likelihood values in Table 8 and Table 5, it is clear that the Gaussian copula t marginals model fits the data much better the Gaussian copula normal marginals model. This conclusion is very reasonable given the heavy tails in the daily returns. Moreover, our method is much cheaper to implement computationally than the FP method, using only 10% of the CPU time.

4.2.3 t copula t marginals

In this model we assume $z_{1t}, z_{2t} \sim t(0, 1, v)$ and $C(\cdot; \delta, \eta)$ to be the t copula function where δ is the correlation coefficient and η captures the tail dependence. Unlike the Gaussian copula, the t copula allows for tail dependence in both tails. The log likelihood function at time t is:

$$\begin{aligned} \ln L_t = & -\ln(2\pi) - \frac{1}{2} \ln(1 - \delta^2) - \frac{\eta + 2}{2} \ln \left(1 + \frac{q_{1t}^2 + q_{2t}^2 - 2\delta q_{1t}q_{2t}}{\eta(1 - \delta^2)} \right) \\ & - \ln f(q_{1t}; \eta) - \ln f(q_{2t}; \eta) + \ln \left(f(z_{1t}; v) h_1^{1/2} \right) + \ln \left(f(z_{2t}; v) h_2^{1/2} \right), \end{aligned}$$

where $\delta \in [-1, 1]$, $q_{it} = F^{-1}(F(z_{it}; v); \eta)$, $z_{it} = (r_{it} - \mu_i)h_i^{1/2}$, $i = 1, 2$. The t copula t marginals model nests the Gaussian copula t marginals model. If $\eta \rightarrow +\infty$, the two models are the same. To validate Assumption 8, we use the transformations $\phi(h_i) = \ln(h_i) \in \mathcal{R}$, $\phi(v) = \ln(v - 2) \in \mathcal{R}$, $\phi(\eta) = \ln(\eta - 2) \in \mathcal{R}$, $\phi(\delta) = \tan(\frac{\pi}{2}\delta) \in \mathcal{R}$. For the prior distributions, we assume $v - 2, \eta - 2 \sim \text{Exp}(1)$, $\delta \sim U[-1, 1]$.

Table 11: Log-marginal likelihood estimates for the t copula t marginals model with $c = 3$

	$S = 20$	$S = 40$	$S = 100$
LML_{FP}	-14694	-14691	-14691
LML_{LWY}	-14689	-14683	-14681

Table 12: CPU time for the two algorithms for the t copula t marginals model

	FP	LWY
$S = 20$	24.45 h	3.50 h
$S = 40$	49.72 h	5.78 h
$S = 100$	120.85 h	12.20 h

The MCMC sampling from the posterior distribution is even more complicated for the t-copula t marginals model. It requires more CPU time (about 1 hour) to draw from the original posterior and the power posteriors for once. The posterior means and posterior standard errors of the parameters are reported in Table 10. Again, these estimates are reasonable. For example, the posterior mean of η is 3.6102, suggesting the evidence of strong tail dependence between the daily returns. The estimates of the marginal likelihood by the FP and LWY algorithms are reported in Table 11 while the CPU time for the two algorithms is reported in Table 12. Both methods provide reliable estimates of the marginal likelihood. Comparing the marginal likelihood values in Table 11 and Table 8, it is clear that the t copula t marginals model fits the data much better than the Gaussian copula t marginals model. This conclusion is very reasonable because there is not only a strong linear relationship but also a strong tail dependence between the two daily returns. However, our method is much cheaper to implement computationally than the FP method, using only 10% of the CPU time. Even with $S = 20$, the computational burden is a major challenge for the FP algorithm, requiring 24 hours of CPU time to run the FP algorithm once. Giving the computational cost, it is impossible to obtain the MCSE using the FP algorithm.

From Table 11, it can be seen that there is a noticeable difference between the log-marginal likelihood values obtained by the two algorithms. With the concern that the difference may be due to a reasonably small value of J being used, we increase J to 20,000, the log-marginal likelihood estimate obtained the LWY algorithm is -14695, -14688 and -14686 for $S = 20, 40, 100$ respectively. However, with the increased J , we cannot obtain the log-marginal likelihood estimate by the FP algorithm as it is too time consuming.

4.2.4 Clayton copula t marginals

In this model we assume $z_{1t}, z_{2t} \sim t(0, 1, v)$ and $C(\cdot; \delta)$ to be the Clayton copula function. The Clayton copula function is given by:

$$C(u_1, u_2; \delta) = \left(u_1^{-\delta} + u_2^{-\delta} - 1\right)^{-1/\delta}, \quad 0 < \delta < \infty,$$

where $\delta > 0$ captures the degree of left tail dependence of the two marginals. This model does not nest or is not nested by any model introduced earlier as the Clayton copula only

Table 13: Posterior means and posterior standard errors of parameters for the Clayton copula t marginals model

Parameters	μ_1	h_1	μ_2	h_2	δ	v
Posterior mean	0.1638	1.9200	0.1920	1.3491	2.2459	2.5487
Posterior sd	0.0110	0.0590	0.0134	0.0386	0.0447	0.0682

Table 14: Log-marginal likelihood estimation for the Clayton copula t marginals model with $c = 3$

	$S = 20$	$S = 40$	$S = 100$
LML_{FP}	-15279	-15280	-15280
LML_{LWY}	-15284	-15278	-15276

allows for dependence at left tails. The log likelihood function at time t is:

$$\begin{aligned} \ln L_t &= \ln(1 + \delta) - (1 + \delta)(\ln u_{1t} + \ln u_{2t}) - (2 + 1/\delta) \ln \left(u_{1t}^{-\delta} + u_{2t}^{-\delta} - 1 \right) \\ &+ \ln(f(z_{1t}; v)h_1^{1/2}) + \ln(f(z_{2t}; v)h_2^{1/2}), \end{aligned}$$

where $z_{it} = (r_{it} - \mu_i)h_i^{1/2}$ and $u_{it} = F(z_{it}; v)$ for $i = 1, 2$. To validate Assumption 8, we use the transformations $\phi(h_i) = \ln(h_i) \in R$, $\phi(v) = \ln(v - 2)$, $\phi(\delta) = \ln \delta$. As for the prior of δ , we assume $\delta \sim \Gamma(1, 1)$.

The posterior means and posterior standard errors of these parameters are reported in Table 13. Again, these estimates are reasonable. For example, the posterior mean of δ is 2.246, suggesting the evidence of strong dependence in the left tails. The estimates of the marginal likelihood by the FP and LWY algorithms are reported in Table 14 while the CPU time for the two algorithms is reported in Table 15. Both methods provide reliable estimates. Comparing the marginal likelihood values of all four models reported in Tables 14, 11, 8 and 5 (some are nested and some are not), it is clear that the t copula t marginals model fits the data much best, followed by the Gaussian copula t marginals model, then by the Clayton copula t marginals model, and finally by the Gaussian copula normal marginals model. Again, our method is much cheaper to implement computationally than the FP method, using only 10% of the CPU time.

Table 15: CPU time for the two algorithms for the Clayton copula t marginals model

	FP	LWY
$S = 20$	7.16 h	1.05 h
$S = 40$	13.86 h	1.70 h
$S = 100$	36.42 h	3.64 h

5 Concluding Remarks

In this paper, under some regularity conditions, we establish the BvM theorem for the power posteriors. Due to the BvM theorem, the power posteriors, when adjusted by the square root of the grid points, converge to the same normal distribution as the original posterior distribution. This large sample theory, therefore, allows us to improve the power-posterior method of Friel and Pettitt (2008) by providing a proposal distribution for importance sampling. Unlike the power-posterior method that requires repeated posterior sampling from the power posteriors, the new method only requires the posterior output of the original posterior. Hence, it is computationally more efficient. Moreover, for models where extra coding efforts are needed to draw MCMC samples from power posteriors, such coding efforts are completely avoided.

The accuracy of the proposed method is examined and compared with the power-posterior method in the Gaussian linear regression model where the true value of the marginal likelihood can be obtained. It suggests that the proposed method provides reliable estimates of the marginal likelihood. It performs as well as the power-posterior method of Friel and Pettitt (2008) in terms of both bias and MCSE. Comparison of computational efficiency between the proposed method against Friel and Pettitt's method is made under a linear regression model with t errors and several copula models. The comparison suggests that when a model is reasonably complicated, Friel and Pettitt's method is very time-consuming for estimating the marginal likelihood and impossible for obtaining the MCSE of the marginal likelihood estimates. Our method can reduce 90% of CPU time of Friel and Pettitt's method.

The marginal likelihood is only well-defined under proper priors. Therefore, it is important to note that, as a method that aims to estimate the marginal likelihood, our method cannot be used in connection to improper priors.

6 Appendix

6.1 Proof of Theorem 3.1

To discriminate the parameter $\boldsymbol{\theta}$ in the original posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$, let $\boldsymbol{\theta}_b$ be the parameter in the power posterior distribution $p(\boldsymbol{\theta}_b|\mathbf{y}, b)$. For the power posterior, when $b \in (0, 1]$, we know that

$$p(\boldsymbol{\theta}_b|\mathbf{y}, b) = \frac{p(\mathbf{y}|\boldsymbol{\theta}_b)^b p(\boldsymbol{\theta}_b)}{p(\mathbf{y}|b)}, p(\mathbf{y}|b) = \int p(\mathbf{y}|\boldsymbol{\theta}_b)^b p(\boldsymbol{\theta}_b) d\boldsymbol{\theta}_b,$$

and $\hat{\theta}_b$ is the MLE of $\hat{\theta}_b$ which is the solution to

$$\hat{\theta}_b = \arg \min_{\theta_b \in \Theta} \ln \left(p(\mathbf{y}|\theta_b)^b \right) = \arg \min_{\theta_b \in \Theta} b \ln p(\mathbf{y}|\theta_b) = \arg \min_{\theta \in \Theta} \ln p(\mathbf{y}|\theta).$$

Hence, $\hat{\theta}_b = \hat{\theta}$ where $\hat{\theta}$ is the MLE of $\hat{\theta}$ in the original model $p(\mathbf{y}|\hat{\theta})$, we can get that

$$\Sigma_n^{-1} = -\frac{\partial^2 \ln p(\mathbf{y}|\hat{\theta}_b)}{\partial \theta \partial \theta'} = -\frac{\partial^2 \ln p(\mathbf{y}|\hat{\theta})}{\partial \theta \partial \theta'}.$$

Let $\mathbf{z}_{nb} := (b^{-1}\Sigma_n)^{-1/2} (\theta_b - \hat{\theta}_b)$ and $A_n := \{\mathbf{z}_{nb} : \hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb} \in \Theta\}$ be the support space of \mathbf{z}_{nb} . Then, based on this transformation, the power posterior density of \mathbf{z}_{nb} , $p(\mathbf{z}_{nb}|\mathbf{y}, b)$, can be written as

$$\begin{aligned} p(\mathbf{z}_{nb}|\mathbf{y}, b) &= \frac{|b^{-1}\Sigma_n|^{1/2} p(\mathbf{y}|\theta_b)^b p(\theta_b)}{p(\mathbf{y}|b)} \\ &= \frac{|b^{-1}\Sigma_n|^{1/2} p(\mathbf{y}|\hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb}) p(\hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb})}{p(\mathbf{y}|b)}. \end{aligned} \quad (23)$$

From (23), to establish the BvM theorem for the power posterior, we only need to prove

$$\lim_{n \rightarrow \infty} P \left(\int_{A_n} \left| p(\mathbf{z}_{nb}|\mathbf{y}, b) - (2\pi)^{-q/2} \exp \left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2} \right) \right| d\mathbf{z}_{nb} < \varepsilon \right) = 1. \quad (24)$$

Based on (24), we can derive that

$$\begin{aligned} & p(\mathbf{z}_{nb}|\mathbf{y}, b) - (2\pi)^{-q/2} \exp \left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2} \right) \\ &= \frac{|b^{-1}\Sigma_n|^{1/2}}{p(\mathbf{y}|b)} p(\mathbf{y}|\hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb}) p(\hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb}) - (2\pi)^{-q/2} \exp \left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2} \right) \\ &= \frac{|b^{-1}\Sigma_n|^{1/2} p(\mathbf{y}|\hat{\theta}_b)}{p(\mathbf{y}|b)} p(\hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb}) \frac{p(\mathbf{y}|\hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb})}{p(\mathbf{y}|\hat{\theta})} \\ &\quad - (2\pi)^{-q/2} \exp \left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2} \right) \\ &= \frac{|b^{-1}\Sigma_n|^{1/2} p(\mathbf{y}|\hat{\theta}_b) p(\theta_0)}{p(\mathbf{y}|b)} \frac{p(\hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb})}{p(\theta_0)} \frac{p(\mathbf{y}|\hat{\theta}_b + (b^{-1}\Sigma_n)^{1/2} \mathbf{z}_{nb})}{p(\mathbf{y}|\hat{\theta})} \\ &\quad - (2\pi)^{-q/2} \exp \left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2} \right). \end{aligned} \quad (25)$$

To prove (25), we first prove that

$$\frac{|b^{-1}\Sigma_n|^{1/2} p(\mathbf{y}|\hat{\theta}) p(\theta_0)}{p(\mathbf{y}|b)} \xrightarrow{P} (2\pi)^{-q/2}.$$

Taking the Taylor expansion to $\ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\boldsymbol{\Sigma}_n)^{1/2}\mathbf{z}_{nb})^b$ at $\hat{\boldsymbol{\theta}}_b$, we get

$$\begin{aligned}
& \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\boldsymbol{\Sigma}_n)^{1/2}\mathbf{z}_{nb})^b = b \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\boldsymbol{\Sigma}_n)^{1/2}\mathbf{z}_{nb}) \\
&= b \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b) + \frac{1}{2}bb^{-1}\mathbf{z}_{nb}'\boldsymbol{\Sigma}_n^{1/2}\frac{\partial^2 \ln p(\mathbf{y}|\tilde{\boldsymbol{\theta}}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}_b'}\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_{nb} \\
&= b \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b) - \frac{1}{2}\mathbf{z}_{nb}'\boldsymbol{\Sigma}_n^{1/2}\left[-\frac{\partial^2 \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}_b'} - \frac{\partial^2 \ln p(\mathbf{y}|\tilde{\boldsymbol{\theta}}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}_b'} + \frac{\partial^2 \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}_b'}\right]\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_{nb} \\
&= b \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b) - \frac{1}{2}\mathbf{z}_{nb}'\boldsymbol{\Sigma}_n^{1/2}\left[\boldsymbol{\Sigma}_n^{-1} - \frac{\partial^2 \ln p(\mathbf{y}|\tilde{\boldsymbol{\theta}}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}_b'} - \boldsymbol{\Sigma}_n^{-1}\right]\boldsymbol{\Sigma}_n^{1/2}\mathbf{z}_{nb} \\
&= b \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b) - \frac{1}{2}\mathbf{z}_{nb}'\left[\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})\right]\mathbf{z}_{nb}, \tag{26}
\end{aligned}$$

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

$$\mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}) = \mathbf{I}_q + \boldsymbol{\Sigma}_n^{1/2}\frac{\partial^2 \ln p(\mathbf{y}|\tilde{\boldsymbol{\theta}}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}_b'}\boldsymbol{\Sigma}_n^{1/2},$$

with $\tilde{\boldsymbol{\theta}}_b$ lies between $\hat{\boldsymbol{\theta}}_b + (b^{-1}\boldsymbol{\Sigma}_n)^{1/2}\mathbf{z}_{nb}$ and $\hat{\boldsymbol{\theta}}_b$.

Based on the regularity conditions, we know that $\exists \delta > 0$, for any $\boldsymbol{\theta}_b$ satisfying $\|\boldsymbol{\theta}_b - \boldsymbol{\theta}_n^0\| \leq \delta$, $\boldsymbol{\theta}_b \in \mathbf{N}_0(\delta) = \{\boldsymbol{\theta}_b : \|\boldsymbol{\theta}_b - \boldsymbol{\theta}_n^0\| \leq \delta\}$. Then, we divide the support space of $\boldsymbol{\theta}$ into two parts, that is,

$$\begin{aligned}
p(\mathbf{y}|b) &= \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}_b) p(\mathbf{y}|\boldsymbol{\theta}_b)^b d\boldsymbol{\theta}_b = K = K_1 + K_2, \\
K_1 &= \int_{N_0(\delta)} p(\boldsymbol{\theta}_b) p(\mathbf{y}|\boldsymbol{\theta}_b)^b d\boldsymbol{\theta}_b, K_2 = \int_{\boldsymbol{\Theta} \setminus N_0(\delta)} p(\boldsymbol{\theta}_b) p(\mathbf{y}|\boldsymbol{\theta}_b)^b d\boldsymbol{\theta}_b. \tag{27}
\end{aligned}$$

Based on (26) and (27), we get

$$\begin{aligned}
K_1 &= \int_{N_0(\delta)} p(\boldsymbol{\theta}_b) p(\mathbf{y}|\boldsymbol{\theta}_b)^b d\boldsymbol{\theta}_b = p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b \int_{N_0(\delta)} p(\boldsymbol{\theta}_b) \exp\left[b\left(\ln p(\mathbf{y}|\boldsymbol{\theta}_b) - \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)\right)\right] d\boldsymbol{\theta}_b \\
&= p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b \int_{N_0(\delta)} p(\boldsymbol{\theta}_b) \exp\left[-\frac{1}{2}(\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b)'(b^{-1}\boldsymbol{\Sigma}_n)^{-\frac{1}{2}}\left[\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})\right](b^{-1}\boldsymbol{\Sigma}_n)^{-\frac{1}{2}}(\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b)\right] d\boldsymbol{\theta}_b.
\end{aligned}$$

For some $\eta \in (0, 1)$ and $\boldsymbol{\theta}_b \in \mathbf{N}_0(\delta)$, based on Assumption 3, we get

$$|p(\boldsymbol{\theta}_b) - p(\boldsymbol{\theta}_n^0)| \leq \eta p(\boldsymbol{\theta}_n^0), \tag{28}$$

so that

$$(1 - \eta)p(\boldsymbol{\theta}_n^0) \leq p(\boldsymbol{\theta}_b) \leq (1 + \eta)p(\boldsymbol{\theta}_n^0).$$

We further get

$$(1 - \eta)p(\boldsymbol{\theta}_n^0)p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b K_{12} < K_1 < (1 + \eta)p(\boldsymbol{\theta}_n^0)p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b K_{12},$$

where

$$\begin{aligned} K_{12} &= \int_{N_0(\delta)} \exp \left[-\frac{1}{2} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b)' (b^{-1} \boldsymbol{\Sigma}_n)^{-\frac{1}{2}} [\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})] (b^{-1} \boldsymbol{\Sigma}_n)^{-\frac{1}{2}} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b) \right] d\boldsymbol{\theta}_b \\ &= \int_{N_0(\delta)} \exp \left[-\frac{1}{2} \mathbf{z}'_{nb} [\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})] \mathbf{z}_{nb} \right] d\boldsymbol{\theta}_b. \end{aligned} \quad (29)$$

Let $\mathbf{r}_0 = \mathbf{z}_{nb} / \|\mathbf{z}_{nb}\|$, so that $\|\mathbf{r}_0\| = 1$. Then, we get

$$\mathbf{r}'_0 \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b) \mathbf{r}_0 = \mathbf{r}'_0 \mathbf{r}_0 + \mathbf{r}'_0 \boldsymbol{\Sigma}_n^{1/2} \frac{\partial^2 \ln p(\mathbf{y}|\tilde{\boldsymbol{\theta}}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}'_b} \boldsymbol{\Sigma}_n^{1/2} \mathbf{r}_0 = 1 + \mathbf{r}'_0 \boldsymbol{\Sigma}_n^{1/2} \frac{\partial^2 \ln p(\mathbf{y}|\tilde{\boldsymbol{\theta}}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}'_b} \boldsymbol{\Sigma}_n^{1/2} \mathbf{r}_0,$$

where $\tilde{\boldsymbol{\theta}}_b$ lies between $\boldsymbol{\theta}_b$ and $\hat{\boldsymbol{\theta}}_b$. Since $\hat{\boldsymbol{\theta}}_b \xrightarrow{p} \boldsymbol{\theta}_n^0$, With probability approaching 1, $\tilde{\boldsymbol{\theta}}_b \in N_0(\delta)$. Hence, $\tilde{\boldsymbol{\theta}}_1 \in N_0(\delta)$ with probability approaching 1. Furthermore, by Assumption 7, for $\boldsymbol{\theta}_b \in N_0(\delta)$, any $\eta_1 > 0$,

$$\lim_{n \rightarrow \infty} P \left(\sup_{\boldsymbol{\theta}_b \in N_0(\delta), \|\mathbf{r}_0\|=1} \left| 1 + \mathbf{r}'_0 \boldsymbol{\Sigma}_n^{1/2} \frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta}_b)}{\partial \boldsymbol{\theta}_b \partial \boldsymbol{\theta}'_b} \boldsymbol{\Sigma}_n^{1/2} \mathbf{r}_0 \right| < \eta_1 \right) = 1,$$

that is

$$\lim_{n \rightarrow \infty} P \left(\sup_{\boldsymbol{\theta}_b \in N_0(\delta), \|\mathbf{r}_0\|=1} \left| \mathbf{r}'_0 \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}) \mathbf{r}_0 \right| < \eta_1 \right) = 1.$$

It is noted that

$$\begin{aligned} K_3 &:= \exp \left[-\frac{1}{2} \mathbf{z}'_{nb} (\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})) \mathbf{z}_{nb} \right] \\ &= \exp \left[-\frac{1}{2} \|\mathbf{z}_{nb}\|^2 \mathbf{r}'_0 (\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})) \mathbf{r}_0 \right] \\ &= \exp \left[-\frac{1}{2} \|\mathbf{z}_{nb}\|^2 (1 - \mathbf{r}'_0 \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}) \mathbf{r}_0) \right]. \end{aligned}$$

With probability approaching 1, we have

$$\begin{aligned} &\int_{N_0(\delta)} \exp \left[-\frac{1+\eta_1}{2} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b)' (b^{-1} \boldsymbol{\Sigma}_n)^{-1} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b) \right] d\boldsymbol{\theta}_b = \int_{N_0(\delta)} \exp \left[-\frac{1+\eta_1}{2} \|\mathbf{z}_{nb}\|^2 \right] d\boldsymbol{\theta}_b \\ &\leq K_3 \leq \int_{N_0(\delta)} \exp \left[-\frac{1-\eta_1}{2} \|\mathbf{z}_{nb}\|^2 \right] d\boldsymbol{\theta}_b \\ &= \int_{N_0(\delta)} \exp \left[-\frac{1-\eta}{2} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b)' (b^{-1} \boldsymbol{\Sigma}_n)^{-1} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b) \right] d\boldsymbol{\theta}_b. \end{aligned}$$

Furthermore, we can derive that

$$\int_{N_0(\delta)} \exp \left[-\frac{1 \pm \eta_1}{2} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b)' (b^{-1} \boldsymbol{\Sigma}_n)^{-1} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b) \right] d\boldsymbol{\theta}_b = (2\pi)^{\frac{q}{2}} (1 \pm \eta_1)^{-\frac{q}{2}} |b^{-1} \boldsymbol{\Sigma}_n|^{1/2} \Phi(c_n),$$

where $\Phi(c_n)$ is the probability that a standard multivariate normal distribution $N_q(0, \mathbf{I}_q)$ is in c_n and

$$c_n = \{t : \hat{\boldsymbol{\theta}}_b + (1 \pm \eta_1)^{-\frac{1}{2}} \boldsymbol{\Sigma}_n^{\frac{1}{2}} t \in N_0(\delta)\}.$$

According to the regularity conditions, $\boldsymbol{\Sigma}_n^{\frac{1}{2}} t = o_p(1)$ for all t and hence $\Phi(c_n) \xrightarrow{p} 1$. Thus, with probability 1, we get

$$\left[(2\pi)^{\frac{q}{2}} \frac{|b^{-1} \boldsymbol{\Sigma}_n|^{\frac{1}{2}}}{(1 + \eta_1)^{\frac{q}{2}}} \right] \leq K_3 \leq \left[(2\pi)^{\frac{q}{2}} \frac{|b^{-1} \boldsymbol{\Sigma}_n|^{\frac{1}{2}}}{(1 - \eta)^{\frac{q}{2}}} \right].$$

Since η is any small positive constant so that we can relate it to any $\epsilon > 0$, we have

$$\lim_{n \rightarrow \infty} P \left[(2\pi)^{\frac{q}{2}} (b^{-1} \boldsymbol{\Sigma}_n)^{\frac{1}{2}} (1 - \epsilon) \leq \frac{K_1}{p(\boldsymbol{\theta}_n^0) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b} \leq (2\pi)^{\frac{q}{2}} (b^{-1} \boldsymbol{\Sigma}_n)^{\frac{1}{2}} (1 + \epsilon) \right] = 1.$$

In other words, we can show that

$$\frac{K_1}{|b^{-1} \boldsymbol{\Sigma}_n|^{\frac{1}{2}} p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b p(\boldsymbol{\theta}_n^0)} \xrightarrow{p} (2\pi)^{\frac{q}{2}}.$$

As to K_2 , we can show that

$$\begin{aligned} K_2 &= \int_{\boldsymbol{\Theta} \setminus N_0(\delta)} p(\boldsymbol{\theta}_b) p(\mathbf{y}|\boldsymbol{\theta}_b)^b d\boldsymbol{\theta}_b \\ &= p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b \int_{\boldsymbol{\Theta} \setminus N_0(\delta)} p(\boldsymbol{\theta}_b) \exp \left[b \left(\ln p(\mathbf{y}|\boldsymbol{\theta}_b) - \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b) \right) \right] d\boldsymbol{\theta}_b \\ &= p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b \exp \left[b \left(\ln p(\mathbf{y}|\boldsymbol{\theta}_n^0) - \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b) \right) \right] \times \\ &\quad \int_{\boldsymbol{\Theta} \setminus N_0(\delta)} p(\boldsymbol{\theta}_b) \exp [b (\ln p(\mathbf{y}|\boldsymbol{\theta}_b) - \ln p(\mathbf{y}|\boldsymbol{\theta}_n^0))] d\boldsymbol{\theta}_b. \end{aligned}$$

According to Assumption 6, when $\boldsymbol{\theta}_b \in \boldsymbol{\Theta} \setminus N_0(\delta)$, $\ln p(\mathbf{y}|\boldsymbol{\theta}_b) - \ln p(\mathbf{y}|\boldsymbol{\theta}_n^0) \leq -\lambda_n^{-1} K(\delta) \leq -|\boldsymbol{\Sigma}_n|^{-\frac{1}{q}} K(\delta)$ with probability approaching 1. Then, with probability approaching 1,

$$K_2 \leq p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b \exp [-b\lambda_n^{-1} K(\delta)] \int_{\boldsymbol{\Theta} \setminus N_0(\delta)} p(\boldsymbol{\theta}) d\boldsymbol{\theta} \leq p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b \exp [-b|\boldsymbol{\Sigma}_n|^{-\frac{1}{q}} K(\delta)],$$

$$\frac{K_2}{|b^{-1}\Sigma_n|^{\frac{1}{2}} p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b} \leq \frac{\exp\left[-b|\Sigma_n|^{-\frac{1}{q}} K(\delta)\right]}{|b^{-1}\Sigma_n|^{\frac{1}{2}}} \rightarrow 0.$$

Noting that $p(\mathbf{y}|b) = K_1 + K_2$, we get

$$\frac{|b^{-1}\Sigma_n|^{\frac{1}{2}} p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)^b p(\boldsymbol{\theta}_n^0)}{p(\mathbf{y}|b)} \xrightarrow{p} (2\pi)^{-\frac{q}{2}}. \quad (30)$$

Based on (24), we can further derive that

$$\begin{aligned} & p(\mathbf{z}_{nb}|\mathbf{y}, b) - (2\pi)^{-q/2} \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \\ = & \frac{|b^{-1}\Sigma_n|^{1/2} p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b) p(\boldsymbol{\theta}_n^0) p(\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb}) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb})}{p(\mathbf{y}|b) p(\boldsymbol{\theta}_n^0) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)} \\ & - (2\pi)^{-q/2} \frac{p(\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb}) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb})}{p(\boldsymbol{\theta}_n^0) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)} \\ & + (2\pi)^{-q/2} \left[\frac{p(\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb}) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb})}{p(\boldsymbol{\theta}_n^0) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)} - \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \right] \\ = & \left[\frac{|b^{-1}\Sigma_n|^{1/2} p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b) p(\boldsymbol{\theta}_n^0)}{p(\mathbf{y}|b)} - (2\pi)^{-q/2} \right] \frac{p(\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb}) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb})}{p(\boldsymbol{\theta}_n^0) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)} \\ & + (2\pi)^{-q/2} \left[\frac{p(\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb}) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb})}{p(\boldsymbol{\theta}_n^0) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)} - \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \right]. \quad (31) \end{aligned}$$

It is noted that

$$\begin{aligned} & \int_{A_n} \frac{p(\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb}) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb})}{p(\boldsymbol{\theta}_n^0) p(\mathbf{y}|\hat{\boldsymbol{\theta}}_b)} d\mathbf{z}_{nb} \\ \leq & \int_{A_n} \frac{p(\hat{\boldsymbol{\theta}}_b + (b^{-1}\Sigma_n)^{1/2}\mathbf{z}_{nb})}{p(\boldsymbol{\theta}_n^0)} d\mathbf{z}_{nb} \leq \frac{1}{p(\boldsymbol{\theta}_n^0)}. \quad (32) \end{aligned}$$

Based on (30), (31) and (32), to prove

$$\lim_{n \rightarrow \infty} P\left(\int_{A_n} \left| p(\mathbf{z}_{nb}|\mathbf{y}, b) - (2\pi)^{-q/2} \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \right| d\mathbf{z}_{nb} < \varepsilon\right) = 1$$

from (23), we only need to prove that

$$P\left(\int_{A_n} \left| \frac{p(\hat{\boldsymbol{\theta}}_b + b^{-1}\Sigma_n^{1/2}\mathbf{z}_{nb})}{p(\boldsymbol{\theta}_n^0)} \exp\left[-\frac{\mathbf{z}_{nb}'[\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})]\mathbf{z}_{nb}}{2}\right] - \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \right| d\mathbf{z}_n < \varepsilon\right) \rightarrow 1.$$

By Assumption 3, it is enough to prove

$$P \left(\int_{A_n} \left| p \left(\hat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{\frac{1}{2}} \mathbf{z}_{nb} \right) \exp \left[-\frac{\mathbf{z}'_{nb} [\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})] \mathbf{z}_{nb}}{2} \right] - p(\boldsymbol{\theta}_n^0) \exp \left(-\frac{\mathbf{z}'_{nb} \mathbf{z}_{nb}}{2} \right) \right| d\mathbf{z}_{nb} < \varepsilon \right) \rightarrow 1.$$

Let

$$A_{1n} = \left\{ \mathbf{z}_n : \hat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb} \in \mathbf{N}_0(\delta) \right\}, A_{2n} = \left\{ \mathbf{z}_{nb} : \hat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb} \in \Theta \setminus \mathbf{N}_0(\delta) \right\},$$

and

$$C_n = \left| p \left(\hat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb} \right) \exp \left[-\frac{1}{2} \mathbf{z}'_{nb} [\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})] \mathbf{z}_{nb} \right] - p(\boldsymbol{\theta}_n^0) \exp \left(-\frac{\mathbf{z}'_{nb} \mathbf{z}_{nb}}{2} \right) \right|.$$

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

$$J = \int_{A_n} C_n d\mathbf{z}_{nb} = \int_{A_{1n}} C_n d\mathbf{z}_{nb} + \int_{A_{2n}} C_n d\mathbf{z}_{nb} := J_1 + J_2.$$

In the following, we try to prove that

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

For J_1 , note that

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

where

$$C_{1n} = \left| p \left(\hat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb} \right) \right| \left| \exp \left[-\frac{1}{2} \mathbf{z}'_{nb} [\mathbf{I}_q - \mathbf{R}_n(\tilde{\boldsymbol{\theta}}_b, \mathbf{y})] \mathbf{z}_{nb} \right] - \exp \left(-\frac{\mathbf{z}'_{nb} \mathbf{z}_{nb}}{2} \right) \right|,$$

$$C_{2n} = \left| p \left(\hat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb} \right) - p(\boldsymbol{\theta}_n^0) \right| \exp \left(-\frac{\mathbf{z}'_{nb} \mathbf{z}_{nb}}{2} \right).$$

Then we have

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

where

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

From (28), we know that $\left| p\left(\widehat{\boldsymbol{\theta}}_b + b^{-1}\Sigma_n^{1/2}\mathbf{z}_{nb}\right) \right| \leq (1+\eta)p\left(\boldsymbol{\theta}_n^0\right)$. Hence, we have

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

Note that for any constant c , $|\exp(|c|) - 1| \leq \exp(|c|)|c|$. Hence, with probability 1, when $\boldsymbol{\theta} \in N_0(\delta)$, we get

$$\begin{aligned} & \left| \exp\left[-\frac{1}{2}\mathbf{z}_{nb}'\left[\mathbf{I}_q - \mathbf{R}_n\left(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}\right)\right]\mathbf{z}_{nb}\right] - \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \right| \\ &= \left| \exp\left[\frac{1}{2}\mathbf{z}_{nb}'\mathbf{R}_n\left(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}\right)\mathbf{z}_{nb}\right] - 1 \right| \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \\ &\leq \exp\left[\left|\frac{1}{2}\mathbf{z}_{nb}'\mathbf{R}_n\left(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}\right)\mathbf{z}_{nb}\right|\right] \left|\frac{1}{2}\mathbf{z}_{nb}'\mathbf{R}_n\left(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}\right)\mathbf{z}_{nb}\right| \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \\ &= \exp\left[\left|\frac{1}{2}\mathbf{z}_{nb}'\mathbf{z}_{nb}\right|\left|\mathbf{r}_0'\mathbf{R}_n\left(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}\right)\mathbf{r}_0\right|\right] \left|\frac{1}{2}\mathbf{z}_{nb}'\mathbf{z}_{nb}\right| \left|\mathbf{r}_0'\mathbf{R}_n\left(\tilde{\boldsymbol{\theta}}_b, \mathbf{y}\right)\mathbf{r}_0\right| \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \\ &\leq \frac{\eta}{2} \exp\left[\left|\frac{\eta}{2}\mathbf{z}_{nb}'\mathbf{z}_{nb}\right|\right] \left|\mathbf{z}_{nb}'\mathbf{z}_{nb}\right| \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) \\ &= \frac{\eta}{2} \|\mathbf{z}_{nb}\|^2 \exp\left(-\frac{(1-\eta)\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right). \end{aligned} \quad (33)$$

Let

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

It follows from (33) that

$$\lim_{n \rightarrow \infty} P\left\{J_{11}^* \leq \frac{\eta}{2} \int_{A_{1n}} \|\mathbf{z}_{nb}\|^2 \exp\left(-\frac{(1-\eta)\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) d\mathbf{z}_{nb}\right\} = 1. \quad (34)$$

Furthermore, we can derive that

$$\begin{aligned} & \int_{A_{1n}} \|\mathbf{z}_{nb}\|^2 \exp\left(-\frac{(1-\eta)\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) d\mathbf{z}_{nb} \\ &\leq \int_{\mathbb{R}^q} \|\mathbf{z}_{nb}\|^2 \exp\left(-\frac{1-\eta}{2}\mathbf{z}_{nb}'\mathbf{z}_{nb}\right) d\mathbf{z}_{nb} = \int_{\mathbb{R}^q} \left(\sum_{i=1}^q \mathbf{z}_{nb,i}^2\right) \exp\left(-\frac{1-\eta}{2}\mathbf{z}_{nb}'\mathbf{z}_{nb}\right) d\mathbf{z}_{nb} \\ &= q \int_{\mathbb{R}} \mathbf{z}_{nb,i}^2 \exp\left(-\frac{1-\eta}{2}\mathbf{z}_{nb,i}^2\right) d\mathbf{z}_{nb,i} = q(2\pi)^{1/2}(1-\eta)^{-3/2}, \end{aligned}$$

where $\mathbf{z}_{nb,i}$ is the i th element of \mathbf{z}_{nb} . Hence, we have

$$\lim_{n \rightarrow \infty} P\left(J_{11} \leq \eta p\left(\boldsymbol{\theta}_n^0\right) \sqrt{\pi} 2^{-1/2} q(1+\eta)^{-1/2}\right) = 1. \quad (35)$$

In the following, we deal with J_{12} . From (28), we have

$$\begin{aligned}
J_{12} &\leq \int_{A_{1n}} \left| p\left(\widehat{\boldsymbol{\theta}}_b + \Sigma_n^{1/2} \mathbf{z}_{nb}\right) - p\left(\boldsymbol{\theta}_n^0\right) \right| \exp\left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2}\right) d\mathbf{z}_{nb} \\
&\leq \eta p\left(\boldsymbol{\theta}_n^0\right) \int_{A_{1n}} \exp\left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2}\right) d\mathbf{z}_{nb} \\
&\leq \eta p\left(\boldsymbol{\theta}_n^0\right) \int_{\mathbb{R}^q} \exp\left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2}\right) d\mathbf{z}_{nb} \\
&= \eta p\left(\boldsymbol{\theta}_n^0\right) (2\pi)^{q/2} \int_{\mathbb{R}^q} (2\pi)^{-q/2} \exp\left(-\frac{\mathbf{z}_{nb}' \mathbf{z}_{nb}}{2}\right) d\mathbf{z}_{nb} \\
&= \eta p\left(\boldsymbol{\theta}_n^0\right) (2\pi)^{q/2}.
\end{aligned}$$

Similarly, we have

$$\lim_{n \rightarrow \infty} P\left\{J_{12} \leq \eta p\left(\boldsymbol{\theta}_n^0\right) (2\pi)^{q/2}\right\} = 1. \quad (36)$$

And from (35) and (36),

$$\lim_{n \rightarrow \infty} P\left\{J_1 = J_{11} + J_{12} \leq \eta p\left(\boldsymbol{\theta}_n^0\right) \left((2\pi)^{q/2} + \sqrt{\pi} 2^{-1/2} q (1 + \eta)^{-1/2}\right)\right\} = 1. \quad (37)$$

By the way of how η and ε are chosen, we get from (37) that

$$\lim_{n \rightarrow \infty} P\{J_1 \leq \varepsilon\} = 1. \quad (38)$$

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

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

Next we show that

$$J_2 \xrightarrow{P} 0. \quad (39)$$

such that

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

where

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

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

For J_{21} , in terms of (26), we have

$$\begin{aligned}
J_{21} &= \int_{A_{2n}} p\left(\widehat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb}\right) \exp\left[-\frac{1}{2} \mathbf{z}_{nb}' \left[\mathbf{I}_q - \mathbf{R}_n\left(\widehat{\boldsymbol{\theta}}_b, \mathbf{y}\right)\right] \mathbf{z}_{nb}\right] d\mathbf{z}_{nb} \\
&= \int_{A_{2n}} p\left(\widehat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb}\right) \exp\left[b\left(\ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb}\right) - \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_b\right)\right)\right] d\mathbf{z}_{nb} \\
&= \int_{A_{2n}} p\left(\widehat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb}\right) \exp\left[b\left(\ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb}\right) - \ln p\left(\mathbf{y}|\boldsymbol{\theta}_n^0\right)\right)\right] d\mathbf{z}_{nb} \\
&\quad \times \exp\left[b\left(\ln p\left(\mathbf{y}|\boldsymbol{\theta}_n^0\right) - \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_b\right)\right)\right]. \tag{40}
\end{aligned}$$

According to Assumption 6, if $\mathbf{z}_n \in A_{2n}$, $\ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb}\right) - \ln p\left(\mathbf{y}|\boldsymbol{\theta}_n^0\right) < -|\boldsymbol{\Sigma}_n|^{-\frac{1}{q}} K(\delta)$ with probability approaching 1. Furthermore, it is noted that $\exp\left[b\left(\ln p\left(\mathbf{y}|\boldsymbol{\theta}_n^0\right) - \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_b\right)\right)\right] \leq 1$. Hence, the integral on the right-hand side of (40) is less than

$$\exp\left[-b|\boldsymbol{\Sigma}_n|^{-\frac{1}{q}} K(\delta)\right] \int_{A_{2n}} p\left(\widehat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb}\right) d\mathbf{z}_{nb},$$

with probability approaching 1. Then, we can have

$$\begin{aligned}
&\exp\left[-b|\boldsymbol{\Sigma}_n|^{-\frac{1}{q}} K(\delta)\right] \int_{A_{2n}} p\left(\widehat{\boldsymbol{\theta}}_b + b^{-1} \boldsymbol{\Sigma}_n^{1/2} \mathbf{z}_{nb}\right) d\mathbf{z}_{nb} \\
&= b^{q/2} \exp\left[-b|\boldsymbol{\Sigma}_n|^{-\frac{1}{q}} K(\delta)\right] \int_{\Theta \setminus \mathbf{N}_0(\delta)} p(\boldsymbol{\theta}) |\boldsymbol{\Sigma}_n|^{-1/2} d\boldsymbol{\theta} \\
&\leq b^{q/2} \exp\left[-b|\boldsymbol{\Sigma}_n|^{-\frac{1}{q}} K(\delta)\right] \int_{\Theta \setminus \mathbf{N}_0(\delta)} p(\boldsymbol{\theta}) |\boldsymbol{\Sigma}_n|^{-1/2} d\boldsymbol{\theta}.
\end{aligned}$$

Note that

$$\exp\left[-|\boldsymbol{\Sigma}_n|^{-\frac{1}{q}} K(\delta)\right] |\boldsymbol{\Sigma}_n|^{-1/2} \xrightarrow{p} 0.$$

Furthermore, $\widehat{\boldsymbol{\theta}}_b - \boldsymbol{\theta}_n^0 \xrightarrow{p} 0$ by Assumptions 1-7. Then we have

$$J_{21} \xrightarrow{p} 0. \tag{41}$$

For J_{22} , we can show that

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

where $\mathbf{z}_{nb,i}$ is the i th element of \mathbf{z}_{nb} and λ_n is the smallest eigenvalue of $-b\frac{1}{n}\ddot{\mathcal{L}}(\hat{\boldsymbol{\theta}})$.

Furthermore, we can derive that

$$\begin{aligned}
& \int_{\cap_{i=1}^q \{|\mathbf{z}_{nb,i}| > \sqrt{\frac{n\lambda_n}{q+1}}\delta\}} (2\pi)^{-q/2} \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) d\mathbf{z}_{nb} \\
&= \int_{R^q} \prod_{i=1}^q 1\left(|\mathbf{z}_{nb,i}| > \sqrt{\frac{n\lambda_n}{q+1}}\delta\right) (2\pi)^{-q/2} \exp\left(-\frac{\mathbf{z}_{nb}'\mathbf{z}_{nb}}{2}\right) d\mathbf{z}_{nb} \\
&= \prod_{i=1}^q \left[\int_R 1\left(|\mathbf{z}_{nb,i}| > \sqrt{\frac{n\lambda_n}{q+1}}\delta\right) (2\pi)^{-1/2} \exp\left(-\frac{\mathbf{z}_{nb,i}^2}{2}\right) d\mathbf{z}_{nb,i} \right] \\
&= \prod_{i=1}^q \left[\int_{|\mathbf{z}_{nb,i}| > \sqrt{\frac{n\lambda_n}{q+1}}\delta} (2\pi)^{-1/2} \exp\left(-\frac{\mathbf{z}_{nb,i}^2}{2}\right) d\mathbf{z}_{nb,i} \right] \\
&\leq \left(\sqrt{q+1} \frac{\exp(-n\lambda_n\delta^2/2(q+1))}{\sqrt{n\lambda_n}2\pi\delta} \right)^q \\
&= 2^{-\frac{q}{2}}(q+1)^{\frac{q}{2}} \left(\frac{1}{\sqrt{\pi}\delta} \right)^q (n\lambda_n)^{-\frac{q}{2}} \exp\left(-\frac{n\lambda_n q\delta^2}{2(q+1)}\right) \xrightarrow{p} 0,
\end{aligned} \tag{42}$$

where the last inequality is due to

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

From (42), we have

$$J_{22} \xrightarrow{p} 0. \tag{43}$$

From (41) and (43), we can get (39). And from (38) and (39), we have

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

Hence, we prove that $p(\mathbf{z}_{nb}|\mathbf{y}, b)$, the posterior distribution $\mathbf{z}_{nb} \left(:= (b^{-1}\boldsymbol{\Sigma}_n)^{-1/2} (\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b) \right)$, converges to a standard multivariate normal distribution. In other words, we prove that the power posterior density of $\boldsymbol{\theta}_b$ also converges to a multivariate normal distribution, i.e.,

$$\sqrt{n}\sqrt{b}(\boldsymbol{\theta}_b - \hat{\boldsymbol{\theta}}_b) | \mathbf{y}, b \xrightarrow{d} N(0, n\boldsymbol{\Sigma}_n).$$

6.2 Proof of Theorem 3.2

Note that the samples $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(J)}\}$ are from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$. Under the parameter transformation, $\{\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}, \dots, \boldsymbol{\phi}^{(J)}\}$ with $\boldsymbol{\phi}^{(j)} = g^{-1}(\boldsymbol{\theta}^{(j)})$ are from the posterior distribution $p_\phi(\boldsymbol{\phi}|\mathbf{y})$. For any $b \in (0, 1]$, by the linear transformation

$$\boldsymbol{\phi}_b^{(j)} = \frac{1}{\sqrt{b}}(\boldsymbol{\phi}^{(j)} - \bar{\boldsymbol{\phi}}) + \bar{\boldsymbol{\phi}}, \quad \bar{\boldsymbol{\phi}} \approx \frac{1}{J} \sum_{j=1}^J \boldsymbol{\phi}^{(j)},$$

we can get the samples $\{\phi_b^{(1)}, \phi_b^{(2)}, \dots, \phi_b^{(J)}\}$ which are from the posterior distribution $p_{A\phi}(\phi_b|\mathbf{y})$.

Again, based on the parameter transformation under Assumption 8*, we get

$$\begin{aligned}\mathcal{U}(b) &= E_{\theta_b|\mathbf{y},b} \ln p(\mathbf{y}|\theta_b) = \int_{\Theta} \ln p(\mathbf{y}|\theta_b) p(\theta_b|\mathbf{y},b) d\theta_b \\ &= \int_{\Phi} \ln p_{\phi}(\mathbf{y}|\phi_b) p_{\phi}(\phi_b|\mathbf{y},b) d\phi_b \\ &= \int_{\Phi} \ln p_{\phi}(\mathbf{y}|\phi_b) \frac{p_{\phi}(\theta_b|\mathbf{y},b)}{p_{A\phi}(\phi_b|\mathbf{y},b)} p_{A\phi}(\phi_b|\mathbf{y},b) d\phi_b \\ &= \int_{\Phi} \ln p_{\phi}(\mathbf{y}|\phi_b) w_{\phi}(\phi_b) p_{A\phi}(\phi_b|\mathbf{y},b) d\phi_b,\end{aligned}$$

where

$$w_{\phi}(\phi_b) = \frac{p_{\phi}(\phi_b|\mathbf{y},b)}{p_{A\phi}(\phi_b|\mathbf{y},b)} = \frac{\frac{p_{\phi}(\mathbf{y}|\phi_b)^b p_{\phi}(\phi_b)}{p(\mathbf{y}|b)}}{\frac{\sqrt{b} p_{\phi}(\mathbf{y}|\phi) p_{\phi}(\phi)}{p(\mathbf{y})}} = \frac{p_{\phi}(\mathbf{y}|\phi_b)^b p_{\phi}(\phi_b)}{p_{\phi}(\mathbf{y}|\phi) p_{\phi}(\phi)} \frac{p(\mathbf{y})}{\sqrt{b} p(\mathbf{y}|b)}.$$

Again, based on importance sampling, we can estimate $\mathcal{U}(b)$ by

$$\mathcal{U}(b) = \int_{\Phi} \ln p_{\phi}(\mathbf{y}|\phi_b) w(\phi_b) p_{A\phi}(\phi_b|\mathbf{y},b) d\phi_b \approx \frac{1}{J} \sum_{j=1}^J \ln p_{\phi}(\mathbf{y}|\phi_b^{(j)}) \hat{w}(\phi_b^{(j)}),$$

where $\hat{w}(\phi_b^{(j)})$ involves some unknown constants. Using the normalized importance sampling technique, we can get another consistent estimate of $\mathcal{U}(b)$ given by

$$\mathcal{U}(b) \approx \hat{\mathcal{U}}_{LWY}^*(b) = \sum_{j=1}^J \ln p_{\phi}(\mathbf{y}|\phi_b^{(j)}) \hat{W}_{\phi}(\phi_b^{(j)}),$$

where

$$\begin{aligned}\hat{W}_{\phi}(\phi_b^{(j)}) &= \frac{w_{\phi}(\phi_b^{(j)})}{\sum_{j=1}^J w_{\phi}(\phi_b^{(j)})} = \frac{\frac{p_{\phi}(\mathbf{y}|\phi_b^{(j)})^b p_{\phi}(\phi_b^{(j)})}{p_{\phi}(\mathbf{y}|\phi^{(j)}) p_{\phi}(\phi^{(j)})}}{\sum_{j=1}^J \frac{p_{\phi}(\mathbf{y}|\phi_b^{(j)})^b p_{\phi}(\phi_b^{(j)})}{p_{\phi}(\mathbf{y}|\phi^{(j)}) p_{\phi}(\phi^{(j)})}} \\ &= \frac{\exp\left\{b \ln p_{\phi}(\mathbf{y}|\phi_b^{(j)}) - \ln p_{\phi}(\mathbf{y}|\phi^{(j)}) + \ln p_{\phi}(\phi_b^{(j)}) - \ln p_{\phi}(\phi^{(j)})\right\}}{\sum_{j=1}^J \left\{b \ln p_{\phi}(\mathbf{y}|\phi_b^{(j)}) - \ln p_{\phi}(\mathbf{y}|\phi^{(j)}) + \ln p_{\phi}(\phi_b^{(j)}) - \ln p_{\phi}(\phi^{(j)})\right\}}.\end{aligned}$$

Under the parameter transformation, we can get that $p_{\phi}(\mathbf{y}|\phi) = p(\mathbf{y}|g(\phi))$ and $p_{\phi}(\phi) = p(g(\phi)) \left| \frac{\partial g(\phi)}{\partial \phi} \right|$. Thus, we can get an consistent estimate of $\mathcal{U}(b)$ as

$$\mathcal{U}(b) \approx \hat{\mathcal{U}}_{LWY}^*(b) = \sum_{j=1}^J \ln p(\mathbf{y}|g(\phi_b^{(j)})) \hat{W}_{\phi}(\phi_b^{(j)}).$$

References

- Chan, Joshua, and Eisenstat, E. (2015). Marginal likelihood estimation with the Cross-Entropy method. *Econometric Reviews*, 34(3), 256-285.
- 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.
- Chib, S. (1995). Marginal Likelihood from the Gibbs Output, *Journal of the American Statistical Association*, 90 (432), 1313-1321.
- Chib, S., and Jeliakov, 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.
- Friel, N., and Pettitt, A. N. (2008). Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3), 589-607.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). Bayesian data analysis, second edition. Chapman & Hall/CRC.
- Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, 57, 1317-1339.
- Ghosh, J., and Ramamoorthi, R. (2003). *Bayesian Nonparametrics*, Springer-Verlag, New York.
- Han, C., and Carlin, B. P. (2001). Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comparative. *Journal of the American Statistical Association*, 96(455), 1122-1132.
- Hoehna, S., Landis, M. L., and Huelsenbeck, J. P. (2017). Parallel power posterior analyses for fast computation of marginal likelihoods in phylogenetics. *Bioinformatics*, forthcoming.
- Hurn, S., Martin, V., Phillips, P. C. B., and Yu, J. (2019). *Financial Econometric Modelling*. Oxford University Press, forthcoming.
- Kass, R. E., and Raftery, A. E. (1995). Bayes Factor. *Journal of the American Statistical Association*, 90, 773-795.
- Koop, G. (2003). *Bayesian Econometrics*. Wiley-Interscience.

- Li, Y., Liu, X. B., Zeng, T., and Yu, J. (2018). A Posterior-Based Wald-Type Statistic for Hypothesis Testing. Economics and Statistics Working Paper, Singapore Management University.
- Schervish, M. J. (2012). *Theory of Statistics*. Springer Science & Business Media.
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). WinBUGS User Manual. Cambridge, UK: MRC Biostatistics Unit.
- Young, K. D. S., and Pettit, L. I. (1996). On priors and Bayes factors. *Journal of Econometrics*, 75(1), 113-119.