ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Deviance Information Criterion for Bayesian model selection: Theoretical justification and applications[☆]

Yong Li [a], Sushanta K. Mallick [b], Nianling Wang [c], Jun Yu [d], Tao Zeng [e,*]

[a] Renmin University of China, China
[b] Queen Mary University of London, United Kingdom
[c] Capital University of Economics and Business, China
[d] Faculty of Business Administration, University of Macau, China
[e] Zhejiang University, China

ARTICLE INFO

JEL classification:
C11
C52
C25
C22
C32

Keywords:
AIC
DIC
Expected loss function
Kullback–Leibler divergence
Model comparison
Plug-in predictive distribution

ABSTRACT

This paper provides a theoretical justification for the Deviance Information Criterion (DIC) as a Bayesian model selection tool using MCMC output. Unlike Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which balance model adequacy against complexity without considering prior information, DIC incorporates priors into this trade-off. The contributions of this paper are two-fold. First, it demonstrates that when a plug-in predictive distribution — obtained by substituting parameter values with their optimal estimates to yield the plug-in estimated sampling distribution — is used under a set of regularity conditions, the DIC serves as an asymptotically unbiased estimator of the expected Kullback–Leibler divergence between the data-generating process and the plug-in predictive distribution. Second, it develops higher-order expansions for DIC and the effective number of parameters, highlighting the effect of the priors. We employ DIC to compare discrete-choice models, stochastic frontier models, and copula models in three empirical applications; the results align with theoretical expectations, showing the utility of DIC as a versatile tool outperforming the traditional model selection criteria. It is found that the logit model is better than the probit model for investigating the marginal effects of parents' education on children's completion of high school. Additionally, the stochastic frontier model with an exponential distribution better fits electricity utility data than the normal distribution. Finally, the chosen copula models for S&P index returns exhibit heavy tails and strong tail dependence. By modelling the effect of priors through higher order expansions, we also find the above empirical models outperforming their benchmark counterparts in terms of predictive accuracy.

☆ We wish to thank Eric Renault, Peter Phillips and David Spiegelhalter, and two anonymous reviewers of this journal for their constructive comments and suggestions in improving the paper. Yong Li, School of Economics, Renmin University of China, Beijing, 100872, China. Sushanta K. Mallick, School of Business and Management, Queen Mary University of London, London E1 4NS, UK. Nianling Wang, School of Finance, Capital University of Economics and Business, Beijing, China. Jun Yu, Faculty of Business Administration, University of Macau, Taipa, Macao, China. Tao Zeng, School of Economics, Academy of Financial Research and Institute of China's State System Research, Zhejiang University, Zhejiang, China 310027. Li gratefully acknowledges the financial support of the National Natural Science Foundation of China (No. 72273142, 72394392). Zeng gratefully acknowledges the financial support of the National Natural Science Foundation of China (No. 72073121) and the National Social Science Fund of China (No. 24&ZD070). Wang gratefully acknowledges the financial support of the National Natural Science Foundation of China (No. 72203151). Yu gratefully acknowledges the financial support from the University of Macau Development Fund.

* Corresponding author.
E-mail address: tzeng@zju.edu.cn (T. Zeng).

https://doi.org/10.1016/j.jeconom.2025.105978
Received 17 May 2024; Received in revised form 26 November 2024; Accepted 23 February 2025

1. Introduction

A highly important statistical inference problem often faced by model builders and empirical researchers in economics is model selection. Many penalty-based information criteria have been proposed to select from a set of candidate models. In the frequentist statistical framework, perhaps the most popular information criterion is the Akaike's Information Criterion (AIC) introduced by Akaike (1973). It balances the model fit and complexity by penalizing the number of parameters and is concerned with how the observed data predict the replicated data. AIC is computationally efficient, requiring only a single model fit, and it does not necessitate splitting the data, allowing the use of the entire dataset to estimate the model, leading to more stable parameter estimates. Compared to cross-validation, which involves repeatedly splitting the data and refitting the model, AIC is less computationally intensive and particularly advantageous when getting access to a separate test set is expensive. Importantly, AIC is asymptotically equivalent to leave-one-out cross-validation (LOOCV) under the assumption of independent and identically distributed (i.i.d.) observations, as demonstrated by Stone (1977). Additionally, Inoue and Kilian (2006) showed that minimizing AIC is equivalent to minimizing the out-of-sample one-step forecast mean squared error (MSE) for time series models under suitable conditions. This property makes AIC an attractive criterion for selecting models for forecasting.

Arguably one of the most important developments for model selection in the Bayesian literature in the last twenty years is the deviance information criterion (DIC) of Spiegelhalter et al. (2002).[1] DIC is understood as a Bayesian version of AIC (Spiegelhalter et al., 2002; Claeskens and Hjort, 2008). Like AIC, it also trades off a measure of model adequacy against a measure of complexity. However, unlike AIC, DIC takes prior information into account.

DIC is constructed based on the posterior mean of the log-likelihood or the deviance and has several desirable features. First, DIC is easy to calculate when the likelihood function is available in closed-form, and the posterior distributions of models are obtained by Markov chain Monte Carlo (MCMC) simulation. Second, it applies to a wide range of statistical models. Third, unlike Bayes factors (BF), it is not subject to Jeffreys-Lindley-Barlett's paradox and can be calculated when vague or even improper priors are used.

However, as acknowledged in Spiegelhalter et al. (2002, 2014), the decision-theoretic justification of DIC is not rigorous in the literature. In fact, in justification given by Spiegelhalter et al. (2002) is heuristic. This point is also acknowledged in Claeskens and Hjort (2008). The first contribution of the present paper is to provide a rigorous decision-theoretic justification to DIC purely in a frequentist setup. It can be shown that DIC is an asymptotically unbiased estimator of the expected Kullback–Leibler (KL) divergence between the data generating process (DGP) and the plug-in predictive distribution when the posterior mean is used. This justification is similar to how AIC has been justified. The second contribution of the present paper is to develop high-order expansions to DIC and the effective number of parameters that allow us to easily see the effect of the prior on DIC and the effective number of parameters.

The rest of the paper is organized as follows. Section 2 explains how to treat the model selection as a decision problem and provides a rigorous decision-theoretic justification to DIC of Spiegelhalter et al. (2002) under a set of regularity conditions. In Section 3, we give two examples to illustrate the effect of the prior distribution on DIC in finite samples. In Section 4, we apply DIC to compare alternative discrete-choice models, alternative stochastic frontier models and alternative copula models. Section 5 concludes the paper. In the Appendix A, Theorem 2.1 is proved and the expressions for the high order terms in Lemma 2.2 are given. The online supplement collects the proof of the lemmas in the paper.[2]

2. Decision-theoretic justification of DIC

There are essentially two strands of literature on model selection.[3] The first strand aims to answer the following question — which model best explains the observed data? The BF (Kass and Raftery, 1995) and its variations belong to this strand. They compare models by examining "posterior probabilities" given the observed data and search for the "true" model. BIC is a large sample approximation to BF, although it is based on the maximum likelihood estimator. The second strand aims to answer the following question: Which model gives the best predictions of future observations generated by the same mechanism that gives the observed data? Clearly, this is a utility-based approach where the utility is set as prediction. Ideally, we would like to choose the model that gives the best overall predictions of future values. Some cross-validation-based criteria have been developed where the original sample is split into a training set and a validation set (Vehtari and Lampinen, 2002; Zhang and Yang, 2015). Unfortunately, different ways of sample splitting often lead to different outcomes. Alternatively, based on replication data generated by the exact mechanism that gives the observed data, some predictive information criteria have been proposed for model selection. They minimize a loss function associated with the predictive decisions. AIC and DIC are two well-known criteria in this framework. After the decision is made about which model should be used for prediction and how predictions should be made, a unique prediction action for future values can be obtained to fulfill the original goal. The latter approach is what we follow in the present paper. Given the relevance of prediction in practice, not surprisingly, such an approach to model selection has been widely used in applications.

---

1 According to Spiegelhalter et al. (2014), Spiegelhalter et al. (2002) was the third most cited paper in international mathematical sciences between 1998 and 2008. Up to March 2025, it has received 15320 citations on Google Scholar.

2 Throughout the paper, we use $:=$, $\mathbf{tr}$, vec, $\otimes$, $o(1)$, $o_p(1)$, $O_p(1)$, $\overset{p}{\to}$ to denote definitional equality, trace, vector operator that converts the matrix into a column vector, Kronecker product, tending to zero, tending to zero in probability, bounded in probability, convergence in probability, respectively. Moreover, we use $\tilde{\theta}_n$, $\overline{\theta}_n$, $\hat{\theta}_n$, θ_n^p to denote a generic estimator, the posterior mean, the quasi maximum likelihood (QML) estimator, and the pseudo true parameter of θ, respectively.

3 For more information about the literature, see Vehtari and Ojanen (2012) and Burnham and Anderson (2002).

2.1. Predictive model selection as a decision problem

Assuming that the probabilistic behavior of observed data, $\mathbf{y} = (y_1, y_2, \ldots, y_n)' \in \mathbf{Y}$, is described by a set of probabilistic models such as $\{M_k\}_{k=1}^K = \{p(\mathbf{y}|\theta_k, M_k)\}_{k=1}^K$ where n is the sample size, θ_k (without confusion, we simply write it as θ) is the set of parameters in candidate model M_k, and $p(\cdot)$ is a probability density function (pdf). Formally, the model selection problem can be taken as a decision problem to select a model among $\{M_k\}_{k=1}^K$ where the action space has K elements, namely, $\{d_k\}_{k=1}^K$, where d_k means M_k is selected.

For the decision problem, a loss function, $\ell(\mathbf{y}, d_k)$, which measures the loss of decision d_k as a function of $\mathbf{y}$, must be specified. Given the loss function, the expected loss (or risk) can be defined as (Berger, 1985)

$$Risk(d_k) = E_{\mathbf{y}}\left[\ell(\mathbf{y}, d_k)\right] = \int \ell(\mathbf{y}, d_k) g(\mathbf{y}) d\mathbf{y},$$

where $g(\mathbf{y})$ is the pdf of the DGP of $\mathbf{y}$. Hence, the model selection problem is equivalent to optimizing the statistical decision,

$$k^* = \arg\min_k Risk(d_k).$$

Based on the set of candidate models $\{M_k\}_{k=1}^K$, the model M_{k^*} with the decision d_{k^*} is selected.

Let $\mathbf{y}_{rep} = (y_{1,rep}, \ldots, y_{n,rep})'$ be the hypothetically replicate data, independently generated by the exact mechanism that gives $\mathbf{y}$. Assume the sample size in $\mathbf{y}_{rep}$ is the same as that in $\mathbf{y}$ (i.e. n). Consider the predictive density of this replicate experiment for a candidate model M_k. The plug-in predictive density can be expressed as $p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)$ for M_k where $\tilde{\theta}_n(\mathbf{y})$ is an estimate of θ based on $\mathbf{y}$ (when there is no confusion we simply write $\tilde{\theta}_n(\mathbf{y})$ as $\tilde{\theta}_n$).

The quantity that has been used to measure the quality of the candidate model in terms of its ability to make predictions is the KL divergence between $g\left(\mathbf{y}_{rep}\right)$ and $p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)$ multiplied by 2,

$$2 \times KL\left[g\left(\mathbf{y}_{rep}\right), p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)\right] = 2\int \ln \frac{g\left(\mathbf{y}_{rep}\right)}{p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)} g\left(\mathbf{y}_{rep}\right) d\mathbf{y}_{rep}.$$

Naturally, the loss function associated with decision d_k is

$$\ell(\mathbf{y}, d_k) = 2 \times KL\left[g\left(\mathbf{y}_{rep}\right), p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)\right] = 2\int \ln \frac{g\left(\mathbf{y}_{rep}\right)}{p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)} g\left(\mathbf{y}_{rep}\right) d\mathbf{y}_{rep}.$$

As a result, the model selection problem is,

$$\begin{aligned} k^* &= \arg\min_k Risk(d_k) = \arg\min_k E_{\mathbf{y}}\left[\ell(\mathbf{y}, d_k)\right] \\ &= \arg\min_k \left\{ E_{\mathbf{y}} E_{\mathbf{y}_{rep}}\left[2\ln g\left(\mathbf{y}_{rep}\right)\right] + E_{\mathbf{y}} E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)\right] \right\}. \end{aligned}$$

Since $g\left(\mathbf{y}_{rep}\right)$ is the DGP, $E_{\mathbf{y}_{rep}}\left[2\ln g\left(\mathbf{y}_{rep}\right)\right]$ is the same across all candidate models, and hence, is dropped from the above equation. Consequently,

$$k^* = \arg\min_k Risk(d_k) = \arg\min_k E_{\mathbf{y}} E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)\right].$$

The smaller $Risk(d_k)$ is, the better the candidate model performs when using $p\left(\mathbf{y}_{rep}|\tilde{\theta}_n(\mathbf{y}), M_k\right)$ to predict $g\left(\mathbf{y}_{rep}\right)$. The optimal decision makes it necessary to evaluate the risk.

2.2. AIC for predictive model selection

When there is no confusion, we simply write a generic candidate model $p\left(\mathbf{y}|\theta, M_k\right)$ as $p(\mathbf{y}|\theta)$ where $\theta \in \Theta \subseteq R^P$ (i.e. the dimension of θ is P). When the candidate model is different, the value of P may be different. Define AIC by

$$\text{AIC} = -2\ln p\left(\mathbf{y}|\hat{\theta}_n(\mathbf{y})\right) + 2P,$$

where $\hat{\theta}_n(\mathbf{y})$ is the QML estimate from $\mathbf{y}$ defined by

$$\hat{\theta}_n(\mathbf{y}) = \arg\max_{\theta \in \Theta} \ln p\left(\mathbf{y}|\theta, M_k\right),$$

which is the global maximum interior to Θ.

Under a set of regularity conditions, it is well known (e.g. Burnham and Anderson (2002)) that AIC is an asymptotically unbiased estimator of $E_{\mathbf{y}} E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\hat{\theta}_n(\mathbf{y}), M_k\right)\right]$, that is, as $n \to \infty$,

$$E_{\mathbf{y}}(\text{AIC}) - E_{\mathbf{y}} E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\hat{\theta}_n(\mathbf{y})\right)\right) \to 0.$$

The decision-theoretic justification of AIC rests on a frequentist framework. Specifically, it requires a careful choice of the KL divergence, the use of QML, and a set of regularity conditions that ensure $\sqrt{n}$ -consistency and the asymptotic normality of QML. The penalty term in AIC arises from two sources. First, the pseudo true parameter value has to be estimated. Second, the estimate obtained from the observed data is not the same as that from the replicate data. Moreover, as pointed out in Burnham and Anderson (2002), the justification of AIC requires the candidate model to be a “good approximation” to the DGP.

2.3. DIC

Spiegelhalter et al. (2002) propose DIC for Bayesian model selection. The criterion is based on the deviance

$$D(\theta) = -2 \ln p(\mathbf{y}|\theta),$$

and takes the form of

$$\text{DIC} = \overline{D(\theta)} + P_D. \tag{1}$$

The first term, interpreted as a Bayesian measure of model fit, is defined as the posterior mean of the deviance, that is,

$$\overline{D(\theta)} = E_{\theta|\mathbf{y}} D(\theta) = E_{\theta|\mathbf{y}} \left[-2 \ln p(\mathbf{y}|\theta)\right].$$

The better the model fits the data, the larger the log-likelihood value, and hence, the smaller the value for $\overline{D(\theta)}$. The second term, used to measure the model complexity and also known as the "effective number of parameters", is defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters:

$$P_D = \overline{D(\theta)} - D\left(\overline{\theta}_n(\mathbf{y})\right) = -2 \int \left[\ln p(\mathbf{y}|\theta) - \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)\right] p(\theta|\mathbf{y})\, d\theta, \tag{2}$$

where $\overline{\theta}_n(\mathbf{y})$ is the posterior mean of θ based on $\mathbf{y}$, defined by $\int \theta p(\theta|\mathbf{y})\, d\theta$. When there is no confusion, we simply write $\overline{\theta}_n(\mathbf{y})$ as $\overline{\theta}_n$.

DIC can be rewritten in two equivalent forms:

$$\text{DIC} = D\left(\overline{\theta}_n\right) + 2P_D, \tag{3}$$

and

$$\text{DIC} = 2\overline{D(\theta)} - D\left(\overline{\theta}_n\right) = -4E_{\theta|\mathbf{y}} \ln p(\mathbf{y}|\theta) + 2 \ln p\left(\mathbf{y}|\overline{\theta}_n\right). \tag{4}$$

DIC defined in Eq. (3) bears similarity to AIC of Akaike (1973) and can be interpreted as a classical "plug-in" measure of fit plus a measure of complexity (i.e. $2P_D$, also known as the penalty term or the "optimism" in the model selection literature). In Eq. (1) the Bayesian measure, $\overline{D(\theta)}$, is the same as $D\left(\overline{\theta}_n\right) + P_D$ that already includes P_D as a penalty for model complexity and, thus, could be better thought of as a measure of model adequacy rather than pure goodness of fit.

Remark 2.1. Unlike AIC that is based on the log-likelihood function (or deviance) with the quasi maximum likelihood (QML) estimate plugged in, DIC is based on the deviance with the posterior mean plugged in. The detachment of DIC from QML is important when candidate models are difficult to estimate by QML. In this case, applied researchers may prefer Bayesian estimation methods over QML. In Bayesian statistics, the recent development of Markov chain Monte Carlo (MCMC) methods has been a key step in making it possible to estimate large hierarchical models, which are hard to estimate by QML, making QML-based model comparison criteria hard to implement. Although the posterior mean may be equivalent to the QML estimate in the limit, they are different in finite samples.

However, as stated explicitly in Spiegelhalter et al. (2002) (Section 7.3 on Page 603 and the first paragraph on Page 605), the justification of DIC is informal and heuristic. It mixes a frequentist setup and a Bayesian setup. In the next subsection, we provide a rigorous decision-theoretic justification of DIC purely in a frequentist setup. Specifically, we show that when a proper loss function is selected, DIC is an unbiased estimator of the expected loss asymptotically.

2.4. Decision-theoretic justification of DIC

When developing DIC, Spiegelhalter et al. (2002) assumes that there is a true distribution for $\mathbf{y}$ in Section 2.2, a pseudo-true parameter value θ_n^p for a candidate model also in Section 2.2, an independent replicate data set $\mathbf{y}_{rep}$ in Section 7.1. All these assumptions are identical to what has been done to justify AIC. Furthermore, as explained in Section 7.1 of Spiegelhalter et al. (2002), the goal for model selection is to estimate the expected loss where the expectation is taken with respect to $\mathbf{y}_{rep}|\theta_n^p$. The assumptions and the goal indicate that a frequentist framework was considered. On the other hand, since the "optimism" associated with the natural estimator depends on a pseudo true parameter value θ_n^p, instead of replacing it with a frequentist estimator and then finding the asymptotic property of the "optimism", Spiegelhalter et al. (2002), in Sections 7.1 and 7.3, replace θ_n^p with a random quantity θ and then calculate the posterior mean of the "optimism". As a result, a Bayesian framework is adopted when studying the behavior of "optimism".

Spiegelhalter et al. (2002) do not explicitly specify the KL divergence function. However, from Eq. (33) on Page 602, the loss function defined in the first paragraph on Page 603, and Eq. (40) on Page 603, one may deduce that the following KL divergence

$$KL\left[p\left(\mathbf{y}_{rep}|\theta\right), p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right] = E_{\mathbf{y}_{rep}|\theta}\left[\ln \frac{p\left(\mathbf{y}_{rep}|\theta\right)}{p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)}\right] \tag{5}$$

was used.[4] Hence,

$$2\times KL\left[p\left(\mathbf{y}_{rep}|\theta\right),p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right]=2E_{\mathbf{y}_{rep}|\theta}\left(\ln p\left(\mathbf{y}_{rep}|\theta\right)\right)+E_{\mathbf{y}_{rep}|\theta}\left(-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right). \quad (6)$$

With this KL function, unfortunately, the first term in the right hand side of Eq. (6) is no longer a constant across candidate models. This is because, when the pseudo-true value is replaced by a random quantity θ, the first term in the right hand side of Eq. (6) is model dependent. This deficiency suggests another KL divergence function is needed.

As in AIC, we first consider the plug-in predictive distribution $p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)$ in the following KL divergence

$$KL\left[g\left(\mathbf{y}_{rep}\right),p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right]=E_{\mathbf{y}_{rep}}\left[\ln\frac{g\left(\mathbf{y}_{rep}\right)}{p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)}\right].$$

The corresponding expected loss function of a statistical decision d_k is

$$\begin{aligned}Risk(d_k)&=E_{\mathbf{y}}\left\{E_{\mathbf{y}_{rep}}\left[2\ln\frac{g\left(\mathbf{y}_{rep}\right)}{p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y}),M_k\right)}\right]\right\}\\&=E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[2\ln g\left(\mathbf{y}_{rep}\right)\right]+E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y}),M_k\right)\right].\end{aligned}$$

Once again, since $E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[2\ln g\left(\mathbf{y}_{rep}\right)\right]$ is the same across candidate models, minimizing the expected loss function $Risk(d_k)$ is equivalent to minimizing

$$E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y}),M_k\right)\right].$$

Denote the selected model by M_{k^*}. Then $p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y}),M_{k^*}\right)$ is used to generate future observations where $\overline{\theta}_n(\mathbf{y})$ is the posterior mean of θ in M_{k^*}.

We are now in the position to provide a rigorous decision-theoretic justification to DIC in a frequentist framework based on a set of regularity conditions. To do so, let us first fix some notations. Let $\mathbf{y}^t=(y_0,y_1,\ldots,y_t)$ for any $0\le t\le n$ and $l_t\left(\mathbf{y}^t,\theta\right)=\ln p\left(\mathbf{y}^t|\theta\right)-\ln p\left(\mathbf{y}^{t-1}|\theta\right)$ be the conditional log-likelihood for the tth observation for any $1\le t\le n$. When there is no confusion, we suppress $l_t\left(\mathbf{y}^t,\theta\right)$ as $l_t(\theta)$ so that the log-likelihood function $\ln p(\mathbf{y}|\theta)$ is $\sum_{t=1}^n l_t(\theta)$.[5] Let $\nabla^j l_t(\theta)$ denote the jth derivative of $l_t(\theta)$ and $\nabla^j l_t(\theta)=l_t(\theta)$ when $j=0$. Furthermore, define

$$\begin{aligned}&\mathbf{s}\left(\mathbf{y}^t,\theta\right)=\frac{\partial\ln p\left(\mathbf{y}^t|\theta\right)}{\partial\theta}=\sum_{i=1}^t\nabla l_i(\theta),\ \mathbf{h}\left(\mathbf{y}^t,\theta\right)=\frac{\partial^2\ln p\left(\mathbf{y}^t|\theta\right)}{\partial\theta\partial\theta'}=\sum_{i=1}^t\nabla^2 l_i(\theta),\\&\mathbf{s}_t(\theta)=\nabla l_t(\theta)=\mathbf{s}\left(\mathbf{y}^t,\theta\right)-\mathbf{s}\left(\mathbf{y}^{t-1},\theta\right),\ \mathbf{h}_t(\theta)=\nabla^2 l_t(\theta)=\mathbf{h}\left(\mathbf{y}^t,\theta\right)-\mathbf{h}\left(\mathbf{y}^{t-1},\theta\right),\\&\mathbf{B}_n(\theta)=Var\left[\frac{1}{\sqrt{n}}\sum_{t=1}^n\nabla l_t(\theta)\right],\bar{\mathbf{H}}_n(\theta)=\frac{1}{n}\sum_{t=1}^n\mathbf{h}_t(\theta),\\&\bar{\mathbf{J}}_n(\theta)=\frac{1}{n}\sum_{t=1}^n\left[\mathbf{s}_t(\theta)-\bar{\mathbf{s}}(\theta)\right]\left[\mathbf{s}_t(\theta)-\bar{\mathbf{s}}(\theta)\right]',\bar{\mathbf{s}}(\theta)=\frac{1}{n}\sum_{t=1}^n\mathbf{s}_t(\theta),\\&L_n(\theta)=\ln p(\theta|\mathbf{y}),\ L_n^{(j)}(\theta)=\partial^j\ln p(\theta|\mathbf{y})/\partial\theta^j,\\&\mathbf{H}_n(\theta)=\int\bar{\mathbf{H}}_n(\theta)g(\mathbf{y})d\mathbf{y},\ \mathbf{J}_n(\theta)=\int\bar{\mathbf{J}}_n(\theta)g(\mathbf{y})d\mathbf{y}.\end{aligned}$$

In this paper, we impose the following regularity conditions.

Assumption 1. $\Theta\subset R^P$ is compact.

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

Assumption 3. For all t, $l_t(\theta)$ satisfies the standard measurability and continuity condition, and the eight-times differentiability condition on Θ almost surely.

[4] In Eq. (33) of Spiegelhalter et al. (2002), the expectation is taken with respect to $\mathbf{y}_{rep}|\theta^t$ which corresponds to the candidate model. In AIC, the expectation is taken with respect to $\mathbf{y}_{rep}$ which corresponds to the DGP.

[5] In the definition of log-likelihood, we ignore the initial condition $\ln p(y_0)$. For weakly dependent data, the impact of ignoring the initial condition is asymptotically negligible.

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

$$\frac{1}{n}\sum_{t=1}^{n}\left(c_t^j(\mathbf{y}^t) - E\left(c_t^j(\mathbf{y}^t)\right)\right) \xrightarrow{p} 0.$$

Assumption 5. For $j = 0, 1, 2, \ldots, 8$, there exist $M_t(y^t)$, $M < \infty$, $\delta > 6$, and $r > 2$ such that for all $\theta \in \Theta$, $\nabla^j l_t(\theta)$ exists, $\sup_{\theta\in\Theta} \left\|\nabla^j l_t(\theta)\right\| \leqslant M_t(y^t)$, $\sup_t E\left\|M_t(\mathbf{y}^t)\right\|^{r+\delta} \leq M$. Moreover, for $j = 2$, there exist a positive random variable $\alpha(y^n)$, a finite constant $\alpha < \infty$, $\delta > 6$, and $r > 2$ such that for all $\theta \in \Theta$, $\nabla^2 l_t(\theta)$ exists, $\inf_{\theta\in\Theta} \left\|\bar{\mathbf{H}}_n(\theta)\right\| \geq \alpha(y^n)$, $E\left\|\alpha(\mathbf{y}^n)\right\|^{r+\delta} \geq \alpha$ and $\sup_{\theta\in\Theta} E\left[\left\|\bar{\mathbf{H}}_n^{-1}(\theta)\right\|^8\right] < \infty$.

Assumption 6. $\left\{\nabla^j l_t(\theta)\right\}$ is L_2-near epoch dependent on $\left\{y_t\right\}_{t=1}^{\infty}$ of size -1 for $j = 0$, L_8-strong near epoch dependent on $\left\{y_t\right\}_{t=1}^{\infty}$ of size -1 for $j = 1$, L_2-strong near epoch dependent on $\left\{y_t\right\}_{t=1}^{\infty}$ of size -1 for $j = 2$ uniformly on Θ.

Assumption 7. Let θ_n^p be the pseudo-true value that minimizes the KL loss between the DGP and the candidate model

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

where $\left\{\theta_n^p\right\}$ is the sequence of minimizers that are interior to Θ uniformly in n. For all $\varepsilon > 0$,

$$\lim_{n\to\infty}\sup \sup_{\Theta\backslash N(\theta_n^p,\varepsilon)} \frac{1}{n}\sum_{t=1}^{n}\left\{E\left[l_t(\theta)\right] - E\left[l_t\left(\theta_n^p\right)\right]\right\} < 0, \tag{7}$$

where $N\left(\theta_n^p, \varepsilon\right)$ is the open ball of radius ε around θ_n^p.

Assumption 8. The sequence $\left\{\mathbf{H}_n\left(\theta_n^p\right)\right\}$ is negative definite and $\left\{\mathbf{B}_n\left(\theta_n^p\right)\right\}$ is positive definite, both uniformly in n.

Assumption 9. $\mathbf{H}_n\left(\theta_n^p\right) + \mathbf{B}_n\left(\theta_n^p\right) = o(1)$.

Assumption 10. The prior density $p(\theta)$ is eight-times continuously differentiable, $p\left(\theta_n^p\right) > 0$ uniformly in n. Moreover, there exists an n^* such that, for any $n > n^*$, the posterior distribution $p(\theta|\mathbf{y})$ is proper, $\int \|\theta\|^2 p(\theta|\mathbf{y})\, d\theta < \infty$.

Remark 2.2. Assumption 1 is the compactness condition. Assumptions 2 and 6 imply weak dependence in y_t and l_t. The first part of Assumption 3 is the continuity condition. Assumption 4 is the Lipschitz condition for l_t first introduced in Andrews (1987) to develop the uniform law of large numbers for dependent and heterogeneous stochastic processes. Assumption 5 contains the domination condition for l_t. Assumption 7 is the identification condition. These assumptions are well-known primitive conditions for developing the QML theory, namely consistency and asymptotic normality, for dependent and heterogeneous data; see, for example, Gallant and White (1988) and Wooldridge (1994).

Remark 2.3. The eight-times differentiability condition in Assumption 3 and the domination condition for up to the eighth derivative of l_t are important to develop a high order stochastic Laplace approximation. In particular, as shown in Kass et al. (1990), these two conditions, together with the well-known consistency condition for QML given by Eq. (8) below, are sufficient for developing the Laplace approximation. This consistency condition requires that, for any $\varepsilon > 0$, there exists $K_1(\varepsilon) > 0$ such that

$$\lim_{n\to\infty} P\left(\sup_{\Theta\backslash N(\theta_n^p,\varepsilon)} \frac{1}{n}\sum_{t=1}^{n}\left[l_t(\theta) - l_t\left(\theta_n^p\right)\right] < -K_1(\varepsilon)\right) = 1. \tag{8}$$

Our Assumption 7 is clearly more primitive than the consistency condition (8). In the following lemma, we show that Assumptions 1–7, including the identification condition (7), are sufficient to ensure (8) as well as the concentration condition around the posterior mode given by Chen (1985). Together with Assumption 10, the concentration condition suggests that the stochastic Laplace approximation can be applied to the posterior distribution, and the asymptotic normality of the posterior distribution can be established. To the best of our knowledge, this is the first time in the literature that primitive conditions have been proposed for the stochastic Laplace approximation. Assumption 10 ensures the second moment of the posterior is bounded. Moreover, it implies that the prior is negligible asymptotically.

Lemma 2.1. *If Assumptions 1–7 hold, then Eq. (8) holds. Furthermore, if Assumptions 1–7 hold, for any $\varepsilon > 0$, there exists $K_2(\varepsilon) > 0$ such that*

$$\lim_{n\to\infty} P\left(\sup_{\Theta\backslash N(\hat{\theta}_n,\varepsilon)} \frac{1}{n}\left[\sum_{t=1}^{n} l_t(\theta) - \sum_{t=1}^{n} l_t(\theta_n^p)\right] < -K_2(\varepsilon)\right) = 1. \tag{9}$$

Let $\vec{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{t=1}^{n} l_t(\theta) + \ln p(\theta)$ be the posterior mode. If, in addition, Assumption 10 holds, then, for any $\varepsilon > 0$, there exists $K_3(\varepsilon) > 0$ such that

$$\lim_{n\to\infty} P\left(\sup_{\Theta\backslash N(\vec{\theta}_n,\varepsilon)} \frac{1}{n}\left(\sum_{t=1}^{n}\left[l_t(\theta) - l_t(\theta_n^p)\right] + \ln p(\theta) - \ln p(\theta_n^p)\right) < -K_3(\varepsilon)\right) = 1. \tag{10}$$

Remark 2.4. Assumption 9 gives the exact requirement for "good approximation". This generalizes the definition of information matrix equality (White, 1996). We now give an example where $\mathbf{H}_n(\theta_n^p) + \mathbf{B}_n(\theta_n^p)$ is $o(1)$ but not zero in finite samples. Let the DGP be

$$y_t = x_{1t}\beta_0 + x_{2t}\gamma_0 + \varepsilon_t,\ \varepsilon_t \overset{iid}{\sim} N(0,\sigma_0^2),$$

where (x_{1t}, x_{2t}) is iid over t and independent of ε_t. Assume that $\gamma_0 = \delta_0/n^{1/2}$, where δ_0 is an unknown constant. Let the candidate model be

$$y_t = x_{1t}\beta + v_t, v_t \overset{iid}{\sim} N(0,\sigma^2).$$

In this case

$$l_t(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{(y_t - x_{1t}\beta)^2}{2\sigma^2},$$

where $\theta = (\beta,\sigma^2)'$. In this case, the pseudo true value is $\theta_n^p = \left(\beta_n^p, \sigma_n^{2p}\right)'$, which maximizes $E\left[l_t(\theta)\right]$, and can be expressed as

$$\beta_n^p = \beta_0 + b\gamma_0, \sigma_n^{2p} = \sigma_0^2 + c\gamma_0^2,$$

where $b = \left[E\left(x_{1t}^2\right)\right]^{-1} E\left(x_{1t}x_{2t}\right)$ and $c = E\left(x_{2t}^2\right) - \left[E\left(x_{1t}x_{2t}\right)\right]^2\left[E\left(x_{1t}^2\right)\right]^{-1}$. Hence,

$$-E\left[\mathbf{h}_t(\theta_n^p)\right] = \begin{bmatrix} \frac{E\left(x_{1t}^2\right)}{\sigma_n^{2p}} & 0 \\ 0 & -\frac{1}{2\left(\sigma_n^{2p}\right)^2} + \frac{\sigma_0^2 + c\gamma_0^2}{\left(\sigma_n^{2p}\right)^3} \end{bmatrix},$$

$$-\mathbf{H}_n(\theta_n^p) = -\frac{1}{n}\sum_{t=1}^{n} E\left[\mathbf{h}_t(\theta_n^p)\right] = -E\left[\mathbf{h}_t(\theta_n^p)\right].$$

From the iid assumption, we have

$$\begin{aligned} &Var\left(s_t(\theta_n^p)\right) = E\left(s_t(\theta_n^p)\, s_t(\theta_n^p)'\right) \\ &= \begin{bmatrix} \frac{\sigma_0^2 E\left(x_{1t}x_{1t}'\right)}{\sigma_n^{2p}} + \frac{d_1\gamma_0^2}{\left(\sigma_n^{2p}\right)^2} & \frac{d_2\gamma_0^3}{2\left(\sigma_n^{2p}\right)^2} \\ \frac{d_2\gamma_0^3}{2\left(\sigma_n^{2p}\right)^2} & -\frac{1}{4\left(\sigma_n^{2p}\right)^2} + \frac{3\sigma_0^2 + 6c\sigma_0^2\gamma_0^2 + d_3\gamma_0^4}{4\left(\sigma_n^{2p}\right)^4} \end{bmatrix}. \end{aligned}$$

where $d_j = E\left[x_{1t}^{4-j-1}\left(x_{2t} - x_{1t}b\right)^{j+1}\right]$ for $j = 1,2,3$ and

$$\begin{aligned} \mathbf{B}_n(\theta_n^p) &= Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} s_t(\theta_n^p)\right) = \frac{1}{n}\sum_{t=1}^{n} Var\left(s_t(\theta_n^p)\right) = \mathbf{J}_n(\theta_n^p) \\ &= \begin{bmatrix} \frac{\sigma_0^2 E\left(x_{1t}^2\right)}{\sigma_n^{2p}} + \frac{d_1\gamma_0^2}{\left(\sigma_n^{2p}\right)^2} & \frac{d_2\gamma_0^3}{2\left(\sigma_n^{2p}\right)^2} \\ \frac{d_2\gamma_0^3}{2\left(\sigma_n^{2p}\right)^2} & -\frac{1}{4\left(\sigma_n^{2p}\right)^2} + \frac{3\left(\sigma_0^2\right)^2 + 6c\sigma_0^2\gamma_0^2 + d_3\gamma_0^4}{4\left(\sigma_n^{2p}\right)^4} \end{bmatrix}. \end{aligned}$$

Hence,

$$\lim_{n\to\infty} \mathbf{B}_n\left(\theta_n^p\right) = \lim_{n\to\infty} \mathbf{J}_n\left(\theta_n^p\right) = \lim_{n\to\infty} -\mathbf{H}_n\left(\theta_n^p\right) = \begin{bmatrix} \frac{E\left(x_{1t}^2\right)}{\sigma_0^2} & 0 \\ 0 & \frac{1}{2\left(\sigma_0^2\right)^2} \end{bmatrix}$$

since $\gamma_0 = \delta_0/n^{1/2}$. Thus, $\mathbf{H}_n\left(\theta_n^p\right) + \mathbf{B}_n\left(\theta_n^p\right) = o(1)$. However, $\mathbf{H}_n\left(\theta_n^p\right) + \mathbf{B}_n\left(\theta_n^p\right) \neq 0$ for any finite n. The violation of this assumption has implications for the expression of DIC and hence, its theoretical justification. This issue has been carefully investigated in Li et al. (2020).

Remark 2.5. The first part of Assumptions 5 and 6 are used to justify AIC; see Li et al. (2024) for more details. The second part of Assumption 5 is used to bound the approximation errors of the Laplace approximations of the posterior mean and the posterior mean variance; see, for example, see Huggins et al. (2018).

To develop the Laplace approximation, we need to fix more notations. For the convenience of exposition, we let $\bar{\mathbf{H}}_n^{(j)}(\theta) = \frac{1}{n}\sum_{t=1}^n \nabla^j l_t(\theta)$ for $j = 3,4,5$. Let $\pi(\theta) = \ln p(\theta)$, $\widehat{p}$, $\widehat{\pi}$, $\nabla^j\widehat{p}$, and $\nabla^j\widehat{\pi}$ be the values of functions, $p(\theta)$, $\pi(\theta)$, $\nabla^j p(\theta)$, and $\nabla^j\pi(\theta)$ evaluated at $\widehat{\theta}_n$. The next lemma extends Theorem 4 of Kass et al. (1990) to a higher order in matrix form.

Lemma 2.2. *Under Assumptions 1–10, we have, as $n \to \infty$,*

$$\begin{aligned} &\frac{\int l_t(\theta)\, p(\theta)\, p(\mathbf{y}|\theta)\, d\theta}{\int p(\theta)\, p(\mathbf{y}|\theta)\, d\theta} \\ &= l_t\left(\widehat{\theta}_n\right) + \frac{1}{n}B_{t,1} + \frac{1}{n^2}\left(B_{t,21}^1 + B_{t,21}^2 + B_{t,22} - B_4 B_{t,1}\right) + O_p\left(n^{-3}\right), \end{aligned} \quad (11)$$

where $B_{t,1}, B_{t,21}^1, B_{t,21}^2, B_{t,22}, B_4$ are all $O_p(1)$ with the expressions given in Appendix A.2.

The following lemma develops a high-order expansion of P_D and DIC.

Lemma 2.3. *Under Assumptions 1–10, we have, as $n \to \infty$,*

$$\begin{aligned} P_D &= P + \frac{1}{n}C_1 - \frac{1}{n}C_2 + O_p\left(n^{-2}\right), \\ DIC &= AIC + \frac{1}{n}D_1 + \frac{1}{n}D_2 + O_p\left(n^{-2}\right), \end{aligned}$$

where

$$\begin{aligned} C_1 &= \frac{1}{4}\mathbf{tr}\left(A_2\right) - \frac{1}{6}A_3 = O_p(1), \\ C_2 &= \mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2\widehat{\pi}\right] = O_p(1), \\ D_1 &= -\frac{1}{4}A_1 + \frac{1}{2}\mathbf{tr}\left(A_2\right) - \frac{1}{3}A_3 = O_p(1), \\ D_2 &= C_{21} - 2C_2 - C_{23} = O_p(1), \end{aligned}$$

with

$$\begin{aligned} A_1 &= vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right), \\ A_2 &= \left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)\right]'\bar{\mathbf{H}}_n^{(4)}\left(\widehat{\theta}_n\right), \\ A_3 &= vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\right)'\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right] vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\right), \\ C_{21} &= \nabla\widehat{\pi}'\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right), \\ C_{23} &= \nabla\widehat{\pi}'\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla\widehat{\pi}. \end{aligned}$$

Theorem 2.1. *Under Assumptions 1–10, we have, as $n \to \infty$,*

$$E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right] = E_{\mathbf{y}}(DIC) + o(1).$$

Remark 2.6. DIC is an unbiased estimator of $E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right]$ asymptotically, according to Theorem 2.1 . Hence, the decision-theoretic justification to DIC is that DIC selects a model that asymptotically minimizes the expected loss, which is the expected KL divergence between the DGP and the plug-in predictive density $p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)$. A key difference between AIC and DIC is that the plug-in predictive density is based on different estimators of θ. In AIC, the QML estimate, $\widehat{\theta}_n(\mathbf{y})$, is used. In DIC, the posterior mean, $\overline{\theta}_n(\mathbf{y})$, is used. In this sense, DIC is the Bayesian version of AIC.

Remark 2.7. The justification of DIC remains valid if the posterior mean is replaced with the posterior mode or with the QML estimator and/or if P_D is replaced with P. This is because the justification of DIC requires the information matrix identity to hold asymptotically, and the posterior distribution to converge to a normal distribution (more specifically, the posterior mean minus the posterior mode converges to zero and the posterior variance converges to zero).

Remark 2.8. In AIC, the number of parameters, P, is used to measure model complexity. When the prior is informative, the prior imposes additional restrictions on the parameter space, and hence, P_D may not be close to P in finite samples. A useful contribution of DIC is to provide a way to measure the model complexity when the prior information is incorporated; see Brooks (2002). From Lemma 2.3, the effect of prior on P_D depends on C_2, which can be thought of as a measure of the ratio of the information in the prior to the information in the likelihood about the parameters. The effect of prior on DIC depends on D_2, which in turn depends on C_{21}, C_2, and C_{23}.

Remark 2.9. If $p(\mathbf{y}|\theta)$ has a closed-form expression, DIC is trivially computable from the MCMC output. The computational tractability, together with the versatility of MCMC and the fact that DIC is incorporated into a Bayesian software, WinBUGS, are among the reasons why DIC has enjoyed a very wide range of applications.

Remark 2.10. Although the theoretic framework under which we justify DIC is general, it requires consistency of the posterior mean, the asymptotic normal approximation to the posterior distribution, and the asymptotic normality to the QML estimator. When there are latent variables in the candidate model under which the number of latent variables grows as n grows, consistency and the asymptotic normality may not hold if the parameter space is enlarged to include latent variables. As a result, our decision-theoretic justification DIC is not applicable. A recent study by Li et al. (2020) provides a modification to DIC to compare latent variable models.

3. Examples

In this section, we use two examples from Spiegelhalter et al. (2002), namely, the normal linear model with known sampling precision and the normal linear model with unknown sampling precision, to illustrate the properties of DIC. In particular, we pay attention to the effect of prior on P_D and DIC.

3.1. The normal linear model with known sampling precision

The general hierarchical normal model described by Lindley and Smith (1972) is

$$\mathbf{y} \sim N\left(F_1\theta_1, G_1\right), \tag{12}$$

and the conjugate prior for θ_1 is

$$\theta_1 \sim N\left(F_2\boldsymbol{\phi}, G_2\right), \tag{13}$$

where F_1 is $n \times P$ matrix, θ_1 is a $P \times 1$ vector, G_1 is $n \times n$ matrix. Assume G_1, F_2, $\boldsymbol{\phi}$, and G_2 are all known. In this case, $\theta = \theta_1$. The log likelihood function is

$$L(\mathbf{y}|\theta) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\left|G_1\right| - \frac{1}{2}\left(\mathbf{y} - F_1\theta_1\right)' G_1^{-1}\left(\mathbf{y} - F_1\theta_1\right).$$

It is easy to see that the QML estimate of θ is

$$\widehat{\theta}_n = \left(F_1' G_1^{-1} F_1\right)^{-1} F_1' G_1^{-1}\mathbf{y}. \tag{14}$$

The log prior density is

$$\pi(\theta) = -\frac{P}{2}\ln 2\pi - \frac{1}{2}\ln\left|G_2\right| - \frac{1}{2}\left(\theta_1 - F_2\boldsymbol{\phi}\right)' G_2^{-1}\left(\theta_1 - F_2\boldsymbol{\phi}\right).$$

It is well-known that the posterior distribution of θ is

$$\theta|\mathbf{y} \sim N(Vb, V),$$

where

$$V = \left(F_1' G_1^{-1} F_1 + G_2^{-1}\right)^{-1}, \tag{15}$$

$$b = F_1' G_1^{-1}\mathbf{y} + G_2^{-1} F_2\boldsymbol{\phi}. \tag{16}$$

By Lemma 2.3, we have

$$\overline{\theta}_n = \widehat{\theta}_n - \frac{1}{n}\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla\pi\left(\widehat{\theta}_n\right) + O_p\left(n^{-2}\right), \tag{17}$$

$$V = -\frac{1}{n}\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1} + O_p\left(n^{-2}\right), \tag{18}$$

$$P_D = P - \frac{1}{n}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla^2\pi\left(\hat{\theta}_n\right)\right] + O_p\left(n^{-2}\right), \tag{19}$$

where $\nabla\pi\left(\hat{\theta}_n\right) = G_2^{-1}\left(\hat{\theta}_n - F_2\boldsymbol{\phi}\right)$ and $\nabla^2\pi\left(\hat{\theta}_n\right) = -G_2^{-1}$.

In (19), one can see the effect of prior on P_D via $\nabla^2\pi\left(\hat{\theta}_n\right)$, which is determined by the curvature of the density of prior at $\hat{\theta}_n$. Note that the third order derivative of the log likelihood function $L(\mathbf{y}|\theta)$ is zero. Thus, $D_1 = C_{21} = 0$ and the effect of prior on DIC is

$$D_2 = -2C_2 - C_{23} = -2\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla^2\pi\left(\hat{\theta}_n\right)\right] - \nabla\pi\left(\hat{\theta}_n\right)'\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla\pi\left(\hat{\theta}_n\right).$$

Hence, by Lemma 2.3, we have

$$\mathrm{DIC} = \mathrm{AIC} + \frac{1}{n}D_2 + O_p\left(n^{-2}\right).$$

Spiegelhalter et al. (2002) express P_D as

$$P_D = \mathbf{tr}\left[F_1'G_1^{-1}F_1V\right] = \mathbf{tr}\left[-L^{(-2)}\left(\bar{\theta}_n\right)V\right] = P - \mathbf{tr}\left[G_2^{-1}V\right], \tag{20}$$

where $L^{(-2)}(\theta)$ is the inverse of $L^{(2)}(\theta)$ and $L^{(2)}(\theta) = n\bar{\mathbf{H}}_n(\theta) = -F_1'G_1^{-1}F_1$. Together with (18), (20) is the same as (19).

3.2. The normal linear models with unknown sampling precision

Suppose the model is

$$\mathbf{y} \sim N\left(F_1\theta_1, \tau^{-1}G_1\right). \tag{21}$$

Assume the conjugate prior for θ_1 is

$$\theta_1 \sim N\left(F_2\boldsymbol{\phi}, \tau^{-1}G_2\right), \tag{22}$$

and the conjugate prior for τ is

$$\tau \sim \Gamma(a, b). \tag{23}$$

Assume G_1, F_2, $\boldsymbol{\phi}$, and G_2 are all known. Let $\theta = \left(\theta_1', \tau\right)'$, where the dimension of θ and θ_1 is $P\times 1$ and $P_1\times 1$, respectively. Clearly, $P = P_1 + 1$. The dimension of F_1, G_1, F_2 and G_2 is $n\times P_1$, $P_1\times P_1$, $P_1\times P_1$, $P_1\times P_1$, respectively. It is well-known that the posterior of θ is

$$\theta_1|\tau,\mathbf{y} \sim N\left(V_1b_1, V_1\right) \text{ and } \tau|\mathbf{y} \sim \Gamma\left(a + \frac{n}{2}, b + \frac{S}{2}\right),$$

where

$$V_1^{-1} = \tau V^{-1}, b_1 = \tau b, S = \left(\mathbf{y}-F_1F_2\boldsymbol{\phi}\right)'\left(G_1 + F_1'G_2F_1\right)^{-1}\left(\mathbf{y}-F_1F_2\boldsymbol{\phi}\right).$$

According to Lemma 2.3, we have

$$\begin{aligned} P_D &= P + \frac{1}{n}C_1 - \frac{1}{n}C_2 + O_p\left(n^{-2}\right) \\ &= P + \frac{1}{n}\left(\frac{1}{4}\mathbf{tr}[A_2] - \frac{1}{6}A_3\right) - \frac{1}{n}C_2 + O_p\left(n^{-2}\right) \\ &= P + \left(-\frac{5}{3n} + \frac{P_1}{n}\right) - \frac{1}{n}C_2 + O_p\left(n^{-2}\right), \end{aligned} \tag{24}$$

since $\mathbf{tr}[A_2] = -12$, and $A_3 = -6P_1 - 8$. The effect of prior on P_D is

$$-\frac{1}{n}C_2 = -\mathbf{tr}\left[L^{(-2)}\left(\hat{\theta}_n\right)\nabla^2\pi\left(\hat{\theta}_n\right)\right] = -\left(\mathbf{tr}\left[\left(F_1'G_1^{-1}F_1\right)^{-1}G_2^{-1}\right] + \frac{P_1}{n} + \frac{2(a-1)}{n}\right). \tag{25}$$

From (24), and (25), we can rewrite P_D as

$$P_D = P - \frac{2a}{n} + \frac{1}{3n} + \frac{1}{n}\mathbf{tr}\left[\bar{\mathbf{H}}_{n,11}\left(\hat{\theta}_n\right)^{-1}\hat{\tau}G_2^{-1}\right] + O_p\left(n^{-2}\right), \tag{26}$$

where $\bar{\mathbf{H}}_{n,11}\left(\hat{\theta}_n\right)^{-1} = -\left(\hat{\tau}F_1'G_1^{-1}F_1\right)^{-1}$ is the submatrix of $\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}$ corresponding to θ_1. In (26), one can see the effect of prior on P_D via a and G_2. The effect of prior on DIC is

$$\frac{1}{n}D_2 = \frac{1}{n}C_{21} - \frac{2}{n}C_2 - \frac{1}{n}C_{23},$$

where

$$\frac{1}{n}C_{21} = -\frac{2\hat{\tau}P_1}{n}C_{21}^*, \quad \frac{1}{n}C_{23} = -C_{22}^{*\prime}\left(F_1'G_1^{-1}F_1\right)^{-1}C_{22}^* - \frac{2}{n}\hat{\tau}_n^2 C_{21}^{*2},$$

with

$$C_{21}^* = \frac{P_1}{2\hat{\tau}_n} - \frac{1}{2}C_{22}^{*\prime}G_2C_{22}^* + \frac{(a-1)}{\hat{\tau}_n} - \frac{1}{b}, \quad C_{22}^* = G_2^{-1}\left(\hat{\theta}_{1,n} - F_2\boldsymbol{\phi}\right).$$

Thus,

$$\text{DIC} = \text{AIC} + \frac{1}{n}D_1 + \frac{1}{n}D_2 + O_p\left(\frac{1}{n^2}\right),$$

where

$$D_1 = \frac{1}{2}P_1^2 + 2P_1 - \frac{4}{3}, \quad A_1 = -\left(2P_1^2 + 8\right).$$

Spiegelhalter et al. (2002) express P_D as

$$P_D = \mathbf{tr}\left[F_1'G_1^{-1}F_1V\right] - n\left\{\psi\left(a + n/2\right) - \log\left(a + n/2\right)\right\}, \tag{27}$$

where $\psi(z)$ is the digamma function that has the asymptotic expansion

$$\psi(z) = \ln z - \frac{1}{2z} - \sum_{j=1}^{\infty}\frac{B_{2j}}{2jz^{2j}} = \ln z - \frac{1}{2z} - \frac{1}{12z^2} + O\left(\frac{1}{z^4}\right), \tag{28}$$

where B_k is the kth Bernoulli number. Thus, the second term of the right-hand side of (27) can be written as

$$n\left\{\psi\left(a + n/2\right) - \log\left(a + n/2\right)\right\} = n\left\{-\frac{1}{(2a+n)} - \frac{1}{3(2a+n)^2} + O\left(\frac{1}{n^4}\right)\right\}. \tag{29}$$

The first term of (27) is

$$\mathbf{tr}\left[F_1'G_1^{-1}F_1V\right] = P_1 + \frac{1}{n}\mathbf{tr}\left[\bar{\mathbf{H}}_{n,11}\left(\hat{\theta}_n\right)^{-1}\hat{\tau}G_2^{-1}\right] + O_p\left(\frac{1}{n^2}\right). \tag{30}$$

Hence, from (27), (29), and (30), we have

$$P_D = P_1 + \frac{1}{n}\mathbf{tr}\left[\bar{\mathbf{H}}_{n,11}\left(\hat{\theta}_n\right)^{-1}\hat{\tau}G_2^{-1}\right] + 1 - \frac{2a - \frac{1}{3}}{2a+n} + O\left(\frac{1}{n^2}\right). \tag{31}$$

Applying the Taylor expansion to $\frac{2a-\frac{1}{3}}{2a+n}$ at $a = 0$, we have

$$\frac{2a - \frac{1}{3}}{2a+n} = -\frac{1}{3n} + \frac{2a}{n} + O\left(\frac{1}{n^2}\right).$$

Substituting this to (31), we can get (26).

From this example, we can see that Lemma 2.3 provides a general and convenient way to measure the effect of prior on P_D. Spiegelhalter et al. (2002) use some specific techniques to derive (26). However, these techniques are problem specific and difficult to use in general.

4. Empirical applications

In this section, we conduct three empirical applications to illustrate the implementation of DIC. The first application compares two alternative discrete choice models to investigate the marginal effects of parents' education level on children's completion of high school. The second application compares two stochastic frontier models under different distributions using electricity utility data. In the third application we compared four copula models using S&P index returns. All three classes of models have been widely applied in economics. Our goal is to demonstrate the extensive applicability of DIC across diverse economic models. For simpler models such as linear regression, DIC and related higher-order expansion terms can be derived in closed form. However, a closed-form expression for DIC becomes infeasible for more complex models such as discrete-choice models, stochastic frontier models, and copula models. Therefore, we provide empirical examples to illustrate how DIC can be practically applied in these more sophisticated models.

In all three applications, the competing models are non-nested, making the hypothesis-testing-based approach to model comparison infeasible. In all three empirical studies, we employ vague priors. Regarding predictive performance, the results indicate that the logit model demonstrates superior effectiveness compared to the probit model in examining the marginal effects of parental education levels on children's likelihood of completing high school. Furthermore, the stochastic frontier model utilizing an exponential distribution exhibits better predictive performance for electricity utility data relative to the one based on the normal distribution. Lastly, among copula models applied to S&P index returns, the t-copula with t marginals model outperforms alternative models in predictive accuracy.

Table 1
Model selection results for the probit model and the logit model.

Model	$D\left(\bar{\theta}\right)$	P_D	DIC	C_2/n
Probit	905.3953	8.0040	921.4032	0.0008
Logit	905.2918	8.0253	921.3424	0.0023

4.1. Discrete choice models

In this section, we compare a binary probit model and a binary logit model. Let $\mathbf{y} = \left(y_{1,}y_{2,} \dots y_n\right)'$ be a vector of dependent variables, where y_i takes a value 0 or 1 for $i = 1, 2, \dots, n$; $X = [\mathbf{x}_1', \mathbf{x}_2', \dots, \mathbf{x}_N']'$ be a matrix of independent variables, where $\mathbf{x}_i$ is a $1 \times P$ vector. The probability of $y_i = 1$ conditional on X is

$$P\left(y_i = 1|X_i, \boldsymbol{\beta}\right) = F\left(X_i\boldsymbol{\beta}\right), \tag{32}$$

where $\boldsymbol{\beta}$ is a $P \times 1$ vector. Assume $\left(y_i, \mathbf{x}_i\right)_{i=1}^n$ are identical and independently distributed. If $F\left(X_i\boldsymbol{\beta}\right) = \Phi\left(X_i\boldsymbol{\beta}\right)$ with $\Phi(\cdot)$ being the CDF of $N(0,1)$, (32) is the probit model. And if choosing $F\left(X_i\boldsymbol{\beta}\right)$ be the CDF of the logistic distribution, that is, $F\left(X_i\boldsymbol{\beta}\right) = \frac{\exp\left(X_i\boldsymbol{\beta}\right)}{1+\exp\left(X_i\boldsymbol{\beta}\right)}$, (32) becomes the logit model.

The latent variable representation of (32) is

$$z_i = X_i\boldsymbol{\beta} + \varepsilon_i, \ y_i = \mathbf{I}\left(z_i > 0\right), \tag{33}$$

where z_i is the latent variable, $\mathbf{I}(\cdot)$ is the indicator function. In this representation, ε_i is a standard normal variate in the probit model and a logistic variate in the logit model.

Albert and Chib (1993) propose a Gibbs sampling algorithm for (33) based on the data augmentation technique of Tanner and Wong (1987). Zens et al. (2022a) apply the marginal data argumentation technique of Liu and Wu (1999) to boost the convergence of the Gibbs sampling algorithm for the probit model. In the logit model, the latent variable follows a linear model with a logistic error term. To approximate the error distribution, Held and Holmes (2006) use the scale mixture normal representation while Polson et al. (2013) use a Pólya-Gamma (UPG) mixture representation. Zens et al. (2022a) combine the UPG representation and the marginal data augmentation technique to improve the efficiency of Gibbs sampler for the logit model. In this paper, we use the algorithm proposed by Zens et al. (2022a) to draw MCMC samplers for the logit model.

We fit the two models to a dataset obtained from the US Panel Study of Income. The dependent variable is a binary variable that takes the value of 1 if a woman participates in the labor force and zero otherwise. The independent variables include the number of children under the age of 5, the number of children between 6 and 18 years, a standardized age index, two binary indicators capturing whether a college degree was obtained by the wife and the husband, the expected log wage of the woman, the logarithm of family income exclusive of the income of the woman. There are 753 observations in the data set.[6] In total, there are eight parameters in both models, including the intercept.

We specify a vague prior distribution for parameters as

$$\boldsymbol{\beta} \sim N\left(0_{k\times 1}, \lambda \times \mathbf{I}_k\right),$$

where $\lambda = 100$ in both models. Here, we draw 5,100,000 random draws from the joint posterior distributions of parameters and latent variables in each model. The first 100,000 draws are used as the burn-in sample. Hence, there are 5,000,000 effective draws. To compute P_D, we need to evaluate $E_{\boldsymbol{\theta}|\mathbf{y}}\left[\ln p(\mathbf{y}|\boldsymbol{\theta})\right]$ where $\boldsymbol{\theta} = \boldsymbol{\beta}$, which does not have a closed-form expression. We approximate it based on the MCMC output as,

$$E_{\boldsymbol{\theta}|\mathbf{y}}\left[\ln p(\mathbf{y}|\boldsymbol{\theta})\right] \approx \frac{1}{5000000}\sum_{m=1}^{M} \ln p\left(\mathbf{y}|\boldsymbol{\theta}^{(m)}\right).$$

Table 1 reports $D\left(\bar{\theta}\right)$, P_D, DIC, and C_2/n for both models. DIC suggests that the logit model is slightly better than the probit model. The difference between the two DIC values is mainly due to the difference between the two $D\left(\bar{\theta}\right)$ values. This is not surprising as the priors are vague. To examine the effect of the priors on P_D, we can compare the two $C_2/n = n^{-1}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla^2\hat{\pi}\right]$ values. It is 0.0008 for the probit model and 0.0023 for the logit model, both being negligible. Not surprisingly, P_D is 8.0040 in the probit model and 8.0253 in the logit model, both values very close to the actual number of parameters.

[6] For more details about the dataset, see Zens et al. (2022b).

4.2. Stochastic frontier models

Since Aigner et al. (1977) the stochastic frontier models have proven useful in analyzing the production efficiency of firms. As there is a latent variable in the model (the inefficiency variable), MCMC was proposed to provide Bayesian analysis of the stochastic frontier models in Koop et al. (1995). See also Kurkalova and Carriquiry (2002), Tsionas (2002), Kumbhakar and Tsionas (2005) and Tsionas and Mallick (2019). With a Cobb–Douglas cost frontier function, the stochastic frontier model can be expressed as,

$$y_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + u_i + v_i, i = 1, \ldots, n, \tag{34}$$

where y_i is the logarithm of the production cost and $\boldsymbol{x}_i = \{\ln Q, \ln \frac{P_l}{P_f}, \ln \frac{P_k}{P_f}, (\ln Q)^2\}'$ contains cost-related variables for firm i, with Q being the output and P_l, P_k and P_f being the three factors (labor, capital and fuel). While the error term u_i captures the production inefficiency that is assumed to be nonnegative, v_i is the error of the production function. We assume v_i and u_i are independent of each other and both are independent of $\boldsymbol{x}_i$. Moreover, we assume $v_i \sim iid N(0, \sigma_v^2)$. For u_i, two well known distributional assumptions have been adopted in the literature: the half normal distribution and the exponential distribution, i.e., $u_i \sim iid N^+(0, \sigma_u^2)$ or $u_i \sim iid Exp(\eta)$ such that $E[u_i] = \eta, Var(u_i) = \eta^2$. We now use DIC to compare these two alternative specifications. Let the composite error $\varepsilon_i = u_i + v_i$.

Under the half normal distribution, according to Kumbhakar and Lovell (2000), the density of ε_i is

$$f(\varepsilon_i) = \frac{2}{\sigma}\phi(\frac{\varepsilon_i}{\sigma})\Phi(\frac{\varepsilon_i\lambda}{\sigma}), \tag{35}$$

where $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, $\lambda = \frac{\sigma_u}{\sigma_v}$, $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and the density of $N(0,1)$. Then, the log likelihood function for the model is

$$\ln L = n \ln 2 - \frac{n}{2}\ln 2\pi - n \ln \sigma + \sum_{i=1}^{n} \ln \Phi(\frac{(y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta})\lambda}{\sigma}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta})^2. \tag{36}$$

Under the exponential distribution, the density of ε_i is

$$f(\varepsilon_i) = \eta \exp(\frac{\eta^2\sigma_v^2}{2} - \eta\varepsilon_i)\Phi(\frac{\varepsilon_i}{\sigma_v} - \eta\sigma_v). \tag{37}$$

Then, the log likelihood function for the model is

$$\ln L = \frac{n\eta^2\sigma_v^2}{2} + n \ln \eta - \eta \sum_{i=1}^{n}(y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta}) + \sum_{i=1}^{n} \ln \Phi(\frac{y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma_v} - \eta\sigma_v). \tag{38}$$

The data we use covers 123 U.S. electric utility firms in 1970 (i.e. $n = 123$), which can be found in Greene (1990). For posterior sampling, following Griffin and Steel (2007), we use the following prior distributions.

$$\alpha \sim N(0, 10^3), \beta \sim N(0, 10^3), \frac{1}{\sigma_v^2} \sim \Gamma(0.001, 0.001), \tag{39}$$

and

$$\begin{cases} \frac{1}{\sigma_u^2} \sim \Gamma(1, 1/37.5), \text{ if } u_i \text{ is the half normal distribution,} \\ \eta \sim Exp(-\ln 0.875), \text{if } u_i \text{ is the exponential distribution.} \end{cases}$$

The MCMC output is obtained from WinBUGS (Spiegelhalter et al., 2002). The total number of iterations is 200,000. The burn-in period is the first 20,000 iterations. One effective sample is taken for every five samples in the remaining iterations, resulting in 36,000 samples for each parameter from their posterior distributions. These effective draws are used for Bayesian parameter estimation and DIC computation.

The parameter estimation results for the two models are reported in Table 2. Table 3 reports $D\left(\overline{\theta}\right)$, P_D, DIC, and C_2/n for both models. The stochastic frontier model with the exponential distribution has a smaller DIC value than that with the half normal distribution (−120.8937 versus −120.0211). This indicates that, for the data, DIC selects the exponential distribution. To see the effect of priors, we report the value of C_2/N. With the vague priors, the values of C_2/N are negligible in both models.

4.3. Copula models

In this section, we compare several copula models based on estimated DIC. Copula models are popular tools in finance to model the joint distribution of multiple asset returns. It consists of the marginal distribution of each random variable and a copula function. Consider a simple case where there are two assets. Let r_{1t} and r_{2t} be daily log returns for asset 1 and asset 2 at time t. Assume

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

Table 2
Parameter estimation results for the stochastic frontier models with the half normal distribution and the exponential distribution.

	SFA model with the half normal distribution						
	α	β_1	β_2	β_3	β_4	$1/\sigma_u^2$	$1/\sigma_v^2$
Posterior mean	−7.3834	0.4056	0.2427	0.0616	0.0307	54.9977	83.0852
Posterior SD	0.3342	0.0393	0.0665	0.0624	0.0027	32.9261	29.0756
	SFA model with the exponential distribution						
	α	β_1	β_2	β_3	β_4	η	$1/\sigma_v^2$
Posterior mean	−7.4656	0.4248	0.2495	0.0480	0.0296	12.4287	81.2437
Posterior SD	0.3403	0.0434	0.0647	0.0616	0.0029	5.4946	24.6652

Table 3
Model selection results for the stochastic frontier models with the half normal distribution and the exponential distribution.

Model	$D\left(\bar{\theta}\right)$	P_D	DIC	C_2/N
Half normal	−131.2210	5.6000	−120.0211	0.0042
Exponential	−133.1431	6.1247	−120.8937	0.0942

Table 4
Four Copula models to be compared.

Gaussian copula normal marginals model (gnc)	
Distributional assumption: $z_{it} \sim N(0,1)$, Gaussian copula function	
Log likelihood	$-n\ln 2\pi - \frac{n}{2}\ln\left(\frac{1-\delta^2}{h_1 h_2}\right) - \sum_{t=1}^{n}\frac{z_{1t}^2+z_{2t}^2-2\delta z_{1t}z_{2t}}{2(1-\delta^2)}$
Parameters	$\boldsymbol{\theta} = (\mu_1 \quad h_1 \quad \mu_2 \quad h_2 \quad \delta)^T, h_i \in (0,+\infty), \delta \in [-1,1]$
Priors	$\mu_i \sim \text{Normal}(0,25), h_i \sim \text{Gamma}(0.1,1), \delta \sim \text{Uniform}[-1,1]$
Gaussian copula t marginals model (gtc)	
Distributional assumption: $z_{it} \sim t(0,1,\upsilon)$, Gaussian copula function	
Log likelihood	$-\frac{n}{2}\ln\frac{1-\delta^2}{h_1 h_2} - \sum_{t=1}^{n}\left[\frac{q_{\phi,1t}^2+q_{\phi,2t}^2-2\delta q_{\phi,1t}q_{\phi,2t}}{2(1-\delta^2)} + \frac{1}{2}(q_{\phi,1t}^2+q_{\phi,2t}^2) + \ln f(z_{1t};\upsilon) + \ln f(z_{2t};\upsilon)\right]$
Parameters	$\boldsymbol{\theta} = (\mu_1 \quad h_1 \quad \mu_2 \quad h_2 \quad \delta \quad \upsilon)^T, h_i \in (0,+\infty), \delta \in [-1,1], \upsilon \in (2,+\infty)$
Priors	$\mu_i \sim \text{Normal}(0,25), h_i \sim \text{Gamma}(0.1,1), \delta \sim \text{Uniform}[-1,1], \upsilon-2 \sim \text{Exponential}(1)$
t copula t marginals model (ttc)	
Distributional assumption: $z_{it} \sim t(0,1,\upsilon)$, t copula function	
Log likelihood	$-n\ln 2\pi - \frac{n}{2}\ln\frac{1-\delta^2}{h_1 h_2} - \frac{\eta+2}{2}\sum_{t=1}^{n}\ln\left(1+\frac{q_{f,1t}^2+q_{f,2t}^2-2\delta q_{f,1t}q_{f,2t}}{\eta(1-\delta^2)}\right)$ $-\sum_{t=1}^{n}\left[\ln f(q_{f,1t};\eta) + \ln f(q_{f,2t};\eta) - \ln f(z_{1t};\upsilon) - \ln f(z_{2t};\upsilon)\right]$
Parameters	$\boldsymbol{\theta} = (\mu_1 \quad h_1 \quad \mu_2 \quad h_2 \quad \delta \quad \upsilon \quad \eta)^T, h_i \in (0,+\infty), \delta \in [-1,1], \upsilon,\eta \in (2,+\infty)$
Priors	$\mu_i \sim \text{Normal}(0,25), h_i \sim \text{Gamma}(0.1,1), \delta \sim \text{Uniform}[-1,1],$ $\upsilon-2 \sim \text{Exponential}(1), \eta-2 \sim \text{Exponential}(1)$
Clayton copula t marginals model (ctc)	
Distributional assumption: $z_{it} \sim t(0,1,\upsilon)$, Clayton copula function	
Log likelihood	$\frac{n}{2}\ln((1+\delta)^2 h_1 h_2) - (1+\delta)\sum_{t=1}^{n}(\ln F(z_{1t};\upsilon) + \ln F(z_{2t};\upsilon))$ $-\sum_{t=1}^{n}\left[(2+\frac{1}{\delta})\ln\left(F(z_{1t};\upsilon)^{-\delta} + F(z_{2t};\upsilon)^{-\delta} - 1\right) - \ln f(z_{1t};\upsilon) - \ln f(z_{2t};\upsilon)\right]$
Parameters	$\boldsymbol{\theta} = (\mu_1 \quad h_1 \quad \mu_2 \quad h_2 \quad \delta \quad \upsilon)^T, h_i \in (0,+\infty), \delta \in (0,+\infty), \upsilon \in (2,+\infty)$
Priors	$\mu_i \sim \text{Normal}(0,25), h_i \sim \text{Gamma}(0.1,1), \delta \sim \text{Gamma}(1,1), \upsilon-2 \sim \text{Exponential}(1)$

where μ_i is mean of return, σ_i is standard deviation, and $z_{it} = (r_{it}-\mu_i)/\sigma_i$ is normalized returns for $i = 1,2$. With different assumptions for marginal distribution of z_{it} and the Copula function, we obtain different Copula models. Particularly, we consider four Copula models in Hurn et al. (2020).

Let $h_i = 1/\sigma_i^2 > 0$ be the precision parameter, $F(z_{it};\upsilon)$, $f(z_{it};\upsilon)$ be the cumulative distribution function (CDF) and probability density function (PDF) of the t distribution with υ degrees of freedom ($\upsilon > 2$) respectively, $\Phi^{-1}(\cdot)$ be the quantile function of the standard normal distribution, $F^{-1}(\cdot;\eta)$ be the quantile function of the t distribution with η degrees of freedom ($\eta > 2$), $q_{\phi,it} = \Phi^{-1}(F(z_{it};\upsilon))$, $q_{f,it} = F^{-1}(F(z_{it};\upsilon);\eta)$. Given the above notations, the log likelihood function, parameters, prior distribution of parameters for considered Copula models are summarized in Table 4. For more model details and model property analysis, one can refer to Hurn et al. (2020).

Table 5
Model selection results for four copula models.

Model	$D\left(\bar{\theta}\right)$	P_D	DIC	C_2/n
gnc	31 378	5.20	31 389	−0.0006
gtc	29 689	5.69	29 700	−0.0014
ttc	29 305	5.60	29 316	−0.0016
ctc	30 490	5.67	30 502	−0.0016

The data we use are daily log returns on the S&P 100 and S&P 600 Indices from 17 August 1995 to 28 December 2018 and the sample size is $n = 5893$. The MCMC output is obtained using "mcmc" package in R, where total iteration is 100,000 times, burn-in iteration is the first 50,000 times and one effective sample is taken for every five samples in the remaining iterations, resulting in 10,000 samples for each parameter from their posterior distributions.

To compute P_D, we need to evaluate $E_{\theta|\mathbf{y}}\left[\ln p(\mathbf{y}|\theta)\right]$. Since it does not has closed form, similar to the discrete choice model example, we approximate it by MCMC output,

$$E_{\theta|\mathbf{y}}\left[\ln p(\mathbf{y}|\theta)\right] \approx \frac{1}{M}\sum_{m=1}^{M} \ln p\left(\mathbf{y}|\theta^{(m)}\right)$$

where M is the number of effective draws.

To compare these four Copula models, we calculate $D\left(\bar{\theta}\right)$, P_D and DIC for all candidate models based on the 10,000 effective draws. The results are summarized in Table 5.

Based on the DIC estimates reported in Table 5, the t copula t marginals model (ttc) outperforms the other models by a large margin. Its DIC is estimated to be around 29316, being the smallest among the candidate models. Then follows the second best model, i.e., the Gaussian copula t marginals model (gtc), with DIC being around 29700. The performance of the remaining Clayton copula t marginals model (ctc) and Gaussian copula normal marginals model (gnc) are not satisfactory. These results are consistent with existing empirical facts that asset returns exhibit heavy tails, and that the two asset returns we choose are expected to have strong tail dependence.

The estimated values of P_D are close to the number of model parameters. This is because we employ vague prior distributions for parameters. In the last column in Table 5, we give the estimated prior effects on P_D, i.e., $C_2/n = n^{-1}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla^2\hat{\pi}\right]$. The prior effects are small.[7]

5. Conclusion

This paper provides a rigorous decision-theoretic justification of DIC based on a set of regularity conditions. To do so, we first specify the underlying loss function to be the KL divergence between the true DGP and plug-in predictive distribution $p\left(\mathbf{y}_{rep}|\bar{\theta}_n(\mathbf{y})\right)$. This loss function is slightly different from that in AIC by using the posterior mean $\bar{\theta}_n(\mathbf{y})$ as the estimator of θ rather than QML. As a result, DIC is easy to calculate when the MCMC output is available.

Under a set of regularity conditions, we then show that DIC is an asymptotically unbiased estimator of the expected loss function as $n \to \infty$. Moreover, we develop expansions to DIC and the penalty term based on the high-order Laplace approximations. These expansions allow us to easily see the effect of prior on DIC and the penalty term. We illustrate how to use DIC to compare some non-nested models widely used in economics.

Although DIC is expected to select the "best" model among a set of candidate models to predict replicate data, as far as out-of-sample forecasting is concerned, model combination has been found to be a fruitful alternative approach to model selection. Useful Bayesian model combination techniques include Bayesian model averaging, Bayesian predictive synthesis, and Bayesian predictive decision synthesis, see Hoeting et al. (1999), McAlinn and West (2019) and Tallman and West (2024) for more details. The topic of Bayesian model combination is beyond the scope of the present paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[7] To obtain a positive prior effect, the $\nabla^2\hat{\pi}$ need to be negative definite. However, this is not necessarily satisfied in practice. Here the estimated C_2/n is negative because $\nabla^2\hat{\pi}$ is not negative definite.

Appendix A

A.1. A proof of Theorem 2.1

We write $\mathbf{H}_n\left(\theta_n^p\right)$ as $\mathbf{H}_n$, $\mathbf{B}_n\left(\theta_n^p\right)$ as $\mathbf{B}_n$, and let $\mathbf{C}_n=\mathbf{H}_n^{-1}\mathbf{B}_n\mathbf{H}_n^{-1}$. Under Assumptions 1–10, we can show that

$$\overline{\theta}_n(\mathbf{y})=\widehat{\theta}_n(\mathbf{y})+O_p\left(n^{-1}\right) \tag{40}$$

by the stochastic expansion, for more details, see the proof of Lemma 2.3 in the online supplement. Then, we have

$$\overline{\theta}_n(\mathbf{y})=\theta_n^p+O_p\left(n^{-1/2}\right),$$

$$\frac{1}{\sqrt{n}}\mathbf{B}_n^{-1/2}\frac{\partial\ln p(\mathbf{y}_{rep}|\theta_n^p)}{\partial\theta}\xrightarrow{d}N\left(0,\mathbf{I}_P\right), \tag{41}$$

and

$$\mathbf{C}_n^{-1/2}\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\xrightarrow{d}N\left(0,\mathbf{I}_P\right). \tag{42}$$

Note that

$$\begin{aligned}
&E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right)\\
=&\Big[\underbrace{E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right)}_{(T_1)}\Big]\\
+&\Big[\underbrace{E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)\right)-E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right)}_{(T_2)}\Big]\\
+&\Big[\underbrace{E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right)-E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)\right)}_{(T_3)}\Big].
\end{aligned}$$

Now let us analyze T_2 and T_3. First, expanding $\ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)$ at $\overline{\theta}_n\left(\mathbf{y}_{rep}\right)$, we have

$$\begin{aligned}
&\ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)\\
=&\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)+\frac{\partial\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\\
&+\frac{1}{2}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)'\frac{\partial^2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\\
&+\frac{1}{6}\left[\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\otimes\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right]'\frac{\partial^3\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'\partial\theta}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)
\end{aligned} \tag{43}$$

where $\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)$ lies between θ_n^p and $\overline{\theta}_n\left(\mathbf{y}_{rep}\right)$. Note that the last term can be written as

$$RT_{1,n}=\frac{1}{6}\frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\otimes\sqrt{n}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right]'\frac{1}{n}\sum_{t=1}^n\nabla^3 l_t\left(\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\sqrt{n}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) \tag{44}$$

where $\sqrt{n}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)=O_p(1)$ by Assumptions 1–10 and

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{t=1}^n\nabla^3 l_t\left(\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\right\|&\leq\frac{1}{n}\sum_{t=1}^n\left\|\nabla^3 l_t\left(\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\right\|\leq\frac{1}{n}\sum_{t=1}^n\sup_{\theta\in\Theta}\left\|\nabla^j l_t(\theta)\right\|\\
&\leqslant\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)
\end{aligned}$$

by Assumption 5. It can be shown that

$$P\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)>C\right)\leq\frac{\frac{1}{n}\sum_{t=1}^n E\left(M_t(\mathbf{y}^t)\right)}{C}\leq\frac{\sup_t E\left(M_t(\mathbf{y}^t)\right)}{C}\leq\frac{M}{C}$$

by the Markov inequality. Let $\varepsilon=M/C$, for any ε, there exists a constant $C=M/\varepsilon$ such that

$$P\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)>C\right)\leq\varepsilon.$$

Thus, $\frac{1}{n}\sum_{t=1}^{n} M_t(\mathbf{y}^t) = O_p(1)$ and $\left\|\frac{1}{n}\sum_{t=1}^{n} \nabla^3 l_t\left(\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\right\| = O_p(1)$. Hence, we have $RT_{1,n} = O_p\left(n^{-1/2}\right)$.

We can rewrite (43) as

$$\begin{aligned}
&\ln p\left(\mathbf{y}_{rep}|\theta_n^p\right) \\
&= \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) + \frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) \\
&\quad + \frac{1}{2}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta \partial \theta'}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) + RT_{1,n} \\
&= \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) + \frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) \\
&\quad + \frac{1}{2}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta \partial \theta'}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) \\
&\quad + \left(\frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'} - \frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'}\right)\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) + RT_{1,n} \\
&= \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) + \frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) \\
&\quad + \frac{1}{2}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta \partial \theta'}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) + RT_n \\
&= \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) + \frac{1}{2}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta \partial \theta'}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) + RT_n
\end{aligned}$$

from (40), where $RT_n = RT_{1,n} + RT_{2,n}$ with

$$RT_{2,n} = \left(\frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'} - \frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'}\right)\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right). \tag{45}$$

We can rewrite the first term on the right-hand side of (45) as

$$\begin{aligned}
&\left(\frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'} - \frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta'}\right) \\
&= \frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)\right)}{\partial \theta \partial \theta'} n\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right) - \widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right) = O_p(1),
\end{aligned}$$

where $\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)$ lies between $\overline{\theta}_n\left(\mathbf{y}_{rep}\right)$ and $\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)$. Thus,

$$RT_{2,n} = O_p(1)\, O_p\left(n^{-1/2}\right) = O_p\left(n^{-1/2}\right).$$

Hence, we have

$$RT_n = RT_{1,n} + RT_{2,n} = O_p\left(n^{-1/2}\right). \tag{46}$$

Now we will consider the expectation of the norm of $RT_{1,n}$ and $RT_{2,n}$. For $RT_{1,n}$, we first consider the term

$$\left[\sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) \otimes \sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right]' \frac{1}{n}\sum_{t=1}^{n} \nabla^3 l_t\left(\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right), \tag{47}$$

and try to prove that the expectation of (47) is bounded. It can be shown that

$$\begin{aligned}
&E\left[\left\|\left[\sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) \otimes \sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right]' \frac{1}{n}\sum_{t=1}^{n} \nabla^3 l_t\left(\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|\right] \\
&\leq \left(E\left[\left\|\left[\sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right) \otimes \sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right]'\right\|^2\right]\right)^{1/2} \\
&\quad \times \left(E\left[\left\|\frac{1}{n}\sum_{t=1}^{n} \nabla^3 l_t\left(\overline{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\sqrt{n}\left(\theta_n^p - \overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^2\right]\right)^{1/2}
\end{aligned} \tag{48}$$

$$= \left(E\left[\left\|\sqrt{n}\left(\theta_n^p - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right]\right)^{1/2}\left(E\left[\left\|\frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\bar{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\sqrt{n}\left(\theta_n^p - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^2\right]\right)^{1/2}$$

by the Cauchy–Schwarz Inequality and the fact that

$$\left\|\left[\sqrt{n}\left(\theta_n^p - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\otimes\sqrt{n}\left(\theta_n^p - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right]'\right\| = \left\|\sqrt{n}\left(\theta_n^p - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^2.$$

To prove that (48) is bounded, we need to prove that

$$E\left[\left\|\sqrt{n}\left(\theta_n^p - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right] \tag{49}$$

and

$$E\left[\left\|\frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\bar{\theta}_n^*\left(\mathbf{y}_{rep}\right)\right)\sqrt{n}\left(\theta_n^p - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^2\right] \tag{50}$$

are both bounded.

For (49), we have

$$\begin{aligned}
&\left(E\left[\left\|\sqrt{n}\left(\theta_n^p - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right]\right)^{1/4}\\
&= \left(E\left[\left\|\sqrt{n}\left(\theta_n^p - \hat{\theta}_n\left(\mathbf{y}_{rep}\right) + \hat{\theta}_n\left(\mathbf{y}_{rep}\right) - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right]\right)^{1/4}\\
&\leq \left(E\left[\left(\left\|\sqrt{n}\left(\theta_n^p - \hat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\| + \left\|\sqrt{n}\left(\hat{\theta}_n\left(\mathbf{y}_{rep}\right) - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|\right)^4\right]\right)^{1/4}\\
&\leq \left(E\left[\left\|\sqrt{n}\left(\theta_n^p - \hat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right]\right)^{1/4} + \left(E\left[\left\|\sqrt{n}\left(\hat{\theta}_n\left(\mathbf{y}_{rep}\right) - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right]\right)^{1/4}
\end{aligned}$$

by the triangular inequality and the Minkowski inequality. To prove that (49) is bounded, it is suffice to show

$$E\left[\left\|\sqrt{n}\left(\theta_n^p - \hat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right] \tag{51}$$

and

$$E\left[\left\|\sqrt{n}\left(\hat{\theta}_n\left(\mathbf{y}_{rep}\right) - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right] \tag{52}$$

are both bounded. Li et al. (2024) have proved that

$$E\left[\left\|\sqrt{n}\left(\theta_n^p - \hat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right] < \infty \tag{53}$$

under Assumptions 1–10.

For (52), following Proposition 6.1 of Huggins et al. (2018), and Lemma A.1 if we use $\hat{\theta}_n\left(\mathbf{y}_{rep}\right)$ to approximate the posterior mean $\bar{\theta}_n\left(\mathbf{y}_{rep}\right)$, the bound of the approximate error is

$$\left\|\sqrt{n}\left(\hat{\theta}_n\left(\mathbf{y}_{rep}\right) - \bar{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\| \leq \frac{1}{\sqrt{n}}C^*, \tag{54}$$

where

$$C^* = \alpha\left(\mathbf{y}^n\right)\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)^2\left(\left(\sum_{j=1}^P\left|\lambda_{\bar{\mathbf{H}}_n^*,j}\right|\right)^2 + \sum_{j=1}^P\lambda_{\bar{\mathbf{H}}_n^*,j}^2\right)\\ +2\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)\left(\sum_{j=1}^P\left|\lambda_{\bar{\mathbf{H}}_n^*,j}\right|\right) + M_0^2\end{array}\right]^{1/2},$$

$\lambda_{\bar{\mathbf{H}}_n^*,j}$ are the eigenvalues of $\bar{\mathbf{H}}_n^* := \bar{\mathbf{H}}_n\left(\hat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)$, M_0 is a finite constant. It can be shown that

$$\begin{aligned}
C^* &= \alpha\left(\mathbf{y}^n\right)\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)^2\left(\left(\sum_{j=1}^P\left|\lambda_{\bar{\mathbf{H}}_n^*,j}\right|\right)^2 + \sum_{j=1}^P\lambda_{\bar{\mathbf{H}}_n^*,j}^2\right)\\ +2\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)\left(\sum_{j=1}^P\left|\lambda_{\bar{\mathbf{H}}_n^*,j}\right|\right) + M_0^2\end{array}\right]^{1/2}\\
&\leq \alpha\left(\mathbf{y}^n\right)\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)^2\left(P^2\times\rho\left(\bar{\mathbf{H}}_n^*\right)^2 + P\times\rho\left(\bar{\mathbf{H}}_n^*\right)^2\right)\\ +2\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)\left(P\times\rho\left(\bar{\mathbf{H}}_n^*\right)\right) + M_0^2\end{array}\right]^{1/2}\\
&= \alpha\left(\mathbf{y}^n\right)\left[\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)^2\left(P^2+P\right)\times\rho\left(\bar{\mathbf{H}}_n^*\right)^2 + 2\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)\left(P\times\rho\left(\bar{\mathbf{H}}_n^*\right)\right) + M_0^2\right]^{1/2}
\end{aligned}$$

$$\leq \alpha\left(\mathbf{y}^{n}\right)\left[\left(\frac{1}{n}\sum_{t=1}^{n}M_{t}(\mathbf{y}^{t})\right)^{2}\left(P^{2}+P\right)\times\left\|\bar{\mathbf{H}}_{n}^{*}\right\|^{2}+2\left(\frac{1}{n}\sum_{t=1}^{n}M_{t}(\mathbf{y}^{t})\right)\times P\times\left\|\bar{\mathbf{H}}_{n}^{*}\right\|+M_{0}^{2}\right]^{1/2},$$

where $\rho\left(\bar{\mathbf{H}}_{n}^{*}\right)=\max_{j}\left|\lambda_{\bar{\mathbf{H}}_{n}^{*},j}\right|$ is the spectral radius of $\bar{\mathbf{H}}_{n}^{*}$ that is smaller than $\|\bar{\mathbf{H}}_{n}^{*}\|$. Therefore, (52) is bounded by

$$E\left[\left\|\sqrt{n}\left(\widehat{\theta}_{n}\left(\mathbf{y}_{rep}\right)-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{4}\right]\leq\frac{1}{n^{2}}E\left(C^{*4}\right)=O\left(n^{-2}\right)<\infty \tag{55}$$

by Assumption 5 since

$$\begin{aligned}E\left(C^{*4}\right)&=E\left[\alpha\left(\mathbf{y}^{n}\right)^{4}\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^{n}M_{t}(\mathbf{y}^{t})\right)^{2}\left(P^{2}+P\right)\times\left\|\bar{\mathbf{H}}_{n}^{*}\right\|^{2}\\+2\left(\frac{1}{n}\sum_{t=1}^{n}M_{t}(\mathbf{y}^{t})\right)\times P\times\left\|\bar{\mathbf{H}}_{n}^{*}\right\|+M_{0}^{2}\end{array}\right]^{2}\right]\\&\leq\left(E\left[\alpha\left(\mathbf{y}^{n}\right)^{8}\right]\right)^{1/2}\left(E\left[\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^{n}M_{t}(\mathbf{y}^{t})\right)^{2}\left(P^{2}+P\right)\times\left\|\bar{\mathbf{H}}_{n}^{*}\right\|^{2}\\+2\left(\frac{1}{n}\sum_{t=1}^{n}M_{t}(\mathbf{y}^{t})\right)\times P\times\left\|\bar{\mathbf{H}}_{n}^{*}\right\|+M_{0}^{2}\end{array}\right]^{4}\right]\right)^{1/2}.\end{aligned}$$

Thus, from (53) and (55), we have

$$\begin{aligned}&\left(E\left[\left\|\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{4}\right]\right)^{1/4}\\&\leq\left(E\left[\left\|\sqrt{n}\left(\theta_{n}^{p}-\widehat{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{4}\right]\right)^{1/4}+\left(E\left[\left\|\sqrt{n}\left(\widehat{\theta}_{n}\left(\mathbf{y}_{rep}\right)-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{4}\right]\right)^{1/4}\\&<\infty.\end{aligned} \tag{56}$$

For (50), we have

$$\begin{aligned}&E\left[\left\|\frac{1}{n}\sum_{t=1}^{n}\nabla^{3}l_{t}\left(\bar{\theta}_{n}^{*}\left(\mathbf{y}_{rep}\right)\right)\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{2}\right]\\&\leq E\left[\left\|\frac{1}{n}\sum_{t=1}^{n}\nabla^{3}l_{t}\left(\bar{\theta}_{n}^{*}\left(\mathbf{y}_{rep}\right)\right)\right\|^{2}\left\|\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{2}\right]\\&\leq\left(E\left[\left\|\frac{1}{n}\sum_{t=1}^{n}\nabla^{3}l_{t}\left(\bar{\theta}_{n}^{*}\left(\mathbf{y}_{rep}\right)\right)\right\|^{4}\right]\right)^{1/2}\left(E\left[\left\|\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{4}\right]\right)^{1/2}\\&<\infty.\end{aligned} \tag{57}$$

by Assumption 5 and (56). Thus, from (47), (48), (56) and (57), we have

$$\begin{aligned}&E\left\|RT_{1,n}\right\|\\&\leq\frac{1}{6}\frac{1}{\sqrt{n}}\left(E\left[\left\|\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{4}\right]\right)^{1/4}\\&\times\left(E\left[\left\|\frac{1}{n}\sum_{t=1}^{n}\nabla^{3}l_{t}\left(\bar{\theta}_{n}^{*}\left(\mathbf{y}_{rep}\right)\right)\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{2}\right]\right)^{1/4}\\&=o(1).\end{aligned} \tag{58}$$

For $RT_{2,n}$, we have

$$\begin{aligned}&E\left\|RT_{2,n}\right\|\\&\leq E\left[\left\|\frac{1}{\sqrt{n}}\left(\frac{\partial\ln p\left(\mathbf{y}_{rep}|\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}-\frac{\partial\ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}\right)\right\|\left\|\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|\right]\\&\leq\left(E\left[\left\|\frac{1}{\sqrt{n}}\left(\frac{\partial\ln p\left(\mathbf{y}_{rep}|\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}-\frac{\partial\ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}\right)\right\|^{2}\right]\right)^{1/2}\\&\times\left(E\left[\left\|\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{2}\right]\right)^{1/2},\end{aligned} \tag{59}$$

where

$$E\left[\left\|\sqrt{n}\left(\theta_{n}^{p}-\bar{\theta}_{n}\left(\mathbf{y}_{rep}\right)\right)\right\|^{2}\right]<\infty$$

by (56). For the first term in the right-hand side of (59)

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}-\frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}\right)\\
&=\frac{1}{\sqrt{n}}\frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'}\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right)-\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\\
&=\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'}\sqrt{n}\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right)-\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right),
\end{aligned}
$$

where $\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)$ lies between $\overline{\theta}_n\left(\mathbf{y}_{rep}\right)$ and $\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)$. Thus, we have

$$
\begin{aligned}
&E\left[\left\|\frac{1}{\sqrt{n}}\left(\frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}-\frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}\right)\right\|^2\right]\\
&=E\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'}\sqrt{n}\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right)-\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^2\right]\\
&\le E\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'}\right\|^2\left\|\sqrt{n}\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right)-\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^2\right]\\
&\le\left(E\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'}\right\|^4\right]\right)^{1/2}\left(E\left[\left\|\sqrt{n}\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right)-\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right]\right)^{1/2}.
\end{aligned}
$$

By Assumption 5 and (55), we have

$$
E\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n^{\#}\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'}\right\|^4\right]<\infty,
$$

and

$$
E\left[\left\|\sqrt{n}\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right)-\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^4\right]=O\left(n^{-2}\right).
$$

Hence,

$$
E\left[\left\|\frac{1}{\sqrt{n}}\left(\frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}-\frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}\right)\right\|^2\right]=o(1).
$$

So we have

$$
\begin{aligned}
&E\left\|RT_{2,n}\right\| \qquad (60)\\
&\le\left(E\left[\left\|\frac{1}{\sqrt{n}}\left(\frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}-\frac{\partial \ln p\left(\mathbf{y}_{rep}|\widehat{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta'}\right)\right\|^2\right]\right)^{1/2}\\
&\times\left(E\left[\left\|\sqrt{n}\left(\theta_n^p-\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right\|^2\right]\right)^{1/2}\\
&=o(1).
\end{aligned}
$$

From (58) and (60), it can be shown that

$$
E\left\|RT_n\right\|\le E\left\|RT_{1,n}\right\|+E\left\|RT_{2,n}\right\|=o(1).
$$

We can further get

$$
T_2 = E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)+2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right]
$$

$$
\begin{aligned}
&= E_{\mathbf{y}} E_{\mathbf{y}_{rep}} \left[-\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right) - \theta_n^p\right)' \frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'} \left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right) - \theta_n^p\right) + RT_n \right] \\
&= E_{\mathbf{y}_{rep}} \left[-\left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right) - \theta_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)}{\partial\theta\partial\theta'} \left(\overline{\theta}_n\left(\mathbf{y}_{rep}\right) - \theta_n^p\right) \right] + o(1) \\
&= E_{\mathbf{y}} \left[-\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'} \left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \right] + o(1).
\end{aligned}
$$

Next, we expand $\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)$ at θ_n^p,

$$
\begin{aligned}
\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right) &= \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right) + \frac{\partial \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta'}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \\
&\quad + \frac{1}{2}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta\partial\theta'}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) + o_p(1).
\end{aligned}
$$

Substituting the above expansion into T_3, we have

$$
\begin{aligned}
T_3 &= E_{\mathbf{y}} E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right] - E_{\mathbf{y}} E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)\right] \\
&= E_{\mathbf{y}} E_{\mathbf{y}_{rep}} \left[\begin{array}{c} -2\frac{\partial \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta'}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) - \\ \left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta\partial\theta'}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) + o_p(1) \end{array} \right] \\
&= E_{\mathbf{y}} E_{\mathbf{y}_{rep}} \left[-2\frac{\partial \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta'}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \right] \\
&\quad + E_{\mathbf{y}} E_{\mathbf{y}_{rep}} \left[-\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta\partial\theta'}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \right] + o(1) \\
&= -2E_{\mathbf{y}_{rep}}\left(\frac{\partial \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta'}\right) E_{\mathbf{y}}\left[\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)\right] \\
&\quad + E_{\mathbf{y}}\left[-\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)' E_{\mathbf{y}_{rep}}\left(\frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \right] + o(1) \\
&= E_{\mathbf{y}}\left[-\sqrt{n}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)' E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p(\mathbf{y}|\theta_n^p)}{\partial\theta\partial\theta'}\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \right] + o(1),
\end{aligned}
$$

since

$$
E_{\mathbf{y}} E_{\mathbf{y}_{rep}}\left[-2\frac{\partial \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta'}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)\right] = E_{\mathbf{y}_{rep}}\left[-2\frac{\partial \ln p\left(\mathbf{y}_{rep}|\theta_n^p\right)}{\partial\theta'}\right] E_{\mathbf{y}}\left[\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)\right] = 0
$$

by (41), (42), and the dominated convergence theorem.

We can rewrite T_2 as

$$
\begin{aligned}
T_2 &= E_{\mathbf{y}}\left[-\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \right] + o(1) \\
&= E_{\mathbf{y}}\left[-\sqrt{n}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)' \frac{1}{n} E_{\mathbf{y}}\left(\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \right] \\
&\quad + E_{\mathbf{y}}\left[\begin{array}{c} -\sqrt{n}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)'\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'} - E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\right) \\ \times\sqrt{n}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \end{array} \right] \\
&\quad + o(1)
\end{aligned}
$$

where

$$
E_{\mathbf{y}}\left[\begin{array}{c} -\sqrt{n}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right)'\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'} - E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\right) \\ \times\sqrt{n}\left(\overline{\theta}_n(\mathbf{y}) - \theta_n^p\right) \end{array} \right] \tag{61}
$$

$$\leq E_{\mathbf{y}}\left[\left\|\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right\|^2\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\right\|\right]$$
$$\leq \left(E_{\mathbf{y}}\left[\left\|\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right\|^4\right]\right)^{1/2}$$
$$\times\left(E_{\mathbf{y}}\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\right\|^2\right]\right)^{1/2}.$$

In (61), we have

$$E_{\mathbf{y}}\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\right\|^2\right] \tag{62}$$
$$= E_{\mathbf{y}}\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-\bar{\mathbf{H}}_n\left(\theta_n^p\right)+\bar{\mathbf{H}}_n\left(\theta_n^p\right)-\mathbf{H}_n\right\|^2\right]$$
$$\leq \left[\left(E_{\mathbf{y}}\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-\bar{\mathbf{H}}_n\left(\theta_n^p\right)\right\|^2\right]\right)^{1/2}+\left(E_{\mathbf{y}}\left[\left\|\bar{\mathbf{H}}_n\left(\theta_n^p\right)-\mathbf{H}_n\right\|^2\right]\right)^{1/2}\right]^2.$$

The first term of (62) can be written as

$$vec\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-\bar{\mathbf{H}}_n\left(\theta_n^p\right)\right)$$
$$= vec\left(\bar{\mathbf{H}}_n\left(\overline{\theta}_n(\mathbf{y})\right)\right)-vec\left(\bar{\mathbf{H}}_n\left(\theta_n^p\right)\right)=\frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widetilde{\theta}_n^{**}(\mathbf{y})\right)\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)$$
$$=\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widetilde{\theta}_n^{**}(\mathbf{y})\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)$$

by vectorization and the Taylor expansion, where $\widetilde{\theta}_n^{**}(\mathbf{y})$ lies between $\overline{\theta}_n(\mathbf{y})$ and θ_n^p. Thus,

$$E_{\mathbf{y}}\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-\bar{\mathbf{H}}_n\left(\theta_n^p\right)\right\|^2\right] \tag{63}$$
$$\leq \frac{1}{n}E_{\mathbf{y}}\left[\left\|\frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widetilde{\theta}_n^{**}(\mathbf{y})\right)\right\|^2\left\|\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right\|^2\right]$$
$$\leq \frac{1}{n}\left(E_{\mathbf{y}}\left[\left\|\frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widetilde{\theta}_n^{**}(\mathbf{y})\right)\right\|^4\right]\right)^{1/2}\left(E_{\mathbf{y}}\left[\left\|\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right\|^4\right]\right)^{1/2}$$
$$=O\left(\frac{1}{n}\right)$$

by Assumption 5 and (56). The second term of (62) can be written as

$$E_{\mathbf{y}}\left[\left\|\bar{\mathbf{H}}_n\left(\theta_n^p\right)-\mathbf{H}_n\right\|^2\right]\leq\frac{1}{n}E_{\mathbf{y}}\left[\left\|\sqrt{n}\left(\bar{\mathbf{H}}_n\left(\theta_n^p\right)-\mathbf{H}_n\right)\right\|^2\right]=O\left(\frac{1}{n}\right) \tag{64}$$

by Assumptions 1–9. From (63) and (62), it can be shown that

$$E_{\mathbf{y}}\left[\left\|\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\right\|^2\right]=o(1).$$

Thus, we have

$$E_{\mathbf{y}}\left[\begin{array}{c}-\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}-E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\right)\\ \times\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\end{array}\right]=o(1). \tag{65}$$

We can further rewrite T_2 as

$$\begin{aligned} T_2 &= E_{\mathbf{y}}\left[-\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta\partial\theta'}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right]+o(1) \\ &= E_{\mathbf{y}}\left[-\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)' \frac{1}{n}E_{\mathbf{y}}\left(\frac{\partial^2 \ln p\left(\mathbf{y}|\theta_n^p\right)}{\partial\theta\partial\theta'}\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right]+o(1) \\ &= T_3+o(1). \end{aligned}$$

Hence, we only need to analyze T_3. Note that

$$\begin{aligned} T_3 &= E_{\mathbf{y}}\left[-\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)' E_{\mathbf{y}}\left(-\frac{1}{n}\frac{\partial^2 \ln p(\mathbf{y}|\theta_n^p)}{\partial\theta\partial\theta'}\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right]+o(1) \\ &= E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'(-\mathbf{H}_n)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right]+o(1) \\ &= E_{\mathbf{y}}\left[\left(\mathbf{C}_n^{-1/2}\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right)'\mathbf{C}_n^{1/2}(-\mathbf{H}_n)\mathbf{C}_n^{1/2}\mathbf{C}_n^{-1/2}\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\right]+o(1) \\ &= E_{\mathbf{y}}\left\{\mathbf{tr}\left[\mathbf{H}_n\mathbf{C}_n^{1/2}\mathbf{C}_n^{-1/2}\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\mathbf{C}_n^{-1/2}\mathbf{C}_n^{1/2}\right]\right\}+o(1) \\ &= \mathbf{tr}\left\{(-\mathbf{H}_n)\mathbf{C}_n^{1/2}E_{\mathbf{y}}\left[\mathbf{C}_n^{-1/2}\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\mathbf{C}_n^{-1/2}\right]\mathbf{C}_n^{1/2}\right\}+o(1). \end{aligned} \tag{66}$$

In (66), we have

$$\begin{aligned} & E_{\mathbf{y}}\left[\mathbf{C}_n^{-1/2}\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\mathbf{C}_n^{-1/2}\right] \\ &= \mathbf{C}_n^{-1/2}E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right]\mathbf{C}_n^{-1/2} \\ &= \mathbf{C}_n^{-1/2}E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right]\mathbf{C}_n^{-1/2}, \end{aligned}$$

where

$$\begin{aligned} & E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right] \\ &= E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\hat{\theta}_n(\mathbf{y})+\hat{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\hat{\theta}_n(\mathbf{y})+\hat{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right] \\ &= E_{\mathbf{y}}\left[\sqrt{n}\left(\hat{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\hat{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right]+E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\hat{\theta}_n(\mathbf{y})\right)\sqrt{n}\left(\hat{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right] \\ &\quad + E_{\mathbf{y}}\left[\sqrt{n}\left(\hat{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\hat{\theta}_n(\mathbf{y})\right)'\right] \\ &\quad + E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\hat{\theta}_n(\mathbf{y})\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\hat{\theta}_n(\mathbf{y})\right)'\right]. \end{aligned} \tag{67}$$

In (67), it can be shown that the last three terms are all $o(1)$ because of (53) and (55). For the first term, we know that

$$E_{\mathbf{y}}\left[\sqrt{n}\left(\hat{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\hat{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right]=\mathbf{H}_n^{-1}\mathbf{B}_n\mathbf{H}_n^{-1}+o(1)=C_n+o(1)$$

by Li et al. (2024). Hence,

$$E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right]=\mathbf{H}_n^{-1}\mathbf{B}_n\mathbf{H}_n^{-1}+o(1)=C_n+o(1).$$

It can be shown that

$$\begin{aligned} T_3 &= \mathbf{tr}\left\{(-\mathbf{H}_n)\mathbf{C}_n^{1/2}\mathbf{C}_n^{-1/2}E_{\mathbf{y}}\left[\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)\sqrt{n}\left(\overline{\theta}_n(\mathbf{y})-\theta_n^p\right)'\right]\mathbf{C}_n^{-1/2}\mathbf{C}_n^{1/2}\right\}+o(1) \\ &= \mathbf{tr}\left\{(-\mathbf{H}_n)\mathbf{C}_n^{1/2}\mathbf{C}_n^{-1/2}\mathbf{C}_n\mathbf{C}_n^{-1/2}\mathbf{C}_n^{1/2}\right\}+o(1) \\ &= \mathbf{tr}\left((-\mathbf{H}_n)\mathbf{C}_n^{1/2}\mathbf{C}_n^{1/2}\right)+o(1)=\mathbf{tr}\left((-\mathbf{H}_n)\mathbf{C}_n\right)+o(1) \\ &= \mathbf{tr}\left((-\mathbf{H}_n)(-\mathbf{H}_n)^{-1}\mathbf{B}_n(-\mathbf{H}_n)^{-1}\right)+o(1) \\ &= \mathbf{tr}\left(\mathbf{B}_n(-\mathbf{H}_n)^{-1}\right)+o(1). \end{aligned}$$

and

$$E_{\mathbf{y}}\left[E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right)\right] \tag{68}$$

$$
\begin{aligned}
&= E_{\mathbf{y}}\left[E_{\mathbf{y}_{rep}}\left(-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n\left(\mathbf{y}_{rep}\right)\right)\right)\right] + 2\mathbf{tr}\left(\mathbf{B}_n\left(-\mathbf{H}_n\right)^{-1}\right) + o(1) \\
&= E_{\mathbf{y}}\left[E_{\mathbf{y}}\left(-2\ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)\right)\right] + 2\mathbf{tr}\left(\mathbf{B}_n\left(-\mathbf{H}_n\right)^{-1}\right) + o(1) \\
&= E_{\mathbf{y}}\left[-2\ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)\right] + 2\mathbf{tr}\left(\mathbf{B}_n\left(-\mathbf{H}_n\right)^{-1}\right) + o(1) \\
&= E_{\mathbf{y}}\left[-2\ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)\right] + 2P + o(1).
\end{aligned}
$$

The last step is due to Assumption 9.

Note that

$$
\begin{aligned}
P_D &= \overline{D(\theta)} - D\left(\overline{\theta}_n(\mathbf{y})\right) \\
&= -2\int\left[\ln p(\mathbf{y}|\theta) - \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)\right] p(\theta|\mathbf{y})\, d\theta.
\end{aligned}
$$

By the Taylor expansion

$$
\begin{aligned}
&\ln p(\mathbf{y}|\theta) - \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right) \\
&= \frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)}{\partial\theta'}\left(\theta - \overline{\theta}_n(\mathbf{y})\right) + \frac{1}{2}\left(\theta - \overline{\theta}_n(\mathbf{y})\right)'\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n^{\#\#}(\mathbf{y},\theta)\right)}{\partial\theta\partial\theta'}\left(\theta - \overline{\theta}_n(\mathbf{y})\right),
\end{aligned}
$$

where $\overline{\theta}_n^{\#\#}(\mathbf{y},\theta)$ lies between θ and $\overline{\theta}_n(\mathbf{y})$. Thus, we have

$$
\begin{aligned}
P_D &= -2\int\left[\ln p(\mathbf{y}|\theta) - \ln p\left(\mathbf{y}|\overline{\theta}_n(\mathbf{y})\right)\right] p(\theta|\mathbf{y})\, d\theta \\
&= -\int\left(\theta - \overline{\theta}_n(\mathbf{y})\right)'\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n^{\#\#}(\mathbf{y},\theta)\right)}{\partial\theta\partial\theta'}\left(\theta - \overline{\theta}_n(\mathbf{y})\right) p(\theta|\mathbf{y})\, d\theta \\
&= -\int\sqrt{n}\left(\theta - \overline{\theta}_n(\mathbf{y})\right)'\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n^{\#\#}(\mathbf{y},\theta)\right)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta - \overline{\theta}_n(\mathbf{y})\right) p(\theta|\mathbf{y})\, d\theta.
\end{aligned}
\tag{69}
$$

From (69), we have

$$
\begin{aligned}
\left\|P_D\right\| &= \left\|\int\sqrt{n}\left(\theta - \overline{\theta}_n(\mathbf{y})\right)'\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\theta}_n^{\#\#}(\mathbf{y},\theta)\right)}{\partial\theta\partial\theta'}\sqrt{n}\left(\theta - \overline{\theta}_n(\mathbf{y})\right) p(\theta|\mathbf{y})\, d\theta\right\| \\
&\le \left\|\frac{1}{n}\sum_{t=1}^{t} M_t\right\|\left\|\int\sqrt{n}\left(\theta - \overline{\theta}_n(\mathbf{y})\right)\sqrt{n}\left(\theta - \overline{\theta}_n(\mathbf{y})\right)' p(\theta|\mathbf{y})\, d\theta\right\| \\
&\le \left\|\frac{1}{n}\sum_{t=1}^{t} M_t\right\|\left\|nV\left(\overline{\theta}_n(\mathbf{y})\right)\right\|,
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left(\left\|\frac{1}{n}\sum_{t=1}^{t} M_t\right\|\left\|nV\left(\overline{\theta}_n(\mathbf{y})\right)\right\|\right) \\
&\le \left(E\left(\left\|\frac{1}{n}\sum_{t=1}^{t} M_t\right\|^2\right)\right)^{1/2}\left(E\left(\left\|nV\left(\overline{\theta}_n(\mathbf{y})\right)\right\|^2\right)\right)^{1/2}
\end{aligned}
\tag{70}
$$

by Assumption 5, where

$$
V\left(\overline{\theta}_n(\mathbf{y})\right) = \int\left(\theta - \overline{\theta}_n(\mathbf{y})\right)'\left(\theta - \overline{\theta}_n(\mathbf{y})\right) p(\theta|\mathbf{y})\, d\theta.
$$

It can be shown that

$$
\left\|V\left(\overline{\theta}_n(\mathbf{y})\right) - \left(-\frac{1}{n}\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right\| \le 3\left\|\frac{1}{n}\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right\|^{1/2} C^{**} + 5.25C^{**}
$$

where

$$
\begin{aligned}
C^{**} &\le \frac{\alpha(\mathbf{y}^n)}{n}\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^{n} M_t(\mathbf{y}^t)\right)^2\left(\left(\sum_{j=1}^{P}\left|\lambda_{\bar{\mathbf{H}}_n^{**},j}\right|\right)^2 + \sum_{j=1}^{P}\lambda^2_{\bar{\mathbf{H}}_n^{**},j}\right) \\ +2\left(\frac{1}{n}\sum_{t=1}^{n} M_t(\mathbf{y}^t)\right)\left(\sum_{j=1}^{P}\left|\lambda_{\bar{\mathbf{H}}_n^{**},j}\right|\right) + M_0^2\end{array}\right]^{1/2} \\
&\le \frac{\alpha(\mathbf{y}^n)}{n}\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^{n} M_t(\mathbf{y}^t)\right)^2\left(P^2\times\rho\left(\bar{\mathbf{H}}_n^{**}\right)^2 + P\times\rho\left(\left(\bar{\mathbf{H}}_n^{**}\right)^2\right)\right) \\ +2\left(\frac{1}{n}\sum_{t=1}^{n} M_t(\mathbf{y}^t)\right)\left(P\times\rho\left(\bar{\mathbf{H}}_n^{**}\right)\right) + M_0^2\end{array}\right]^{1/2}
\end{aligned}
$$

$$= \frac{\alpha(\mathbf{y}^n)}{n}\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)^2 (P^2+P)\times\rho\left(\bar{\mathbf{H}}_n^{**}\right)^2 + 2\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)\left(P\times\rho\left(\bar{\mathbf{H}}_n^{**}\right)\right)\\ +M_0^2\end{array}\right]^{1/2}$$

$$\leq \frac{\alpha(\mathbf{y}^n)}{n}\left[\begin{array}{c}\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)^2 (P^2+P)\times\|\bar{\mathbf{H}}_n^{**}\|^2 + 2\left(\frac{1}{n}\sum_{t=1}^n M_t(\mathbf{y}^t)\right)\times P\times\|\bar{\mathbf{H}}_n^{**}\|\\ +M_0^2\end{array}\right]^{1/2}$$

by Theorem 4.1, Proposition 6.2 in Huggins et al. (2018) and Lemma A.1 with $\bar{\mathbf{H}}_n^{**} = \bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)$. By Assumptions 5 and 6, we have

$$E\left(C^{**2}\right) = O\left(n^{-2}\right) < \infty,\ E\left(C^{**4}\right) = O\left(n^{-4}\right) < \infty.$$

Thus,

$$\begin{aligned}
&E\left(\left\|nV\left(\overline{\theta}_n(\mathbf{y})\right) - \left(-\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right\|^2\right)\\
&\leq n^2 E\left(\left[3\left\|\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right\|^{1/2} C^{**} + 5.25C^{**}\right]^2\right)\\
&\leq n^2\left[E\left(\left[9\left\|\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right\| C^{**2}\right]\right)^{1/2} + \left(E\left[5.25^2\|C^{**}\|^2\right]\right)^{1/2}\right]^2\\
&< \infty.
\end{aligned}$$

Hence, we can further get

$$\begin{aligned}
&E\left(\left\|nV\left(\overline{\theta}_n(\mathbf{y})\right)\right\|^2\right)\\
&= E\left(\left\|nV\left(\overline{\theta}_n(\mathbf{y})\right) - \left(-\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right) + \left(-\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right\|^2\right)\\
&\leq E\left(\left[\left\|nV\left(\overline{\theta}_n(\mathbf{y})\right) - \left(-\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right\| + \left\|\left(-\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right\|\right]^2\right)\\
&\leq \left[\left(E\left(\left\|nV\left(\overline{\theta}_n(\mathbf{y})\right) - \left(-\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right\|^2\right)\right)^{1/2} + \left(E\left(\left\|\left(-\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right\|^2\right)\right)^{1/2}\right]^2\\
&< \infty.
\end{aligned} \tag{71}$$

From (70) and (71), we have

$$E\left(P_D\right) = P + o(1) \tag{72}$$

by the dominated convergence theorem and $P_D = P + o_p(1)$ following Lemma 2.3.

Finally, by (68), the fact that $P_D = P + o_p(1)$, and (72), we have

$$\begin{aligned}
&E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[-2\ln p\left(\mathbf{y}_{rep}|\overline{\theta}_n(\mathbf{y})\right)\right] = E_{\mathbf{y}}\left[-2\ln p\left(\mathbf{y}|\overline{\theta}_n\right) + 2P\right] + o(1)\\
&= E_{\mathbf{y}}\left[D\left(\overline{\theta}_n\right) + 2P_D + o_p(1)\right] = E_{\mathbf{y}}\left[\mathrm{DIC} + o_p(1)\right] = E_{\mathbf{y}}\left[\mathrm{DIC}\right] + o(1).
\end{aligned}$$

Lemma A.1 (*Asymptotic Laplace Approximation Error*). *Let the Laplace approximation of the posterior distribution be*

$$\hat{\Pi}_{Laplace} = (2\pi)^{-P/2}\det\left(-n\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)\right)^{1/2}\exp\left(-\frac{1}{2}\left(\theta-\hat{\theta}_n\right)'\left(-n\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)\right)\left(\theta-\hat{\theta}_n\right)\right).$$

The, under Assumptions 1–10, the p-Wasserstein distance ($p = 2$) between $\hat{\Pi}_{Laplace}$ and the posterior density $p\left(\theta|\mathbf{y}_n\right)$ is bounded as

$$\mathcal{W}_p\left(\hat{\Pi}_{Laplace}, p\left(\theta|\mathbf{y}_n\right)\right) \leq \frac{\alpha(\mathbf{y}^n)\left(M^2L_2^2 + 2M_0ML_1 + M_0^2\right)}{n},$$

where $L_1 = \sum_{j=1}^P\left|\lambda_{\bar{\mathbf{H}}_n(\hat{\theta}_n),j}\right|$ and $L_2 = \left(2\sum_{j=1}^P\lambda^2_{\bar{\mathbf{H}}_n(\hat{\theta}_n),j} + L_1^2\right)^{1/2}$, $\left\{\lambda_{\bar{\mathbf{H}}_n(\hat{\theta}_n),j}\right\}$ are the eigenvalues of $\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)$, M, M_0 are two finite constants.

Proof. Let $b(\theta) = \nabla\ln p\left(\theta|\mathbf{y}_n\right)$ and $b_{\text{Laplace}}(\theta) = n\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)\left(\theta-\hat{\theta}_n\right)$. By the Taylor expansion, the ith component of $b(\theta)$ can be rewritten as

$$b_i(\theta) = \nabla\ln p\left(\theta|\mathbf{y}_n\right)_i + \nabla\left(\nabla\ln p\left(\hat{\theta}_n|\mathbf{y}_n\right)_i\right)'\left(\theta-\hat{\theta}_n\right)$$

$$
\begin{aligned}
&+ R\left(\nabla \ln p\left(\theta|\mathbf{y}_n\right)_i, \theta\right) \\
&= \nabla\pi\left(\hat{\theta}_n\right)_i + \nabla\left(\nabla \ln p\left(\hat{\theta}_n|\mathbf{y}_n\right)_i\right)'\left(\theta-\hat{\theta}_n\right) \\
&+ R\left(\nabla \ln p\left(\theta|\mathbf{y}_n\right)_i, \theta\right) \\
&= \nabla\pi\left(\hat{\theta}_n\right)_i + \nabla\left(\nabla\pi\left(\hat{\theta}_n\right)_i\right)\left(\theta-\hat{\theta}_n\right) \\
&+ \nabla\left(\nabla \ln p\left(\mathbf{y}_n|\hat{\theta}_n\right)_i\right)'\left(\theta-\hat{\theta}_n\right) + R\left(\nabla \ln p\left(\theta|\mathbf{y}_n\right)_i, \theta\right),
\end{aligned}
$$

where

$$
R(\phi,\theta) = \left(\theta-\hat{\theta}_n\right)'\left\{\int_0^1 (1-t)\nabla^2\phi\left(\hat{\theta}_n + t\left(\theta-\hat{\theta}_n\right)\right)\mathrm{d}t\right\}\left(\theta-\hat{\theta}_n\right).
$$

Note that $\nabla\pi_i$ and $\nabla\left(\nabla\pi_i\right)$ are both continuous functions on a compact set by Assumptions 1 and 10. Thus, there exits a finite upper bound M_0 for $\nabla\pi\left(\hat{\theta}_n\right)_i + \nabla\left(\nabla\pi\left(\hat{\theta}_n\right)_i\right)\left(\theta-\hat{\theta}_n\right)$. Hence,

$$
\begin{aligned}
&\left\|b(\theta) - b_{\text{Laplace}}(\theta)\right\|^2 \\
&= \sum_{i=1}^{P}\left(R\left(\nabla \ln p\left(\theta|\mathbf{y}_n\right)_i, \theta\right) + \nabla\pi\left(\hat{\theta}_n\right)_i + \nabla\left(\nabla\pi\left(\hat{\theta}_n\right)_i\right)\left(\theta-\hat{\theta}_n\right)\right)^2 \\
&\le \sum_{i=1}^{P}\left(R\left(\nabla \ln p\left(\theta|\mathbf{y}_n\right)_i, \theta\right) + M_0\right)^2 \\
&= \sum_{i=1}^{P} R\left(\nabla \ln p\left(\theta|\mathbf{y}_n\right)_i, \theta\right)^2 + 2M_0\sum_{i=1}^{P} R\left(\nabla \ln p\left(\theta|\mathbf{y}_n\right)_i, \theta\right) + M_0^2 \\
&\le \sup_{t\in[0,1]}\sum_{i=1}^{P}\left\|\theta-\hat{\theta}_n\right\|^4\left\|\nabla^2\left(\nabla \ln p\left(\hat{\theta}_n + t\left(\theta-\hat{\theta}_n\right)|\mathbf{y}_n\right)_i\right)\right\|^2 \\
&+ M_0\sup_{t\in[0,1]}\sum_{i=1}^{P}\left\|\theta-\hat{\theta}_n\right\|^2\left\|\nabla^2\left(\nabla \ln p\left(\hat{\theta}_n + t\left(\theta-\hat{\theta}_n\right)|\mathbf{y}_n\right)_i\right)\right\| \\
&+ M_0^2 \\
&\le n^2M^2\left\|\theta-\hat{\theta}_n\right\|^4 + 2nM_0M\left\|\theta-\hat{\theta}_n\right\|^2 + M_0^2
\end{aligned}
$$

by Assumptions 5 and 6. Following Huggins et al. (2018), the condition

$$
\left\{\int\left\|\theta-\hat{\theta}_n\right\|^{2p} p\left(\theta|\mathbf{y}_n\right)d\theta\right\}^{1/p} \le \frac{L_p}{n} \quad (p=1,2)
$$

is satisfied under Assumptions 1–10. Thus, we have

$$
\begin{aligned}
\mathcal{W}_2\left(\hat{\Pi}_{\text{Laplace}}, p\left(\theta|\mathbf{y}_n\right)\right) &\le \frac{\alpha(\mathbf{y}^n)\left(\int\left\|b(\theta)-b_{\text{Laplace}}(\theta)\right\|^2 p\left(\theta|\mathbf{y}_n\right)d\theta\right)^{1/2}}{n} \\
&\le \frac{\alpha(\mathbf{y}^n)\left(M^2L_2^2 + 2M_0ML_1 + M_0^2\right)^{1/2}}{n}
\end{aligned}
$$

by Theorem 4.1, Theorem 5.2 and Proposition 6.2 in Huggins et al. (2018). ■

A.2. *Expressions for* $B_{t,1}, B_{t,21}^1, B_{t,21}^2, B_{t,22}, B_4$

For $B_{t,1}$, we have

$$
\begin{aligned}
B_{t,1} = &-\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\hat{\theta}_n\right)\right] - \nabla l_t\left(\hat{\theta}_n\right)'\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\frac{\nabla\hat{p}}{\hat{p}} \\
&+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right).
\end{aligned}
\tag{73}
$$

For $B_{t,21}^1$, we have

$$
\begin{aligned}
&B_{t,21}^1 \\
&= -\frac{1}{8}\left(\nabla l_t\left(\hat{\theta}_n\right)\right)'\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\left(\bar{\mathbf{H}}_n^{(5)}\left(\hat{\theta}_n\right)\right)'vec\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right] \\
&+\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(4)}\left(\hat{\theta}_n\right)'
\end{aligned}
\tag{74}
$$

$$
\begin{aligned}
&\times\left(vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\otimes\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\right)\right)\\
&+\frac{1}{6}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\right)'\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right]\bar{\mathbf{H}}_n^{(4)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\\
&+\frac{1}{16}vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\\
&\times\mathbf{tr}\left[\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right)'\bar{\mathbf{H}}_n^{(4)}\left(\hat{\theta}_n\right)\right]\\
&+\frac{1}{4}\mathbf{tr}\left[\begin{pmatrix}\left(\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\left(\nabla l_t\left(\hat{\theta}_n\right)'\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right)\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\\ \otimes vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\end{pmatrix}'\bar{\mathbf{H}}_n^{(4)}\left(\hat{\theta}_n\right)\right]\\
&-\frac{3}{8}vec\left(\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)'\\
&\times vec\begin{bmatrix}\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)'\\ \times\left(vec\left(\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)^{-1}\right)\otimes\left(\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\right)\right)\end{bmatrix}\\
&-\frac{1}{4}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\right)'\begin{bmatrix}\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\\ \begin{pmatrix}\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)'\\ \times\left(vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\otimes\left(\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\right)\right)\end{pmatrix}\end{bmatrix}\\
&\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\right)\\
&-\frac{1}{16}vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)'vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\\
&\times vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\\
&-\frac{1}{24}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\right)'\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\right)\\
&\times vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\\
&-\frac{5}{8}\left(\nabla l_t\left(\hat{\theta}_n\right)\right)'\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\frac{\nabla\hat{p}}{\hat{p}}\mathbf{tr}\left[\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right]\left(\bar{\mathbf{H}}_n^{(4)}\left(\hat{\theta}_n\right)\right)'\right]\\
&-\frac{1}{8}\mathbf{tr}\left[\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right]\left(\bar{\mathbf{H}}_n^{(4)}\left(\hat{\theta}_n\right)\right)'\right]\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\frac{\nabla\hat{p}'}{\hat{p}}\right]\\
&-\frac{1}{2}\mathbf{tr}\left[\left[\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\nabla l_t\left(\hat{\theta}_n\right)\frac{\nabla\hat{p}'}{\hat{p}}\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\otimes vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right]\left(\bar{\mathbf{H}}_n^{(4)}\left(\hat{\theta}_n\right)\right)'\right]\\
&+\frac{1}{8}vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\left(\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\right)'vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\\
&\times\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\frac{\nabla\hat{p}}{\hat{p}}\nabla l_t\left(\hat{\theta}_n\right)\right]\\
&+\frac{1}{12}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\right)'\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\right)\\
&\times\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\frac{\nabla\hat{p}}{\hat{p}}\nabla l_t\left(\hat{\theta}_n\right)\right]\\
&+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)'\\
&\times vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\frac{\nabla\hat{p}}{\hat{p}}\nabla l_t\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right),
\end{aligned}
$$

For $B_{t,21}^2$, we have

$$B_{t,21}^2=\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\frac{\nabla\hat{p}}{\hat{p}}\nabla l_t\left(\hat{\theta}_n\right) \tag{75}$$

$$\times\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)$$
$$+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\right)'\left[\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}\nabla l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right]$$
$$\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\right)$$
$$-\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla l_t\left(\widehat{\theta}_n\right)$$
$$-\frac{1}{4}\mathbf{tr}\left[\frac{\nabla^2\widehat{p}}{\widehat{p}}\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla l_t\left(\widehat{\theta}_n\right)$$
$$-\frac{1}{2}vec\left(\left(\frac{\nabla^2\widehat{p}}{\widehat{p}}\right)^{-1}\otimes\nabla l_t\left(\widehat{\theta}_n\right)\right)'\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\right)$$
$$+\frac{1}{2}\left(\nabla l_t\left(\widehat{\theta}_n\right)\right)'\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{(\nabla^3\widehat{p})'}{\widehat{p}}\left[vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)\right],$$

For $B_{t,22}$, we have

$$B_{t,22}=-\frac{1}{16}\mathbf{tr}\left[A_2\right]\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\right] \tag{76}$$
$$-\frac{1}{4}\mathbf{tr}\left[\left[\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)\right]\bar{\mathbf{H}}_n^{(4)}\left(\widehat{\theta}_n\right)'\right]$$
$$+\frac{1}{16}A_1\times\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\right]+\frac{1}{24}A_3\times\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\right]$$
$$+\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)'$$
$$\times vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)$$
$$+\frac{1}{8}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}$$
$$\times\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)$$
$$+\frac{1}{4}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\right)'\begin{bmatrix}\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)\\ \otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\end{bmatrix}$$
$$\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\right)$$
$$-\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^3 l_t\left(\widehat{\theta}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)$$
$$-\frac{1}{6}vec\left(\nabla^2 l_t\left(\widehat{\theta}_n\right)\right)\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right]$$
$$\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\right)$$
$$+\frac{1}{8}\mathbf{tr}\left[\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)\right]\nabla^4 l_t\left(\widehat{\theta}_n\right)'\right]$$
$$-\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}$$
$$-\frac{1}{4}\mathbf{tr}\left[\nabla^2 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}$$
$$-\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}$$
$$+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\right)'\nabla^3 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}$$
$$+\frac{1}{4}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\right]\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\right]$$
$$+\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\theta}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\right]$$

For B_4, we have

$$B_4 = -\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\frac{\nabla^2\hat{p}}{\hat{p}}\right] + \frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\frac{\nabla\hat{p}}{\hat{p}} - \frac{1}{8}A_1 - \frac{1}{12}A_3 + \frac{1}{8}\mathbf{tr}\left[A_2\right], \quad (77)$$

where

$$A_1 = vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)'vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)$$
$$= \mathbf{tr}\left[vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)'A_4\right],$$

$$A_2 = \left[\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\right)\right]'\bar{\mathbf{H}}_n^{(4)}\left(\hat{\theta}_n\right),$$

$$A_4 = \bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)\bar{\mathbf{H}}_n\left(\hat{\theta}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\hat{\theta}_n\right)'.$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2025.105978.

References

Aigner, D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. J. Econometrics 6, 21–37.
Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory. Vol. 1, Springer Verlag, pp. 267–281.
Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. J. Amer. Statist. Assoc. 88 (422), 669–679.
Andrews, D.W.K., 1987. Consistency in nonlinear econometric models: A generic uniform law of large numbers. Econometrica 55 (6), 1465–1471.
Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, second ed. Springer-Verlag.
Brooks, S., 2002. Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde 2002. J. R. Stat. Soc. Ser. B 64, 616–618.
Burnham, K., Anderson, D., 2002. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. Springer.
Chen, C., 1985. On asymptotic normality of limiting density functions with Bayesian implications. J. R. Stat. Soc. Ser. B 47, 540–546.
Claeskens, G., Hjort, N., 2008. Model Selection and Model Averaging. Cambridge University Press, Cambridge.
Gallant, A.R., White, H., 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Blackwell.
Greene, W.H., 1990. A gamma-distributed stochastic frontier model. J. Econometrics 46, 141–163.
Griffin, J.E., Steel, M.F.J., 2007. Bayesian stochastic frontier analysis using winbugs. J. Prod. Anal. 27, 163–176.
Held, L., Holmes, C.C., 2006. Bayesian auxiliary variable models for binary and multinomial regression. Bayesian Anal. 1 (1), 145–168.
Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and EI George, and a rejoinder by the authors). Statist. Sci. 14 (4), 382–417.
Huggins, J.H., Campbell, T., Kasprzak, M., Broderick, T., 2018. Practical bounds on the error of Bayesian posterior approximations: A nonasymptotic approach. arXiv preprint arXiv:1809.09505.
Hurn, S., Martin, V., Phillips, P.C.B., Yu, J., 2020. Financial Econometric Modeling. Oxford University Press.
Inoue, A., Kilian, L., 2006. On the selection of forecasting models. J. Econometrics 130 (2), 273–306.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Kass, R., Tierney, L., Kadane, J., 1990. The validity of posterior expansions based on Laplace's method. In: Geisser, S., Hodges, J.S., Press, S.J., Zellner, A. (Eds.), Bayesian and Likelihood Methods in Statistics and Econometrics. Elsevier Science Publishers B.V., North-Holland.
Koop, G., Steel, M.F.J., Osiewalski, J., 1995. Posterior analysis of stochastic frontier models using gibbs sampling. Comput. Statist. 10, 353–373.
Kumbhakar, S.C., Lovell, C.A.K., 2000. Stochastic Frontier Analysis, First Cambridge University Press.
Kumbhakar, S.C., Tsionas, E.G., 2005. Measuring technical and allocative inefficiency in the translog cost system: a Bayesian approach. J. Econometrics 126, 355–384.
Kurkalova, L.A., Carriquiry, A., 2002. An analysis of grain production decline during the early transition in Ukraine: a Bayesian inference. Am. J. Agric. Econ. 84, 1256–1263.
Li, Y., Wu, Z., Yu, J., Zeng, T., 2024. A Note on AIC and TIC for Model Selection. Working Paper.
Li, Y., Yu, J., Zeng, T., 2020. Deviance information criterion for latent variable models and misspecified models. J. Econometrics 216 (2), 450–493.
Lindley, D.V., Smith, A.F., 1972. Bayes estimates for the linear model. J. R. Stat. Soc. Ser. B 34 (1), 1–18.
Liu, J.S., Wu, Y.N., 1999. Parameter expansion for data augmentation. J. Amer. Statist. Assoc. 94 (448), 1264–1274.
McAlinn, K., West, M., 2019. Dynamic Bayesian predictive synthesis in time series forecasting. J. Econometrics 210 (1), 155–169.
Polson, N.G., Scott, J.G., Windle, J., 2013. Bayesian inference for logistic models using Pólya–Gamma latent variables. J. Amer. Statist. Assoc. 108 (504), 1339–1349.
Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B 64, 583–639.
Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2014. The deviance information criterion: 12 years on. J. R. Stat. Soc. Ser. B 76, 485–493.
Stone, M., 1977. An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1), 44–47.
Tallman, E., West, M., 2024. Bayesian predictive decision synthesis. J. R. Stat. Soc. Ser. B Stat. Methodol. 86 (2), 340–363.
Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. J. Amer. Statist. Assoc. 82 (398), 528–540.
Tsionas, E.G., 2002. Stochastic frontier models with random coefficients. J. Appl. Econometrics 17, 127–147.
Tsionas, M.G., Mallick, S.K., 2019. A Bayesian semiparametric approach to stochastic frontiers and productivity. European J. Oper. Res. 274 (1), 391–402.
Vehtari, A., Lampinen, J., 2002. Bayesian model assessment and comparison using cross-validation predictive densities. Neural Comput. 14, 2439–2468.
Vehtari, A., Ojanen, J., 2012. A survey of Bayesian predictive methods for model assessment, selection and comparison. Stat. Surv. 6, 142–228.

White, H., 1996. Estimation, Inference and Specification Analysis. Cambridge University Press., Cambridge, UK..

Wooldridge, J.M., 1994. Estimation and inference for dependent processes. Handb. Econ. 4, 2639–2738.

Zens, G., Frühwirth-Schnatter, S., Wagner, H., 2022a. Ultimate Pólya Gamma Samplers–Efficient MCMC for possibly imbalanced binary and categorical data. arXiv preprint, arXiv:2011.06898.

Zens, G., Frühwirth-Schnatter, S., Wagner, H., 2022b. Efficient Bayesian modeling of binary and categorical data in r: the UPG package. arXiv preprint, arXiv:2101.02506.

Zhang, Y., Yang, Y., 2015. Cross-validation for selecting a model selection procedure. J. Econometrics 187 (1), 95–112.