Deviance Information Criterion for Model Selection: Theoretical Justification and Applications*

Yong Li

Renmin University of China

Nianling Wang

Capital University of Economics and Business

Jun Yu

Singapore Management University

Tao Zeng

Zhejiang University

October 12, 2022

Abstract

This paper gives a rigorous justification to the Deviance information criterion (DIC), which has been extensively used for model selection based on MCMC output. It is shown that, when a plug-in predictive distribution is used and under a set of regularity conditions, DIC is an asymptotically unbiased estimator of

*We wish to thank Eric Renault, Peter Phillips and David Spiegelhalter for their helpful comments.
Yong Li, School of Economics, Renmin University of China, Beijing, 100872, P.R. China. Nianling Wang, School of Finance, Capital University of Economics and Business, Beijing, China. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, Singapore 178903. Tao Zeng, School of Economics and Academy of Financial Research, Zhejiang University, Zhejiang, China 310027.

the expected Kullback-Leibler divergence between the data generating process and the plug-in predictive distribution. High-order expansions to DIC and the effective number of parameters are developed, facilitating investigating the effect of the prior. DIC is used to compare alternative discrete-choice models, alternative GARCH-type models, and alternative copula models in three empirical applications.

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

JEL: C11, C52, C25, C22, C32

1 Introduction

A highly important statistical inference 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 AIC. 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. Like AIC, it trades off a measure of model adequacy against a measure of complexity and is concerned with how hypothetically replicate data predict the observed data. 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

[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 September 2022, it has received 13524 citations on Google Scholar.

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 the heuristic justification given by Spiegelhalter et al. (2002), the frequentist framework and the Bayesian framework were mixed. 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 GARCH-type models and alternative copula models. Section 5 concludes the paper. In Appendix, 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.

Throughout the paper, we use $:=$, $\mathbf{tr}$, vec, $\otimes$, $o(1)$, $o_p(1)$, $O_p(1)$, $\stackrel{p}{\to}$ 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 $\tilde{\boldsymbol{\theta}}_n$, $\overline{\boldsymbol{\theta}}_n$, $\widehat{\boldsymbol{\theta}}_n$, $\boldsymbol{\theta}_n^p$ to denote a generic estimator, the posterior mean, the quasi maximum likelihood (QML) estimator, and the pseudo true parameter of $\boldsymbol{\theta}$, respectively.

2 Decision-theoretic Justification of DIC

There are essentially two strands of literature on model selection.[2] 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 hypothetically replicate 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

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

widely used in applications.

2.1 Predictive model selection as a decision problem

Assuming that the probabilistic behavior of observed data, $\mathbf{y} = (y_1, y_2, \cdots, y_n)' \in \mathbf{Y}$, is described by a set of probabilistic models such as $\{M_k\}_{k=1}^K = \{p(\mathbf{y}|\boldsymbol{\theta}_k, M_k)\}_{k=1}^K$ where n is the sample size, $\boldsymbol{\theta}_k$ (without confusion, we simply write it as $\boldsymbol{\theta}$) 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 $\mathbf{y}_{rep} = (y_{1,rep}, \cdots, y_{n,rep})'$ 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

$p\left(\mathbf{y}_{rep}|\tilde{\boldsymbol{\theta}}_n(\mathbf{y}), M_k\right)$ for M_k where $\tilde{\boldsymbol{\theta}}_n(\mathbf{y})$ is an estimate of $\boldsymbol{\theta}$ based on $\mathbf{y}$ (when there is no confusion we simply write $\tilde{\boldsymbol{\theta}}_n(\mathbf{y})$ as $\tilde{\boldsymbol{\theta}}_n$).

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 $p\left(\mathbf{y}_{rep}|\tilde{\boldsymbol{\theta}}_n(\mathbf{y}), M_k\right)$ multiplied by 2,

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

Naturally, the loss function associated with decision d_k is

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

As a result, the model selection problem is,

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

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,

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

The smaller $Risk(d_k)$ is, the better the candidate model performs when using $p\left(\mathbf{y}_{rep}|\tilde{\boldsymbol{\theta}}_n(\mathbf{y}), M_k\right)$ 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 $p(\mathbf{y}|\boldsymbol{\theta}, M_k)$ as $p(\mathbf{y}|\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq R^P$ (i.e. the dimension of $\boldsymbol{\theta}$ is P). When the candidate model is different, the value of P may be different. Define AIC by

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

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

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

which is the global maximum interior to $\boldsymbol{\Theta}$.

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

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

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

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

and takes the form of

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

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

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

The better the model fits the data, the larger the log-likelihood value, and hence, the smaller the value for $\overline{D(\boldsymbol{\theta})}$. 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:

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

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

DIC can be rewritten in two equivalent forms:

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

and

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

DIC defined in Equation (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 Equation (1) the Bayesian measure, $\overline{D(\boldsymbol{\theta})}$, is the same as $D\left(\overline{\boldsymbol{\theta}}_n\right) + P_D$ 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.

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 $\boldsymbol{\theta}_n^p$ 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 $\mathbf{y}_{rep}|\boldsymbol{\theta}_n^p$. 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 $\boldsymbol{\theta}_n^p$, instead of replacing it with a frequentist estimator and then finding the asymptotic property of the "optimism", in Sections 7.1 and 7.3 of Spiegelhalter et al. (2002), $\boldsymbol{\theta}_n^p$ is replaced with a random quantity $\boldsymbol{\theta}$ and then calculates 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 Equation (33) on Page 602, the loss function defined in the first paragraph on Page 603, and Equation (40) on Page 603, one may deduce that the following KL divergence

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

was used.[3] Hence,

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

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

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

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

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

The corresponding expected loss function of a statistical decision d_k is

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

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

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

Denote the selected model by M_{k^*}. Then $p\left(\mathbf{y}_{rep}|\overline{\boldsymbol{\theta}}_n(\mathbf{y}),M_{k^*}\right)$ is used to generate future observations where $\overline{\boldsymbol{\theta}}_n(\mathbf{y})$ is the posterior mean of $\boldsymbol{\theta}$ 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 $l_t\left(\mathbf{y}^t,\boldsymbol{\theta}\right)=\ln p\left(\mathbf{y}^t|\boldsymbol{\theta}\right)-\ln p\left(\mathbf{y}^{t-1}|\boldsymbol{\theta}\right)$ be the conditional log-likelihood for the t^{th} observation for any $1\le t\le n$. When there is no confusion, we suppress $l_t\left(\mathbf{y}^t,\boldsymbol{\theta}\right)$ as $l_t\left(\boldsymbol{\theta}\right)$ so that the log-likelihood function $\ln p\left(\mathbf{y}|\boldsymbol{\theta}\right)$ is $\sum_{t=1}^n l_t\left(\boldsymbol{\theta}\right)$.[4] Let $\nabla^j l_t\left(\boldsymbol{\theta}\right)$ denote the j^{th} derivative of $l_t\left(\boldsymbol{\theta}\right)$ and $\nabla^j l_t\left(\boldsymbol{\theta}\right)=l_t\left(\boldsymbol{\theta}\right)$ when $j=0$. Furthermore, define

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

[4]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.

$$\mathbf{H}_n(\boldsymbol{\theta}) = \int \bar{\mathbf{H}}_n(\boldsymbol{\theta})\, g(\mathbf{y})\, d\mathbf{y},\ \mathbf{J}_n(\boldsymbol{\theta}) = \int \bar{\mathbf{J}}_n(\boldsymbol{\theta})\, g(\mathbf{y})\, d\mathbf{y}.$$

In this paper, we impose the following regularity conditions.

Assumption 1: $\boldsymbol{\Theta} \subset R^P$ is compact.

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

Assumption 3: For all t, $l_t(\boldsymbol{\theta})$ satisfies the standard measurability and continuity condition, and the eight-times differentiability condition on $\boldsymbol{\Theta}$ almost surely**.**

Assumption 4: For $j = 0, 1, 2$, for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$, $\|\triangledown^j l_t(\boldsymbol{\theta}) - \triangledown^j l_t(\boldsymbol{\theta}')\| \le c_t^j(\mathbf{y}^t)\, \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|$, 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 $j = 0, 1, \ldots, 8$, there exist $M_t(\mathbf{y}^t)$ and $M < \infty$ such that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\triangledown^j l_t(\boldsymbol{\theta})$ exists, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\triangledown^j l_t(\boldsymbol{\theta})\| \leqslant M_t(\mathbf{y}^t)$, $\sup_t E\|M_t(\mathbf{y}^t)\|^{r+\delta} \le M$ for some $\delta > 0$, where r is the same as that in Assumption 2.

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

Assumption 7: Let $\boldsymbol{\theta}_n^p$ be the pseudo-true value that minimizes the KL loss between the DGP and the candidate model

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

where $\{\boldsymbol{\theta}_n^p\}$ is the sequence of minimizers that are interior to $\boldsymbol{\Theta}$ uniformly in n. For all $\varepsilon > 0$,

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

where $N(\boldsymbol{\theta}_n^p, \varepsilon)$ is the open ball of radius ε around $\boldsymbol{\theta}_n^p$.

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

Assumption 9: $\mathbf{H}_n(\boldsymbol{\theta}_n^p) + \mathbf{B}_n(\boldsymbol{\theta}_n^p) = o(1)$.

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

Remark 2.1 *Assumption 1 is the compactness condition. Assumption 2 and Assumption 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 2.2 *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 Equation (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*

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

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 Equation (8) holds. Furthermore, if Assumptions 1-7 hold, for any* $\varepsilon > 0$*, there exists* $K_2(\varepsilon) > 0$ *such that*

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

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

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

Remark 2.3 *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* $\mathbf{H}_n(\boldsymbol{\theta}_n^p) + \mathbf{B}_n(\boldsymbol{\theta}_n^p)$ *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

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

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

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

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

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

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

From the iid assumption, we have

$$Var(s_t(\boldsymbol{\theta}_n^p)) = E\left(s_t(\boldsymbol{\theta}_n^p)\, s_t(\boldsymbol{\theta}_n^p)'\right)$$

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

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*

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

Hence,

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

since $\gamma_0 = \delta_0/n^{1/2}$*. Thus,* $\mathbf{H}_n\left(\boldsymbol{\theta}_n^p\right) + \mathbf{B}_n\left(\boldsymbol{\theta}_n^p\right) = o\left(1\right)$*. However,* $\mathbf{H}_n\left(\boldsymbol{\theta}_n^p\right) + \mathbf{B}_n\left(\boldsymbol{\theta}_n^p\right) \neq 0$ *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).*

To develop the Laplace approximation, we need to fix more notations. For the convenience of exposition, we let $\bar{\mathbf{H}}_n^{(j)}\left(\boldsymbol{\theta}\right) = \frac{1}{n}\sum_{t=1}^n \nabla^j l_t\left(\boldsymbol{\theta}\right)$ for $j = 3, 4, 5$. Let $\pi\left(\boldsymbol{\theta}\right) = \ln p\left(\boldsymbol{\theta}\right)$, $\widehat{p}$, $\widehat{\pi}$, $\nabla^j\widehat{p}$, and $\nabla^j\widehat{\pi}$ be the values of functions, $p\left(\boldsymbol{\theta}\right)$, $\pi\left(\boldsymbol{\theta}\right)$, $\nabla^j p\left(\boldsymbol{\theta}\right)$, and $\nabla^j\pi\left(\boldsymbol{\theta}\right)$ evaluated at $\widehat{\boldsymbol{\theta}}_n$. 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* $n \to \infty$*,*

$$\frac{\int l_t\left(\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}\right)p\left(\mathbf{y}|\boldsymbol{\theta}\right)d\boldsymbol{\theta}}{\int p\left(\boldsymbol{\theta}\right)p\left(\mathbf{y}|\boldsymbol{\theta}\right)d\boldsymbol{\theta}} \tag{11}$$

$$= \quad l_t\left(\widehat{\boldsymbol{\theta}}_n\right)+\frac{1}{n}B_{t,1}+\frac{1}{n^2}\left(B_{t,21}^1+B_{t,21}^2+B_{t,22}-B_4B_{t,1}\right)+O_p\left(n^{-3}\right),$$

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 5.2.*

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

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

with

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

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

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

Remark 2.4 *DIC is an unbiased estimator of $E_{\mathbf{y}} E_{\mathbf{y}_{rep}} \left[-2 \ln p \left(\mathbf{y}_{rep} | \overline{\boldsymbol{\theta}}_n(\mathbf{y}) \right) \right]$ 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 $p \left(\mathbf{y}_{rep} | \overline{\boldsymbol{\theta}}_n(\mathbf{y}) \right)$. A key difference between AIC and DIC is that the plug-in predictive density is based on different estimators of $\boldsymbol{\theta}$. In AIC, the QML estimate, $\widehat{\boldsymbol{\theta}}_n(\mathbf{y})$, is used. In DIC, the posterior mean, $\overline{\boldsymbol{\theta}}_n(\mathbf{y})$, is used. In this sense, DIC is the Bayesian version of AIC.*

Remark 2.5 *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 2.6 *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 2.7 *If $p \left(\mathbf{y} | \boldsymbol{\theta} \right)$ 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.

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

$$\mathbf{y} \sim N\left(F_1\boldsymbol{\theta}_1, G_1\right), \tag{12}$$

and the conjugate prior for $\boldsymbol{\theta}_1$ is

$$\boldsymbol{\theta}_1 \sim N\left(F_2\boldsymbol{\phi}, G_2\right), \tag{13}$$

where F_1 is $n \times P$ matrix, $\boldsymbol{\theta}_1$ 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, $\boldsymbol{\theta} = \boldsymbol{\theta}_1$. The log likelihood function is

$$L\left(\mathbf{y}|\boldsymbol{\theta}\right) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\left|G_1\right| - \frac{1}{2}\left(\mathbf{y} - F_1\boldsymbol{\theta}_1\right)' G_1^{-1}\left(\mathbf{y} - F_1\boldsymbol{\theta}_1\right).$$

It is easy to see that the QML estimate of $\boldsymbol{\theta}$ is

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

The log prior density is

$$\pi(\boldsymbol{\theta}) = -\frac{P}{2}\ln 2\pi - \frac{1}{2}\ln|G_2| - \frac{1}{2}(\boldsymbol{\theta}_1 - F_2\boldsymbol{\phi})' G_2^{-1}(\boldsymbol{\theta}_1 - F_2\boldsymbol{\phi}).$$

It is well-known that the posterior distribution of $\boldsymbol{\theta}$ is

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

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

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

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

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

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

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

DIC is

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

Hence, by Lemma 2.3, we have

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

Spiegelhalter et al. (2002) express P_D as

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

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

3.2 The normal linear models with unknown sampling precision

Suppose the model is

$$\mathbf{y} \sim N\left(F_1\boldsymbol{\theta}_1, \tau^{-1}G_1\right). \tag{21}$$

Assume the conjugate prior for $\boldsymbol{\theta}_1$ is

$$\boldsymbol{\theta}_1 \sim N\left(F_2\boldsymbol{\phi}, \tau^{-1}G_2\right), \tag{22}$$

and the conjugate prior for τ is

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

Assume G_1, F_2, $\boldsymbol{\phi}$, and G_2 are all known. Let $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \tau\right)'$, where the dimension of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_1$ 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 $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta}_1 | \tau, \mathbf{y} \sim N\left(V_1 b_1, V_1\right) \text{ and } \tau | \mathbf{y} \sim \Gamma\left(a + \frac{n}{2}, b + \frac{S}{2}\right),$$

where

$$V_1^{-1} = \tau V^{-1}, b_1 = \tau b, S = \left(\mathbf{y} - F_1 F_2 \boldsymbol{\phi}\right)' \left(G_1 + F_1' G_2 F_1\right)^{-1} \left(\mathbf{y} - F_1 F_2 \boldsymbol{\phi}\right).$$

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

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

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

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

where $\bar{\mathbf{H}}_{n,11}\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1} = -\left(\hat{\tau}F_1'G_1^{-1}F_1\right)^{-1}$ is the submatrix of $\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}$ corresponding to $\boldsymbol{\theta}_1$. 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

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

with

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

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

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

The first term of (27) is

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

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

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

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 GARCH-type models using the USD/euro exchange rate returns. In the third application we compares four copula models using S&P indexes returns. All three classes of models have been widely applied in economics. 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.

4.1 Discrete choice models

In this section, we compare a binary probit model and a binary logit model. Let $\mathbf{y} = (y_{1,}y_{2,} \dots y_n)'$ 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

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

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

The latent variable representation of (32) is

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

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, Holmes and Held (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.[5] In total, there are eight parameters in both models, including the intercept.

We specify a vague prior distribution for parameters as

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

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

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 $E_{\boldsymbol{\theta}|\mathbf{y}}[\ln p(\mathbf{y}|\boldsymbol{\theta}))]$ where $\boldsymbol{\theta} = \boldsymbol{\beta}$, which does not have a closed-form expression. We approximate it based on the MCMC output as,

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

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

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

Table 1 reports $D\left(\overline{\boldsymbol{\theta}}\right)$, 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 $D\left(\overline{\boldsymbol{\theta}}\right)$ 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 $C_2/n = n^{-1}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2\widehat{\pi}\right]$ 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.

4.2 GARCH-type models

Pioneered by Engle (1982), the GARCH-type models have proven highly useful in modeling time-varying conditional variance. Two GARCH-type models are particularly popular, namely, the GARCH model of Bollerslev (1986) and the EGARCH model of Nelson (1991). These two models are non-nested. A way of comparing them is to investigate their out-of-sample forecasting performance, as did in Hansen and Lunde (2005). In this paper, we use DIC to compare the following GARCH(1,1) and EGARCH(1,1):

$$\text{GARCH(1,1)}: \begin{cases} y_t = \sigma_t \varepsilon_t, \varepsilon_t \overset{iid}{\sim} N(0,1) \\ \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{cases},$$

$$\text{EGARCH(1,1)}: \begin{cases} y_t = \exp\left(\frac{h_t}{2}\right) e_t, e_t \overset{iid}{\sim} N(0,1) \\ h_t = \delta_0 + \delta_1 h_{t-1} + \gamma_1 \left|e_{t-1}\right| \end{cases}.$$

In this application, we assume y_t is the demeaned daily log return of the USD/euro exchange rate over the period between January 2, 2020 and September 16, 2022. The data is downloaded from the European Central Bank (ECB) website. The sample size $n = 698$.

To do posterior sampling and DIC computation, we impose vague priors under parameter restriction as follows. For the prior distribution in GARCH(1,1):

$$\begin{cases} \alpha_0, \alpha_1, \beta_1 \overset{iid}{\sim} N(0,100), \text{ if } \alpha_0 > 0 \text{ and } \alpha_1 \geq 0 \text{ and } \beta_1 \geq 0 \text{ and } \alpha_1 + \beta_1 < 1, \\ \text{negative infinity prior density, if } \alpha_0 \leq 0 \text{ or } \alpha_1 < 0 \text{ or } \beta_1 < 0 \text{ or } \alpha_1 + \beta_1 \geq 1. \end{cases}$$

For the prior distribution in EGARCH(1,1):

$$\begin{cases} \delta_0, \delta_1, \gamma_1 \overset{iid}{\sim} N(0,100), \text{ if } -1 < \delta_1 < 1, \\ \text{negative infinity prior density, if } \delta_1 \geq 1 \text{ or } \delta_1 \leq -1. \end{cases}$$

The MCMC output are obtained using "MitISEM" package in R (Basturk et al., 2017). The total number of iterations is 100,000. The burn-in period is the first 50,000 iterations. 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. 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 $D\left(\overline{\boldsymbol{\theta}}\right)$, P_D, DIC, and C_2/n for both models. According to DIC, GARCH(1,1) model has a smaller DIC value than EGARCH(1,1) (842.26 versus 849.58). This indicates that DIC prefers GARCH(1,1) over EGARCH(1,1). To see the effect of the prior, we report the value of C_2/n. With vague priors, the values of C_2/n are negligible in both models.

Table 2: Parameter estimation results for GARCH(1,1) and EGARCH(1,1)

	GARCH(1,1)			EGARCH(1,1)		
	α_0	α_1	β_1	δ_0	δ_1	γ_1
Posterior Mean	0.0047	0.0662	0.9157	-0.2130	0.9288	0.2861
Posterior SD	0.0025	0.0185	0.0246	0.0549	0.0210	0.0694

Table 3: Model selection results for GARCH(1,1) and EGARCH(1,1)

Model	$D\left(\overline{\boldsymbol{\theta}}\right)$	P_D	DIC	C_2/n
GARCH(1,1)	837.36	2.45	842.26	7.89×10^{-6}
EGARCH(1,1)	844.15	2.71	849.58	6.91×10^{-5}

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

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

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, $F(z_{it}; v)$, $f(z_{it}; v)$ be the cumulative distribution function (CDF) and probability density function (PDF) of the t distribution with v degrees of freedom ($v > 2$) 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 ($\eta > 2$), $q_{\phi,it} = \Phi^{-1}(F(z_{it}; v))$, $q_{f,it} = F^{-1}(F(z_{it}; v); \eta)$. 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).

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

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

where M is the number of effective draws.

To compare these four Copula models, we calculate $D\left(\overline{\boldsymbol{\theta}}\right)$, 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

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

Table 5: Model selection results for four copula models

Model	$D\left(\overline{\boldsymbol{\theta}}\right)$	P_D	DIC	C_2/n
gnc	31378	5.20	31389	-0.0006
gtc	29689	5.69	29700	-0.0014
ttc	29305	5.60	29316	-0.0016
ctc	30490	5.67	30502	-0.0016

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., $C_2/n = n^{-1}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2\widehat{\pi}\right]$. The prior effects are small.[6]

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 $p\left(\mathbf{y}_{rep}|\overline{\boldsymbol{\theta}}_n(\mathbf{y})\right)$. This loss function is slightly different from that in AIC by using the posterior mean $\overline{\boldsymbol{\theta}}_n(\mathbf{y})$ as the estimator of $\boldsymbol{\theta}$ 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

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

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 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. Moreover, when the data are nonstationary, the asymptotic normality may not hold. In this case, it remains unknown whether or not DIC is justified.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, Springer Verlag, **1**, 267-281.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, **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.

Basturk, N., Grassi, S., Hoogerheide, L., Opschoor, A. and Van Dijk, H. K. (2017). The R package MitISEM: Efficient and robust simulation procedures for Bayesian inference. *Journal of Statistical Software*, **79(1)**, 1-39.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd edition. Springer-Verlag.

Brooks, S. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde (2002). *Journal of the Royal Statistical Society Series B*, **64**, 616–618.

Burnham, K. and 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. *Journal of the Royal Statistical Society Series B*, **47**, 540–546.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50(4)**, 987–1007.

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

Hansen, P.R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, **20**, 873-889.

Held, L. and Holmes, C. C. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian analysis*, **1(1)**, 145-168.

Hurn, S., Martin,V., Phillips, P. C. B. and Yu, J. (2020). *Financial Econometric Modeling*. Oxford University Press.

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

Kass, R., Tierney, L. and Kadane, J. (1990). The validity of posterior expansions based on Laplace's Method. in *Bayesian and Likelihood Methods in Statistics and Econometrics*, ed. by S. Geisser, J.S. Hodges, S.J. Press and A. Zellner. Elsevier Science Publishers B.V.: North-Holland.

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

Lindley, D. V. and Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34(1)**, 1-18.

Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, **94(448)**, 1264-1274.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, **59(2)**, 347-370.

Polson, N. G., Scott, J. G. and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American statistical Association*, **108(504)**, 1339-1349.

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

Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society Series B*, **76**, 485-493.

Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, **82(398)**, 528-540.

Vehtari, A. and Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation*, **14**, 2439-2468.

Vehtari, A. and Ojanen, J. (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, **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. *Handbook of Econometrics*, **4**, 2639-2738.

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

Zhang, Y. and Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, **187(1)**, 95-112.

Appendix

5.1 A Proof of Theorem 2.1

We write $\mathbf{H}_n(\boldsymbol{\theta}_n^p)$ as $\mathbf{H}_n$, $\mathbf{B}_n(\boldsymbol{\theta}_n^p)$ 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

$$\overline{\boldsymbol{\theta}}_n(\mathbf{y}) = \widehat{\boldsymbol{\theta}}_n(\mathbf{y}) + O_p\left(n^{-1}\right) \tag{34}$$

by (86). Then, we have

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

$$\frac{1}{\sqrt{n}}\mathbf{B}_n^{-1/2}\frac{\partial \ln p(\mathbf{y}_{rep}|\boldsymbol{\theta}_n^p)}{\partial \boldsymbol{\theta}} \stackrel{d}{\to} N\left(0, \mathbf{I}_P\right), \tag{35}$$

and

$$\mathbf{C}_n^{-1/2}\sqrt{n}\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y}) - \boldsymbol{\theta}_n^p\right) \stackrel{d}{\to} N\left(0, \mathbf{I}_P\right). \tag{36}$$

Note that

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

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

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

from (34). Then, we have

$$\begin{aligned}
T_2 &= E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[-2\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta}_n^p) + 2\ln p\left(\mathbf{y}_{rep}|\overline{\boldsymbol{\theta}}_n(\mathbf{y}_{rep})\right)\right] \\
&= E_{\mathbf{y}}E_{\mathbf{y}_{rep}}\left[-\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y}_{rep}) - \boldsymbol{\theta}_n^p\right)' \frac{\partial \ln p\left(\mathbf{y}_{rep}|\overline{\boldsymbol{\theta}}_n(\mathbf{y}_{rep})\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y}_{rep}) - \boldsymbol{\theta}_n^p\right) + o_p(1)\right] \\
&= E_{\mathbf{y}_{rep}}\left[-\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y}_{rep}) - \boldsymbol{\theta}_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}_{rep}|\overline{\boldsymbol{\theta}}_n(\mathbf{y}_{rep})\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y}_{rep}) - \boldsymbol{\theta}_n^p\right)\right] + o(1) \\
&= E_{\mathbf{y}}\left[-\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y}) - \boldsymbol{\theta}_n^p\right)' \frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\boldsymbol{\theta}}_n(\mathbf{y})\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y}) - \boldsymbol{\theta}_n^p\right)\right] + o(1),
\end{aligned}$$

by Assumptions 5-6 and the dominated convergence theorem (Chung, 2001 and Das-Gupta, 2008). Next, we expand $\ln p\left(\mathbf{y}_{rep}|\overline{\boldsymbol{\theta}}_n(\mathbf{y})\right)$ at $\boldsymbol{\theta}_n^p$:

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

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

since

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

by (35), (36), and by the dominated convergence theorem.

Note that

$$\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\overline{\boldsymbol{\theta}}_n(\mathbf{y})\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = E_{\mathbf{y}}\left(\frac{1}{n}\frac{\partial^2 \ln p\left(\mathbf{y}|\boldsymbol{\theta}_n^p\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) + o_p(1),$$

by Assumptions 1-10 and the uniform law of large numbers. Hence, we get

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

Hence, we only need to analyze T_3. Note that

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

and

$$E_{\mathbf{y}}\left[\mathbf{C}_n^{-1/2}\sqrt{n}\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y})-\boldsymbol{\theta}_n^p\right)\sqrt{n}\left(\overline{\boldsymbol{\theta}}_n(\mathbf{y})-\boldsymbol{\theta}_n^p\right)'\mathbf{C}_n^{-1/2}\right]=\mathbf{I}_P+o(1).$$

Hence,

$$\begin{aligned}
T_3 &= \mathbf{tr}\left((-\mathbf{H}_n)\mathbf{C}_n^{1/2}\mathbf{C}_n^{1/2}\right)+o(1)=\mathbf{tr}\left((-\mathbf{H}_n)\mathbf{C}_n\right)+o(1)\\
&= \mathbf{tr}\left((-\mathbf{H}_n)(-\mathbf{H}_n)^{-1}\mathbf{B}_n(-\mathbf{H}_n)^{-1}\right)+o(1)\\
&= \mathbf{tr}\left((-\mathbf{H}_n)(-\mathbf{H}_n)^{-1}\mathbf{B}_n(-\mathbf{H}_n)^{-1}\right)+o(1)\\
&= \mathbf{tr}\left(\mathbf{B}_n(-\mathbf{H}_n)^{-1}\right)+o(1),
\end{aligned}$$

and

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

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

The last step is due to the assumption that $\mathbf{H}_n\left(\boldsymbol{\theta}_n^p\right) + \mathbf{B}_n\left(\boldsymbol{\theta}_n^p\right) = o\left(1\right)$.

Following Lemma 2.3, we get $P_D = P + o_p(1)$. Finally, by Assumption 10 and the dominated convergence theorem, we have

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

5.2 Expressions for $B_{t,1}, B_{t,21}^1, B_{t,21}^2, B_{t,22}, B_4$

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

$$\begin{aligned} B_{t,1} \quad = \quad & -\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right] - \nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)'\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}} \\ & +\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right). \end{aligned} \tag{37}$$

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

$$\begin{aligned} & B_{t,21}^1 \\ = \quad & -\frac{1}{8}\left(\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\left(\bar{\mathbf{H}}_n^{(5)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'vec\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\right] \\ & +\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n^{(4)}\left(\widehat{\boldsymbol{\theta}}_n\right)' \\ & \times\left(vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\otimes\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right)\right) \\ & +\frac{1}{6}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right]\bar{\mathbf{H}}_n^{(4)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right) \\ & +\frac{1}{16}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right) \\ & \times\mathbf{tr}\left[\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\right)'\bar{\mathbf{H}}_n^{(4)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right] \end{aligned} \tag{38}$$

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

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

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

$$B_{t,21}^2=\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right) \tag{39}$$
$$\times\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)'vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)$$
$$+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'\left[\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right]$$
$$\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)$$
$$-\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)$$
$$-\frac{1}{4}\mathbf{tr}\left[\frac{\nabla^2\widehat{p}}{\widehat{p}}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)$$
$$-\frac{1}{2}vec\left(\left(\frac{\nabla^2\widehat{p}}{\widehat{p}}\right)^{-1}\otimes\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)$$
$$+\frac{1}{2}\left(\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\left(\nabla^3\widehat{p}\right)'}{\widehat{p}}\left[vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\right],$$

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

$$B_{t,22}=-\frac{1}{16}\mathbf{tr}\left[A_2\right]\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right] \tag{40}$$
$$-\frac{1}{4}\mathbf{tr}\left[\left[\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\right]\bar{\mathbf{H}}_n^{(4)}\left(\widehat{\boldsymbol{\theta}}_n\right)'\right]$$
$$+\frac{1}{16}A_1\times\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right]+\frac{1}{24}A_3\times\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right]$$
$$+\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)'$$
$$\times vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)$$

$$
\begin{aligned}
&+\frac{1}{8}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1} \\
&\times\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)'vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right) \\
&+\frac{1}{4}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'\left[\begin{array}{c}\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right) \\ \otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\end{array}\right] \\
&\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right) \\
&-\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^3 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)'vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right) \\
&-\frac{1}{6}vec\left(\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right)\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right] \\
&\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right) \\
&+\frac{1}{8}\mathbf{tr}\left[\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\right]\nabla^4 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)'\right] \\
&-\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}} \\
&-\frac{1}{4}\mathbf{tr}\left[\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}} \\
&-\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}} \\
&+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\nabla^3 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}} \\
&+\frac{1}{4}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right]\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\right] \\
&+\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\right]
\end{aligned}
$$

For B_4, we have

$$
\begin{aligned}
B_4 &= -\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\right]+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}} \\
&\quad -\frac{1}{8}A_1-\frac{1}{12}A_3+\frac{1}{8}\mathbf{tr}\left[A_2\right], \qquad (41)
\end{aligned}
$$

where

$$\begin{aligned} A_1 &= vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)' \bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right) \bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1} \bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right) \\ &= \mathbf{tr}\left[vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)' A_4\right], \end{aligned}$$

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

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

Online Supplement (Not for Publication)

5.3 Proof of Lemma 2.1

Note that

$$\begin{aligned} & \frac{1}{n}\sum_{t=1}^{n}\left[l_t\left(\boldsymbol{\theta}\right)-l_t\left(\boldsymbol{\theta}_n^p\right)\right] \\ = & \frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)+\frac{1}{n}\sum_{t=1}^{n}\left(E\left[l_t\left(\boldsymbol{\theta}\right)\right]-E\left[l_t\left(\boldsymbol{\theta}_n^p\right)\right]\right)+\frac{1}{n}\sum_{t=1}^{n}\left(E\left[l_t\left(\boldsymbol{\theta}_n^p\right)\right]-l_t\left(\boldsymbol{\theta}_n^p\right)\right). \end{aligned}$$

From (7), we know that for any $\varepsilon > 0$, there exists $\delta_1(\varepsilon) > 0$ and $N(\varepsilon) > 0$, for all $n > N(\varepsilon)$,

$$\frac{1}{n}\sum_{t=1}^{n}\left\{E\left[l_t\left(\boldsymbol{\theta}\right)\right]-E\left[l_t\left(\boldsymbol{\theta}_n^p\right)\right]\right\} < -\delta_1(\varepsilon),$$

if $\boldsymbol{\theta} \in \boldsymbol{\Theta}\backslash N\left(\boldsymbol{\theta}_n^p, \varepsilon\right)$. Thus, for any $\varepsilon > 0$, if $\boldsymbol{\theta} \in \boldsymbol{\Theta}\backslash N\left(\boldsymbol{\theta}_n^p, \varepsilon\right)$, for all $n > N(\varepsilon)$,

$$\frac{1}{n}\sum_{t=1}^{n}\left[l_t\left(\boldsymbol{\theta}\right)-l_t\left(\boldsymbol{\theta}_n^p\right)\right] < \frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)-\delta_1\left(\varepsilon\right)+\frac{1}{n}\sum_{t=1}^{n}\left(E\left[l_t\left(\boldsymbol{\theta}_n^p\right)\right]-l_t\left(\boldsymbol{\theta}_n^p\right)\right),$$

and

$$\begin{aligned} & \sup_{\boldsymbol{\Theta}\backslash N\left(\boldsymbol{\theta}_n^p,\varepsilon\right)}\frac{1}{n}\sum_{t=1}^{n}\left[l_t\left(\boldsymbol{\theta}\right)-l_t\left(\boldsymbol{\theta}_n^p\right)\right] \\ \leq & \sup_{\boldsymbol{\Theta}\backslash N\left(\boldsymbol{\theta}_n^p,\varepsilon\right)}\frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)-\delta_1\left(\varepsilon\right)+\frac{1}{n}\sum_{t=1}^{n}\left(E\left[l_t\left(\boldsymbol{\theta}_n^p\right)\right]-l_t\left(\boldsymbol{\theta}_n^p\right)\right) \\ \leq & \sup_{\boldsymbol{\Theta}\backslash N\left(\boldsymbol{\theta}_n^p,\varepsilon\right)}\left|\frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)\right|-\delta_1\left(\varepsilon\right)+\left|\frac{1}{n}\sum_{t=1}^{n}\left(E\left[l_t\left(\boldsymbol{\theta}_n^p\right)\right]-l_t\left(\boldsymbol{\theta}_n^p\right)\right)\right| \\ \leq & 2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)\right|-\delta_1\left(\varepsilon\right). \end{aligned} \tag{42}$$

Under Assumptions 1-6, the uniform convergence condition is satisfied, that is,

$$P\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)\right|<\varepsilon\right)\to 1, \tag{43}$$

From the uniform convergence, if we choose δ_2 such that $0<\delta_2<\delta_1\left(\varepsilon\right)/2$, we have

$$P\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)\right|<\delta_2\right]\to 1.$$

Hence,

$$P\left[2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)\right|-\delta_1\left(\varepsilon\right)<2\delta_2-\delta_1\left(\varepsilon\right)\right]\to 1.$$

From (42), we have

$$\begin{aligned}
& P\left[2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{t=1}^{n}\left(l_t\left(\boldsymbol{\theta}\right)-E\left[l_t\left(\boldsymbol{\theta}\right)\right]\right)\right|-\delta_1\left(\varepsilon\right)<2\delta_2-\delta_1\left(\varepsilon\right)\right] \\
\leqslant\ & P\left[\sup_{\boldsymbol{\Theta}\backslash N\left(\boldsymbol{\theta}_n^p,\varepsilon\right)}\frac{1}{n}\left[\sum_{t=1}^{n}l_t\left(\boldsymbol{\theta}\right)-\sum_{t=1}^{n}l_t\left(\boldsymbol{\theta}_n^p\right)\right]<2\delta_2-\delta_1\left(\varepsilon\right)\right].
\end{aligned}$$

Letting $K_1\left(\varepsilon\right)=-\left(2\delta_2-\delta_1\left(\varepsilon\right)\right)>0$, we have, for any $\varepsilon>0$,

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

which proves the consistency condition given by (8). The proof of the other two concentration conditions (9) and (10) can be done similarly, and hence, omitted.

5.4 Proof of Lemma 2.2

Before we prove Lemma 2.2, we need to prove the three lemmas. Lemma 5.1 is about the high-order analytical expansions while Lemma 5.3 and Lemma 5.2 are about the high-order stochastic expansions.

5.4.1 High-order analytical expansions

Suppose $\mathbf{\Theta}$ is a compact subset of R^P. For any $\boldsymbol{\theta} \in \mathbf{\Theta}$, let $\{h_n(\boldsymbol{\theta}) : n = 1, 2, \dots\}$ be a sequence of eight-times continuously differentiable functions of $\boldsymbol{\theta}$, having an interior global minimum at $\left\{\widehat{\boldsymbol{\theta}}_n : n = 1, 2, \dots\right\}$, $b(\boldsymbol{\theta})$ be a six-times continuously differentiable real function of $\boldsymbol{\theta}$. For any function $f(\boldsymbol{\theta})$, let $\widehat{f}$ be the value of function f evaluated at $\widehat{\boldsymbol{\theta}}_n$ (i.e., $\widehat{f} = f\left(\widehat{\boldsymbol{\theta}}_n\right)$). When there is no confusion, we write $h_n(\boldsymbol{\theta})$ as $h(\boldsymbol{\theta})$ or h_n or even h and $b(\boldsymbol{\theta})$ as b. We use $B_\delta(\boldsymbol{\theta})$ to denote the open ball of radius δ centered at $\boldsymbol{\theta}$. So $B_{\sqrt{n}\delta}(0)$ is an open ball of radius $\sqrt{n}\delta$ centered at the origin. For convenience of exposition, we write $\frac{\partial^d}{\partial \boldsymbol{\theta}_{j_1} \partial \boldsymbol{\theta}_{j_2} \cdots \partial \boldsymbol{\theta}_{j_d}} f(\boldsymbol{\theta})$ as $f_{j_1 \cdots j_d}$. The Hessian of h_n at $\boldsymbol{\theta}$ is denoted by $\nabla^2 h_n(\boldsymbol{\theta})$, and its (i, j)-component is written as h_{ij} while the component of its inverse is written as h^{ij}. Let μ^4_{ijkq}, μ^6_{ijkqrs}, $\mu^8_{ijkqrstw}$, $\mu^{10}_{ijkqrstwv\beta}$, $\mu^{12}_{ijkqrstwv\beta\tau\phi}$ be the fourth, sixth, eighth, tenth, and twelfth central moments of a multivariate normal distribution whose covariance matrix is $\left(\nabla^2 \widehat{h}\right)^{-1} = \left(\nabla^2 h_n(\boldsymbol{\theta})\right)^{-1} |_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_n}$. Note that we require $h_n(\boldsymbol{\theta})$ be eight-times continuously differentiable and $b(\boldsymbol{\theta})$ be six-times continuously differentiable. These two conditions are stronger than what have typically been assumed in the literature on the Laplace approximation as we would like to develop higher order expansions.

Following Kass et al. (1990), we call the pair $(\{h_n\}, b)$ satisfies the analytical assumptions for Laplace's method if the following assumptions are met. There exists positive numbers ε, M and η, and an integer n_0 such that $n \geq n_0$ implies (i) for all $\boldsymbol{\theta} \in B_\varepsilon\left(\widehat{\boldsymbol{\theta}}_n\right)$ and all $1 \leq j_1, \cdots, j_d \leq P$ with $0 \leq d \leq 8$, $\|h_n(\boldsymbol{\theta})\| < M$ and $\|h_{j_1 \cdots j_d}(\boldsymbol{\theta})\| < M$; (ii) $\nabla^2 \widehat{h}$ is positive definite and $\det\left(\nabla^2 \widehat{h}\right) > \eta$; (iii) $\int_{\mathbf{\Theta}} b(\boldsymbol{\theta}) \exp\left[-nh(\boldsymbol{\theta})\right] d\boldsymbol{\theta}$ exists and is finite, and for all δ for which $0 < \delta < \varepsilon$ and $B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right) \subseteq \mathbf{\Theta}$,

$$\left[\det\left(n\nabla^2 \widehat{h}\right)\right]^{\frac{1}{2}} \int_{\mathbf{\Theta} - B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta}) \exp\left[-nh_n(\boldsymbol{\theta}) - n\widehat{h}\right] d\boldsymbol{\theta} = O\left(n^{-3}\right).$$

If one sets $-nh_n$ to be the sequence of the log-likelihood functions of a model (as a sequence of n), we say the model is Laplace regular. Lemma 5.1 below and Lemma 5.2 in the next subsection extend Theorem 1 and Theorem 5 of Kass et al. (1990) to a higher order.

Lemma 5.1 *If $(\{h_n\}, b)$ satisfy the analytical assumptions for Laplace's method, then*

$$\int_{\boldsymbol{\Theta}} b(\boldsymbol{\theta}) \exp[-nh(\boldsymbol{\theta})]\, d\boldsymbol{\theta} = (2\pi)^{\frac{P}{2}} \left[\det\left(n\nabla^2 \widehat{h}\right)\right]^{-\frac{1}{2}} \exp\left(-n\widehat{h}\right) \left(\widehat{b} + \frac{1}{n}Q_1 + \frac{1}{n^2}Q_2 + O\left(n^{-3}\right)\right),$$

where

$$Q_1 = -\frac{1}{24}\sum_{ijkq} \widehat{h}_{ijkq}\mu^4_{ijkq}\widehat{b} + \frac{1}{72}\sum_{ijkqrs} \widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs}\widehat{b} - \frac{1}{6}\sum_{ijk\zeta} \widehat{h}_{ijk}\mu^4_{ijk\zeta}\widehat{b}_\zeta + \frac{1}{2}\sum_{\zeta\eta} \widehat{b}_{\zeta\eta}\widehat{h}^{\zeta\eta},$$

$$\begin{aligned}
Q_2 \quad = \quad & -\frac{1}{720}\sum_{ijkqrs} \widehat{h}_{ijkqrs}\mu^6_{ijkqrs}\widehat{b} + \frac{1}{1152}\sum_{ijkqrstw} \widehat{h}_{ijkq}\widehat{h}_{rstw}\mu^8_{ijkqrstw}\widehat{b} \\
& + \frac{1}{720}\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrstw}\mu^8_{ijkqrstw}\widehat{b} - \frac{1}{1728}\sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv\beta}\mu^{10}_{ijkqrstwv\beta}\widehat{b} \\
& + \frac{1}{31104}\sum_{ijkqrstwv\beta\tau\phi} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}\mu^{12}_{ijkqrstwv\beta\tau\phi}\widehat{b} - \frac{1}{120}\sum_{ijkqr\zeta} \widehat{h}_{ijkqr}\mu^6_{ijkqr\zeta}\widehat{b}_\zeta \\
& + \frac{1}{144}\sum_{ijkqrst\zeta} \widehat{h}_{ijk}\widehat{h}_{qrst}\mu^8_{ijkqrst\zeta}\widehat{b}_\zeta - \frac{1}{1296}\sum_{ijkqrstwv\zeta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\mu^{10}_{ijkqrstwv\zeta}\widehat{b}_\zeta \\
& - \frac{1}{48}\sum_{ijkq\zeta\eta} \widehat{h}_{ijkq}\mu^6_{ijkq\zeta\eta}\widehat{b}_{\zeta\eta} + \frac{1}{144}\sum_{ijkqrs\zeta\eta} \widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrs\zeta\eta}\widehat{b}_{\zeta\eta} \\
& - \frac{1}{36}\sum_{ijk\zeta\eta\xi} \widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{\zeta\eta\xi} + \frac{1}{24}\sum_{\zeta\eta\xi\omega} \widehat{b}_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}.
\end{aligned}$$

Proof. Note that, by the third analytical assumption for Laplace's method, we have

$$\int_{\boldsymbol{\Theta}} b(\boldsymbol{\theta}) \exp[-nh(\boldsymbol{\theta})]\, d\boldsymbol{\theta} = \int_{B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta}) \exp[-nh(\boldsymbol{\theta})]\, d\boldsymbol{\theta} + \int_{\boldsymbol{\Theta} - B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta}) \exp[-nh(\boldsymbol{\theta})]\, d\boldsymbol{\theta},$$

and

$$\begin{aligned}
&\int_{\boldsymbol{\Theta}-B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta})\exp\left(-nh(\boldsymbol{\theta})\right)d\boldsymbol{\theta} \\
&= \left[\det\left(n\nabla^2\widehat{h}\right)\right]^{-\frac{1}{2}}\exp\left(-n\widehat{h}\right)\left[\det\left(n\nabla^2\widehat{h}\right)\right]^{\frac{1}{2}}\int_{\boldsymbol{\Theta}-B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta})\exp\left(-n\left(h(\boldsymbol{\theta})-\widehat{h}\right)\right)d\boldsymbol{\theta} \\
&= \left[\det\left(n\nabla^2\widehat{h}\right)\right]^{-\frac{1}{2}}\exp\left(-n\widehat{h}\right)O\left(n^{-3}\right).
\end{aligned}$$

Let $u=\sqrt{n}\left(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_n\right)$. Applying the Taylor expansion to $h(\boldsymbol{\theta})$ at $\widehat{\boldsymbol{\theta}}_n$, we have

$$\begin{aligned}
nh(\boldsymbol{\theta}) &= n\widehat{h}+\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j+\frac{1}{6}n^{-\frac{1}{2}}\sum_{ijk}\widehat{h}_{ijk}u_iu_ju_k+\frac{1}{24}n^{-1}\sum_{ijkq}\widehat{h}_{ijkq}u_iu_ju_ku_q \\
&+\frac{1}{120}n^{-\frac{3}{2}}\sum_{ijkqr}\widehat{h}_{ijkqr}u_iu_ju_ku_qu_r+\frac{1}{720}n^{-2}\sum_{ijkqrs}\widehat{h}_{ijkqrs}u_iu_ju_ku_qu_ru_s \\
&+\frac{1}{5040}n^{-\frac{5}{2}}\sum_{ijkqrst}\widehat{h}_{ijkqrst}u_iu_ju_ku_qu_ru_su_t+r_n(u),
\end{aligned}$$

where

$$r_n(u)=\frac{1}{40320}n^{-3}\sum_{ijkqrstw}h_{ijkqrstw}\left(\boldsymbol{\theta}'\right)u_iu_ju_ku_qu_ru_su_tu_w,$$

and $\boldsymbol{\theta}'$ lies between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}_n$.

Define

$$\begin{aligned}
x &= \frac{1}{6}n^{-\frac{1}{2}}\sum_{ijk}\widehat{h}_{ijk}u_iu_ju_k+\frac{1}{24}n^{-1}\sum_{ijkq}\widehat{h}_{ijkq}u_iu_ju_ku_q+\frac{1}{120}n^{-\frac{3}{2}}\sum_{ijkqr}\widehat{h}_{ijkqr}u_iu_ju_ku_qu_r \\
&+\frac{1}{720}n^{-2}\sum_{ijkqrs}\widehat{h}_{ijkqrs}u_iu_ju_ku_qu_ru_s+\frac{1}{5040}n^{-\frac{5}{2}}\sum_{ijkqrst}\widehat{h}_{ijkqrst}u_iu_ju_ku_qu_ru_su_t+r_n(u).
\end{aligned}$$

Applying the Taylor expansion to $\exp(-x)$ at the origin, we have

$$\exp(-nh) = \exp\left\{-n\widehat{h}\right\}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right)\times$$

$$\left(1+\Xi_1+\frac{1}{2}\Xi_2+\frac{1}{6}\Xi_3+\frac{1}{24}\Xi_4+\frac{1}{120}\Xi_5+R_{1,n}\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)\right),$$

where

$$\begin{aligned}
\Xi_1 &= -\frac{1}{6}n^{-\frac{1}{2}}\sum_{ijk}\widehat{h}_{ijk}u_iu_ju_k-\frac{1}{24}n^{-1}\sum_{ijkq}\widehat{h}_{ijkq}u_iu_ju_ku_q-\frac{1}{120}n^{-\frac{3}{2}}\sum_{ijkqr}\widehat{h}_{ijkqr}u_iu_ju_ku_qu_r\\
&\quad -\frac{1}{720}n^{-2}\sum_{ijkqrs}\widehat{h}_{ijkqrs}u_iu_ju_ku_qu_ru_s-\frac{1}{5040}n^{-\frac{5}{2}}\sum_{ijkqrst}\widehat{h}_{ijkqrst}u_iu_ju_ku_qu_ru_su_t,\\
\Xi_2 &= \frac{1}{36}n^{-1}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}u_iu_ju_ku_qu_ru_s+\frac{1}{24^2}n^{-2}\sum_{ijkqrstw}\widehat{h}_{ijkq}\widehat{h}_{rstw}u_iu_ju_ku_qu_ru_su_tu_w\\
&\quad +\frac{1}{72}n^{-\frac{3}{2}}\sum_{ijkqrst}\widehat{h}_{ijk}\widehat{h}_{qrst}u_iu_ju_ku_qu_ru_su_t+\frac{1}{360}n^{-2}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrstw}u_iu_ju_ku_qu_ru_su_tu_w\\
&\quad +\frac{1}{1440}n^{-\frac{5}{2}}\sum_{ijkqrstwv}\widehat{h}_{ijk}\widehat{h}_{qrstwv}u_iu_ju_ku_qu_ru_su_tu_wu_v\\
&\quad +\frac{1}{2160}n^{-\frac{5}{2}}\sum_{ijkqrstwv}\widehat{h}_{ijk}\widehat{h}_{qrstwv}u_iu_ju_ku_qu_ru_su_tu_wu_v,\\
\Xi_3 &= -\frac{1}{216}n^{-\frac{3}{2}}\sum_{ijkqrstwv}h_{ijk}h_{qrs}h_{twv}u_iu_ju_ku_qu_ru_su_tu_wu_v\\
&\quad -\frac{1}{288}n^{-2}\sum_{ijkqrstwv\beta}h_{ijk}h_{qrs}h_{twv\beta}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta\\
&\quad -\frac{1}{1152}n^{-\frac{5}{2}}\sum_{ijkqrstwv\beta\tau}h_{ijk}h_{qrst}h_{wv\beta\tau}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau\\
&\quad -\frac{1}{1440}n^{-\frac{5}{2}}\sum_{ijkqrstwv\beta\tau}h_{ijk}h_{qrs}h_{twv\beta\tau}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau,\\
\Xi_4 &= \frac{1}{1296}n^{-2}\sum_{ijkqrstwv\beta\tau\phi}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau u_\phi\\
&\quad +\frac{1}{1296}n^{-\frac{5}{2}}\sum_{ijkqrstwv\beta\tau\phi\alpha}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi\alpha}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau u_\phi u_\alpha,\\
\Xi_5 &= -\frac{1}{7776}n^{-\frac{5}{2}}\sum_{ijkqrstwv\beta\tau\phi\alpha\varkappa\varrho}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}\widehat{h}_{\alpha\varkappa\varrho}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau u_\phi u_\alpha u_\varkappa u_\varrho.
\end{aligned}$$

Applying the Taylor expansion to $b(\boldsymbol{\theta})$ at $\widehat{\boldsymbol{\theta}}_n$, we have

$$b(\boldsymbol{\theta}) = \widehat{b}+n^{-\frac{1}{2}}\sum_i\widehat{b}_iu_i+\frac{1}{2}n^{-1}\sum_{ij}\widehat{b}_{ij}u_iu_j+\frac{1}{6}n^{-\frac{3}{2}}\sum_{ijk}\widehat{b}_{ijk}u_iu_ju_k$$

$$+\frac{1}{24}n^{-2}\sum_{ijkq}\widehat{b}_{ijkq}u_iu_ju_ku_q+\frac{1}{120}n^{-\frac{5}{2}}\sum_{ijkqr}\widehat{b}_{ijkqr}u_iu_ju_ku_qu_r+R_{2,n}\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right).$$

Hence,

$$\begin{aligned}
b(\boldsymbol{\theta})\exp\left(-nh(\boldsymbol{\theta})\right) &= \exp\left(-n\hat{h}\right)\exp\left[-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right]\left\{I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)\right\} \\
&= (2\pi)^{\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{-\frac{1}{2}}\exp\left(-n\hat{h}\right)\left(\frac{1}{2\pi}\right)^{\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right) \\
&\quad\times\left\{I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)\right\},
\end{aligned}$$

where $I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)$ and $R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)$ will be specified below. Thus,

$$\begin{aligned}
&\int_{B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta})\exp\left[-nh(\boldsymbol{\theta})\right]d\boldsymbol{\theta} \\
=\ & (2\pi)^{\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{-\frac{1}{2}}\exp\left(-n\widehat{h}\right)\times \\
&\int_{B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)}(2\pi)^{-\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right)\left\{I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)\right\}d\boldsymbol{\theta} \\
=\ & (2\pi)^{\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{-\frac{1}{2}}\exp\left(-n\widehat{h}\right)n^{-\frac{P}{2}}\times \\
&\int_{B_{\sqrt{n}\delta}(0)}(2\pi)^{-\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right)\left\{I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)\right\}du \\
=\ & (2\pi)^{\frac{P}{2}}\left|n\nabla^2\widehat{h}\right|^{-\frac{1}{2}}\exp\left(-n\widehat{h}\right)\times \\
&\int_{B_{\sqrt{n}\delta}(0)}(2\pi)^{-\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right)\left\{I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)\right\}du,
\end{aligned}$$

where $R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)$ is the sum of the terms involving $R_{1,n}\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)$, $R_{2,n}\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)$, and the terms whose order is equal to or smaller than $O(n^{-3})$. Furthermore, we can get

$$R_{2,n}\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)=\frac{1}{720}n^{-3}\sum b_{ijkqrs}\left(\widetilde{\boldsymbol{\theta}}\right)u_iu_ju_ku_qu_ru_s,$$

where $\widetilde{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}_n$. Thus, the leading term of $R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)$ is $\widehat{b}R_{1n}\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+$

$R_{2n}\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)$ that include $r_n(u)$. The integral of $r_n(u)$ over $B_\varepsilon\left(\widehat{\boldsymbol{\theta}}_n\right)$ can be expressed as

$$\begin{aligned}
&\left|n^{-3}\int_{B_{\sqrt{n}\delta}(0)}(2\pi)^{-\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right)\sum_{ijkqrstw}h_{ijkqrstw}\left(\boldsymbol{\theta}'\right)u_iu_ju_ku_qu_ru_su_tu_wdu\right| \\
\leq\ & n^{-3}\int_{B_{\sqrt{n}\delta}(0)}(2\pi)^{-\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right)\sum_{ijkqrstw}\left|h_{ijkqrstw}\left(\boldsymbol{\theta}'\right)\right|\left|u_iu_ju_ku_qu_ru_su_tu_w\right|du \\
\leq\ & n^{-3}M\int_{R^P}(2\pi)^{-\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right)\sum_{ijkqrstw}\left|u_iu_ju_ku_qu_ru_su_tu_w\right|du \\
=\ & O\left(n^{-3}\right),
\end{aligned}$$

where $\int_{R^P}(2\pi)^{-\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right)\sum_{ijkqrstw}\left|u_iu_ju_ku_qu_ru_su_tu_w\right|du$ is the eighth order moment of folded multivariate folded normal distribution with mean 0 and covariance $\left(\nabla^2\widehat{h}\right)^{-1}$ that is finite; see Kamat (1953) and Kan and Robotti (2017). Then, we have

$$\begin{aligned}
& (2\pi)^{\frac{P}{2}}\left|n\nabla^2\widehat{h}\right|^{-\frac{1}{2}}\exp\left(-n\widehat{h}\right)\int_{B_{\sqrt{n}\delta}(0)}\left[I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}\right)+R_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}\right)\right]f(u)\,du \\
=\ & (2\pi)^{\frac{P}{2}}\left|n\nabla^2\widehat{h}\right|^{-\frac{1}{2}}\exp\left(-n\widehat{h}\right)\left[\int_{B_{\sqrt{n}\delta}(0)}I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}\right)f(u)\,du+O\left(n^{-3}\right)\right].
\end{aligned}$$

For $I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)$, we have

$$I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)=I_n^0\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+I_n^1\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+I_n^2\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+I_n^3\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+I_n^4\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right)+I_n^5\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right),$$

where

$$\begin{aligned}
I_n^0\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) &= \widehat{b}\left(1+\Xi_1+\frac{1}{2}\Xi_2+\frac{1}{6}\Xi_3+\frac{1}{24}\Xi_4+\frac{1}{120}\Xi_5\right) \\
&= \widehat{b}\left\{1+n^{-\frac{1}{2}}I_n^{01}+n^{-1}I_n^{02}+n^{-\frac{3}{2}}I_n^{03}+n^{-2}I_n^{04}+n^{-\frac{5}{2}}I_n^{05}\right\}, \\
I_n^1\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) &= n^{-\frac{1}{2}}\sum\widehat{b}_\zeta u_\zeta\left(1+\Xi_1+\frac{1}{2}\Xi_2+\frac{1}{6}\Xi_3+\frac{1}{24}\Xi_4+\frac{1}{120}\Xi_5\right) \\
&= n^{-\frac{1}{2}}\sum\widehat{b}_\zeta u_\zeta+n^{-1}I_n^{11}+n^{-\frac{3}{2}}I_n^{12}+n^{-2}I_n^{13}+n^{-\frac{5}{2}}I_n^{14},
\end{aligned}$$

$$
\begin{aligned}
I_n^2\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) &= n^{-1}\frac{1}{2}\sum\widehat{b}_{\zeta\eta}u_\zeta u_\eta\left(1+\Xi_1+\frac{1}{2}\Xi_2+\frac{1}{6}\Xi_3+\frac{1}{24}\Xi_4+\frac{1}{120}\Xi_5\right)\\
&= \frac{1}{2}\left[n^{-1}\sum\widehat{b}_{\zeta\eta}u_\zeta u_\eta+n^{-\frac{3}{2}}I_n^{21}+n^{-2}I_n^{22}+n^{-\frac{5}{2}}I_n^{23}\right],\\
I_n^3\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) &= n^{-\frac{3}{2}}\frac{1}{6}\sum\widehat{b}_{\zeta\eta\xi}u_\zeta u_\eta u_\xi\left(1+\Xi_1+\frac{1}{2}\Xi_2+\frac{1}{6}\Xi_3+\frac{1}{24}\Xi_4+\frac{1}{120}\Xi_5\right)\\
&= \frac{1}{6}\left[n^{-\frac{3}{2}}\sum\widehat{b}_{\zeta\eta\xi}u_\zeta u_\eta u_\xi+n^{-2}I_n^{31}+n^{-\frac{5}{2}}I_n^{32}\right],\\
I_n^4\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) &= n^{-2}\frac{1}{24}\sum\widehat{b}_{\zeta\eta\xi\omega}u_\zeta u_\eta u_\xi u_\omega\left(1+\Xi_1+\frac{1}{2}\Xi_2+\frac{1}{6}\Xi_3+\frac{1}{24}\Xi_4+\frac{1}{120}\Xi_5\right)\\
&= \frac{1}{24}\left[n^{-2}\sum\widehat{b}_{\zeta\eta\xi\omega}u_\zeta u_\eta u_\xi u_\omega+n^{-\frac{5}{2}}I_n^{41}\right],\\
I_n^5\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) &= n^{-\frac{5}{2}}\frac{1}{120}\sum_{ijkqr}\widehat{b}_{ijkqr}u_iu_ju_ku_qu_r\left(1+\Xi_1+\frac{1}{2}\Xi_2+\frac{1}{6}\Xi_3+\frac{1}{24}\Xi_4+\frac{1}{120}\Xi_5\right)\\
&= \frac{1}{120}\left[n^{-\frac{5}{2}}\sum_{ijkqr}\widehat{b}_{ijkqr}u_iu_ju_ku_qu_r\right].
\end{aligned}
$$

with

$$
\begin{aligned}
I_n^{01} &= -\frac{1}{6}\sum_{ijk}\widehat{h}_{ijk}u_iu_ju_k,\\
I_n^{02} &= -\frac{1}{24}\sum_{ijkq}\widehat{h}_{ijkq}u_iu_ju_ku_q+\frac{1}{72}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}u_iu_ju_ku_qu_ru_s,\\
I_n^{03} &= -\frac{1}{120}\sum_{ijkqr}\widehat{h}_{ijkqr}u_iu_ju_ku_qu_r+\frac{1}{144}\sum_{ijkqrst}\widehat{h}_{ijk}\widehat{h}_{qrst}u_iu_ju_ku_qu_ru_su_t\\
&\quad -\frac{1}{1296}\sum_{ijkqrstwv}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}u_iu_ju_ku_qu_ru_su_tu_wu_v,\\
I_n^{04} &= -\frac{1}{720}\sum_{ijkqrs}\widehat{h}_{ijkqrs}u_iu_ju_ku_qu_ru_s+\frac{1}{1152}\sum_{ijkqrstw}\widehat{h}_{ijkq}\widehat{h}_{rstw}u_iu_ju_ku_qu_ru_su_tu_w\\
&\quad +\frac{1}{720}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrstw}u_iu_ju_ku_qu_ru_su_tu_w\\
&\quad -\frac{1}{1728}\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv\beta}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta\\
&\quad +\frac{1}{31104}\sum_{ijkqrstwv\beta\tau\phi}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau u_\phi,\\
I_n^{05} &= -\frac{1}{5040}\sum_{ijkqrst}\widehat{h}_{ijkqrst}u_iu_ju_ku_qu_ru_su_t+\frac{1}{2880}\sum_{ijkqrstwv}\widehat{h}_{ijk}\widehat{h}_{qrstwv}u_iu_ju_ku_qu_ru_su_tu_wu_v
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{5120}n^{-\frac{5}{2}}\sum_{ijkqrstwv}\widehat{h}_{ijk}\widehat{h}_{qrstwv}u_iu_ju_ku_qu_ru_su_tu_wu_v\\
&-\frac{1}{6912}\sum_{ijkqrstwv\beta\tau}h_{ijk}h_{qrst}h_{wv\beta\tau}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau\\
&-\frac{1}{8640}\sum_{ijkqrstwv\beta\tau}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv\beta\alpha}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau\\
&+\frac{1}{31104}\sum_{ijkqrstwv\beta\tau\phi\alpha}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi\alpha}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau u_\phi u_\alpha\\
&-\frac{1}{933120}\sum_{ijkqrstwv\beta\tau\phi\alpha\varkappa\varrho}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}\widehat{h}_{\alpha\varkappa\varrho}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau u_\phi u_\alpha u_\varkappa u_\varrho,
\end{aligned}
$$

$$
\begin{aligned}
I_n^{11} &= -\frac{1}{6}\sum_{ijk\zeta}\widehat{h}_{ijk}u_iu_ju_ku_\zeta\widehat{b}_\zeta,\\
I_n^{12} &= -\frac{1}{24}\sum_{ijkq\zeta}\widehat{h}_{ijkq}u_iu_ju_ku_qu_\zeta\widehat{b}_\zeta+\frac{1}{72}\sum_{ijkqrs\zeta}\widehat{h}_{ijk}\widehat{h}_{qrs}u_iu_ju_ku_qu_ru_su_\zeta\widehat{b}_\zeta,\\
I_n^{13} &= -\frac{1}{120}\sum_{ijkqr}\widehat{h}_{ijkqr}u_iu_ju_ku_qu_ru_\zeta\widehat{b}_\zeta+\frac{1}{144}\sum_{ijkqrst}\widehat{h}_{ijk}\widehat{h}_{qrst}u_iu_ju_ku_qu_ru_su_tu_\zeta\widehat{b}_\zeta\\
&\quad-\frac{1}{1296}\sum_{ijkqrstwv}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\zeta\widehat{b}_\zeta,\\
I_n^{14} &= -\frac{1}{720}\sum_{ijkqrs\zeta}\widehat{h}_{ijkqrs}u_iu_ju_ku_qu_ru_su_\zeta\widehat{b}_\zeta+\frac{1}{1152}\sum_{ijkqrstw\zeta}\widehat{h}_{ijkq}\widehat{h}_{rstw}u_iu_ju_ku_qu_ru_su_tu_wu_\zeta\widehat{b}_\zeta\\
&\quad+\frac{1}{720}\sum_{ijkqrstw\zeta}\widehat{h}_{ijk}\widehat{h}_{qrstw}u_iu_ju_ku_qu_ru_su_tu_wu_\zeta\widehat{b}_\zeta\\
&\quad-\frac{1}{1728}\sum_{ijkqrstwv\beta\zeta ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv\beta}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\zeta\widehat{b}_\zeta\\
&\quad+\frac{1}{31104}\sum_{ijkqrstwv\beta\tau\phi\zeta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\beta u_\tau u_\phi u_\zeta\widehat{b}_\zeta,\\
I_n^{21} &= -\frac{1}{6}\sum_{ijk\zeta\eta}\widehat{h}_{ijk}u_iu_ju_ku_\zeta u_\eta\widehat{b}_{\zeta\eta},\\
I_n^{22} &= -\frac{1}{24}\sum_{ijkq\zeta\eta}\widehat{h}_{ijkq}u_iu_ju_ku_qu_\zeta u_\eta\widehat{b}_{\zeta\eta}+\frac{1}{72}\sum_{ijkqrs\zeta\eta}\widehat{h}_{ijk}\widehat{h}_{qrs}u_iu_ju_ku_qu_ru_su_\zeta u_\eta\widehat{b}_{\zeta\eta},\\
I_n^{23} &= -\frac{1}{120}\sum_{ijkqr\zeta\eta}\widehat{h}_{ijkqr}u_iu_ju_ku_qu_ru_\zeta u_\eta\widehat{b}_{\zeta\eta}+\frac{1}{144}\sum_{ijkqrst\zeta\eta}\widehat{h}_{ijk}\widehat{h}_{qrst}u_iu_ju_ku_qu_ru_su_tu_\zeta u_\eta\widehat{b}_{\zeta\eta}\\
&\quad-\frac{1}{1296}\sum_{ijkqrstwv\zeta\eta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}u_iu_ju_ku_qu_ru_su_tu_wu_vu_\zeta u_\eta\widehat{b}_{\zeta\eta},
\end{aligned}
$$

$$
\begin{aligned}
I_n^{31} &= -\frac{1}{6}\sum_{ijk\zeta\eta}\widehat{h}_{ijk}u_iu_ju_ku_\zeta u_\eta u_\xi \widehat{b}_{\zeta\eta\xi},\\
I_n^{32} &= -\frac{1}{24}\sum_{ijkq\zeta\eta}\widehat{h}_{ijkq}u_iu_ju_ku_qu_\zeta u_\eta u_\xi \widehat{b}_{\zeta\eta\xi}+\frac{1}{72}\sum_{ijkqrs\zeta\eta}\widehat{h}_{ijk}\widehat{h}_{qrs}u_iu_ju_ku_qu_ru_su_\zeta u_\eta u_\xi \widehat{b}_{\zeta\eta\xi},\\
I_n^{41} &= -\frac{1}{6}\sum_{ijk\zeta\eta}\widehat{h}_{ijk}u_iu_ju_ku_\zeta u_\eta u_\xi u_\omega \widehat{b}_{\zeta\eta\xi\omega},
\end{aligned}
$$

Let $f(u)$ be the pdf of the multivariate normal distribution with mean 0 and covariance matrix $\left(\nabla^2\widehat{h}\right)^{-1}$. Then, we have

$$
\begin{aligned}
&\int_{R^p} I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) f(u)\,du\\
=\;& \widehat{b}+\frac{1}{n}\left(-\frac{1}{24}\sum_{ijkq}\widehat{h}_{ijkq}\mu^4_{ijkq}+\frac{1}{72}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs}-\frac{1}{6}\sum_{ijk\zeta}\widehat{h}_{ijk}\mu^4_{ijk\zeta}\widehat{b}_\zeta+\frac{1}{2}\sum_{\zeta\eta}\widehat{b}_{\zeta\eta}\widehat{h}^{\zeta\eta}\right)\\
&+\frac{1}{n^2}\left(\begin{array}{c}
-\frac{1}{720}\sum_{ijkqrs}\widehat{h}_{ijkqrs}\mu^6_{ijkqrs}+\frac{1}{1152}\sum_{ijkqrstw}\widehat{h}_{ijkq}\widehat{h}_{rstw}\mu^8_{ijkqrstw}\\
+\frac{1}{720}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrstw}\mu^8_{ijkqrstw}-\frac{1}{1728}\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv\beta}\mu^{10}_{ijkqrstwv\beta}\\
+\frac{1}{31104}\sum_{ijkqrstwv\beta\tau\phi}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}\mu^{12}_{ijkqrstwv\beta\tau\phi}\\
-\frac{1}{120}\sum_{ijkqr\zeta}\widehat{h}_{ijkqr}\mu^6_{ijkqr\zeta}\widehat{b}_\zeta+\frac{1}{144}\sum_{ijkqrst\zeta}\widehat{h}_{ijk}\widehat{h}_{qrst}\mu^8_{ijkqrst\zeta}\widehat{b}_\zeta\\
-\frac{1}{1296}\sum_{ijkqrstwv\zeta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\mu^{10}_{ijkqrstwv\zeta}\widehat{b}_\zeta\\
-\frac{1}{48}\sum_{ijkq\zeta\eta}\widehat{h}_{ijkq}\mu^6_{ijkq\zeta\eta}\widehat{b}_{\zeta\eta}+\frac{1}{144}\sum_{ijkqrs\zeta\eta}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrs\zeta\eta}\widehat{b}_{\zeta\eta}\\
-\frac{1}{36}\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{\zeta\eta\xi}+\frac{1}{24}\sum_{\zeta\eta\xi\omega}\widehat{b}_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}
\end{array}\right)\\
&+O\left(n^{-3}\right).
\end{aligned}
$$

Note that the odd order central moments of the multivariate normal distribution are all zero. By expanding the domain from $B_{\sqrt{n}\delta}(0)$ to R^P, the error can be expressed as

$$
\begin{aligned}
&\int_{R^p-B_{\sqrt{n}\delta}(0)}\sum_{ijkq}\widehat{h}_{ijkq}u_iu_ju_ku_q f(u)\,du\\
=\;&\sum_{ijkq}\widehat{h}_{ijkq}\int_{R^p-B_{\sqrt{n}\delta}(0)}u_iu_ju_ku_q f(u)\,du\\
=\;&(2\pi)^{-\frac{P}{2}}\left|\nabla^2\widehat{h}\right|^{\frac{1}{2}}\sum_{ijkq}\widehat{h}_{ijkq}\int_{R^P-B_{\sqrt{n}\delta}(0)}u_iu_ju_ku_q\exp\left[-\frac{1}{2}\sum_{ij}\widehat{h}_{ij}u_iu_j\right]du
\end{aligned}
$$

$$
\begin{aligned}
&= (2\pi)^{-\frac{P}{2}} \left|\nabla^2 \widehat{h}\right|^{\frac{1}{2}} \sum_{ijkq} \widehat{h}_{ijkq} \int_{R^P - B_{\sqrt{n}\delta}(0)} u_i u_j u_k u_q \exp\left[-\frac{1}{4}\sum_{ij} \widehat{h}_{ij} u_i u_j\right] \exp\left[-\frac{1}{4}\sum_{ij} \widehat{h}_{ij} u_i u_j\right] du \\
&\leq (2\pi)^{-\frac{P}{2}} \left|\nabla^2 \widehat{h}\right|^{\frac{1}{2}} \sum_{ijkq} \widehat{h}_{ijkq} \int_{R^P - B_{\sqrt{n}\delta}(0)} u_i u_j u_k u_q \exp\left[-\frac{1}{4}\sum_{ij} \lambda_{\min} u_i u_j\right] \exp\left[-\frac{1}{4}\sum_{ij} \widehat{h}_{ij} u_i u_j\right] du \\
&\leq (2\pi)^{-\frac{P}{2}} \left|\nabla^2 \widehat{h}\right|^{\frac{1}{2}} \exp\left[-\frac{1}{4}\lambda_{\min}\delta^2 n\right] \sum_{ijkq} \widehat{h}_{ijkq} \int_{R^P} u_i u_j u_k u_q \exp\left[-\frac{1}{2}\sum_{ij}\left(\frac{1}{2}\widehat{h}_{ij}\right) u_i u_j\right] du \\
&= M' \exp\left[-\frac{1}{4}\lambda_{\min}\delta^2 n\right],
\end{aligned}
$$

where $\lambda_{\min} > 0$ is the smallest eigenvalue of $\nabla^2 \widehat{h}$ and

$$
M' = (2\pi)^{-\frac{P}{2}} \left|\nabla^2 \widehat{h}\right|^{\frac{1}{2}} \sum_{ijkq} \widehat{h}_{ijkq} \int_{R^P} u_i u_j u_k u_q \exp\left[-\sum_{ij}\left(\frac{1}{4}\widehat{h}_{ij}\right) u_i u_j\right] du < \infty.
$$

The first inequality follows from the fact that

$$
\lambda_{\min} = \min_{\|e\|=1} f(e) = e' A e,
$$

where A is a positive definite matrix, and $\lambda_{\min}$ is the smallest eigenvalue of A. Here, we only express one term in $I_n\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_n\right)$. The other terms can be analyzed in the same way.

Hence, we have

$$
\int_{B_{\sqrt{n}\delta}(0)} I_n\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_n\right) f(u)\, du = \int_{R^P} I_n\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_n\right) f(u)\, du - \int_{R^P - B_{\sqrt{n}\delta}(0)} I_n\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_n\right) f(u)\, du,
$$

and

$$
\begin{aligned}
&\int_{\boldsymbol{\Theta}} b(\boldsymbol{\theta}) \exp\left[-nh(\boldsymbol{\theta})\right] d\boldsymbol{\theta} \\
&= \int_{B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta}) \exp\left[-nh(\boldsymbol{\theta})\right] d\boldsymbol{\theta} + \int_{\boldsymbol{\Theta} - B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta}) \exp\left[-nh(\boldsymbol{\theta})\right] d\boldsymbol{\theta} \\
&= (2\pi)^{\frac{P}{2}} \left|n\nabla^2 \widehat{h}\right|^{-\frac{1}{2}} \exp\left(-n\widehat{h}\right) \left[\int_{B_{\sqrt{n}\delta}(0)} \left[I_n\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_n\right) + R_n\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_n\right)\right] f(u)\, du + O\left(n^{-3}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= (2\pi)^{\frac{P}{2}} \left|n\nabla^2\widehat{h}\right|^{-\frac{1}{2}} \exp\left(-n\widehat{h}\right)\left[\int_{B_{\sqrt{n}\delta}(0)} I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) f(u)\,du + O\left(n^{-3}\right)\right] \\
&= (2\pi)^{\frac{P}{2}} \left|n\nabla^2\widehat{h}\right|^{-\frac{1}{2}} \exp\left(-n\widehat{h}\right)\left[\begin{array}{c} \int_{R^P} I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) f(u)\,du \\ -\int_{R^P - B_{\sqrt{n}\delta}(0)} I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) f(u)\,du \end{array} + O\left(n^{-3}\right)\right] \\
&= (2\pi)^{\frac{P}{2}} \left|n\nabla^2\widehat{h}\right|^{-\frac{1}{2}} \exp\left(-n\widehat{h}\right)\left[\int_{R^P} I_n\left(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}_n\right) f(u)\,du + O\left(n^{-3}\right)\right].
\end{aligned}
$$

Hence, this lemma is proved. ■

5.4.2 High-order stochastic expansions

In this subsection we develop high-order stochastic Laplace expansions. Suppose $\mathbf{y}$ is defined on a common probability space $\left\{\Omega, \mathcal{F}, \wp_{\boldsymbol{\theta}}\right\}$, where Ω is a sample space, $\mathcal{F}$ is a sigma-algebra, and $\wp_{\boldsymbol{\theta}}$ is a probability measure that depends on parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, a compact subset of R^P. Assume $\{y_i, i = 1, 2, \cdots\}$ take values in the same subset of R. Let $h_n(\mathbf{y},\boldsymbol{\theta})$ be a sequence of functions, each of which is eight-times continuously differentiable with respect to $\boldsymbol{\theta}$ and has an interior global minimum $\left\{\widehat{\boldsymbol{\theta}}_n\right\}$ and $b(\boldsymbol{\theta})$ be a six-times continuously differentiable real function of $\boldsymbol{\theta}$. When there is no confusion, we write $h_n(\mathbf{y},\boldsymbol{\theta})$ as $h_n(\boldsymbol{\theta})$ or h_n or even h.

We call the pair $(\{h_n\}, b)$ satisfies the analytical assumptions for the stochastic Laplace method on $\wp_{\boldsymbol{\theta}}$ if the following assumptions are satisfied. There exists positive numbers ε, M and η such that (i) with probability approach one (w.p.a.1), for all $\boldsymbol{\theta} \in B_\varepsilon\left(\widehat{\boldsymbol{\theta}}_n\right)$ and all $1 \leq j_1, \cdots, j_d \leq P$ with $0 \leq d \leq 8$, $\|h_n(\boldsymbol{\theta})\| < M$ and $\|h_{j_1 \cdots j_d}(\boldsymbol{\theta})\| < M$; (ii) w.p.a.1, $\nabla^2\widehat{h}$ is positive definite and $\det\left(\nabla^2\widehat{h}\right) > \eta$; (iii) $\int_{\boldsymbol{\Theta}} b(\boldsymbol{\theta}) \exp\left(-nh_n(\boldsymbol{\theta})\right) d\boldsymbol{\theta}$ exists and is finite, and for all δ for which $0 < \delta < \varepsilon$ and $B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right) \subseteq \boldsymbol{\Theta}$,

$$
\left[\det\left(n\nabla^2\widehat{h}\right)\right]^{\frac{1}{2}} \int_{\boldsymbol{\Theta} - B_\delta\left(\widehat{\boldsymbol{\theta}}_n\right)} b(\boldsymbol{\theta}) \exp\left[-n\left(h_n(\boldsymbol{\theta}) - \widehat{h}\right)\right] d\boldsymbol{\theta} = O_p\left(n^{-3}\right).
$$

Note that the assumptions above are related to but slightly different from those in Section 3 of Kass et al. (1990). They differ in two aspects. First, we require $h_n(\boldsymbol{\theta})$ be eight-times continuously differentiable and $b(\boldsymbol{\theta})$ be six-times continuously differentiable. Second, for conditions (ii) and (iii), instead of almost sure boundedness and almost sure convergence, we assume they hold w.p.a.1. We do so because we are interested in convergence in probability only. Following the result in Theorem 7 of Kass et al. (1990), $(\{h_n\}, b)$ satisfy the analytical assumptions for stochastic Laplace's method on $\wp_{\boldsymbol{\theta}}$ and Lemma 5.1 above, it is straightforward to show that

$$\int_{\boldsymbol{\Theta}} b(\boldsymbol{\theta}) \exp\left[-nh_n(\boldsymbol{\theta})\right] d\boldsymbol{\theta} = (2\pi)^{\frac{P}{2}} \left[\det\left(n\nabla^2\widehat{h}\right)\right]^{-\frac{1}{2}} \exp\left(-n\widehat{h}\right)\left(\widehat{b} + \frac{1}{n}Q_1 + \frac{1}{n^2}Q_2 + O_p\left(n^{-3}\right)\right), \tag{44}$$

where the expressions for Q_1 and Q_2 are given in Lemma 5.1.

Lemma 5.2 *If both* $(\{h_n\}, g \times b_D)$ *and* $(\{h_n\}, b_D)$ *satisfy the analytical assumptions for the stochastic Laplace method on* $\wp_{\boldsymbol{\theta}}$*, then*

$$\frac{\int g(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) \exp\left(-nh_n(\boldsymbol{\theta})\right) d\boldsymbol{\theta}}{\int b_D(\boldsymbol{\theta}) \exp\left(-nh_n(\boldsymbol{\theta})\right) d\boldsymbol{\theta}} = \widehat{g} + \frac{1}{n}B_1 + \frac{1}{n^2}\left(B_2 - B_3\right) + O_p\left(\frac{1}{n^3}\right),$$

where

$$\begin{aligned}
B_1 &= \frac{1}{2}\sum_{ij}\widehat{\sigma}_{ij}\widehat{g}_{ij} + \frac{\sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{D,j}\widehat{g}_i}{\widehat{b}_D} - \frac{1}{6}\sum_{ijkq}\widehat{h}_{ijk}\mu^4_{ijkq}\widehat{g}_q, \\
B_2 &= -\frac{1}{120}\sum_{ijkqrs}\widehat{h}_{ijkqr}\mu^6_{ijkqrs}\widehat{g}_s + \frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\mu^8_{ijkqrstw}\widehat{g}_w \\
&\quad -\frac{1}{1296}\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\mu^{10}_{ijkqrstwv\beta}\widehat{g}_\beta - \frac{1}{24}\frac{\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{b}_{D,s}\widehat{g}_r}{\widehat{b}_D} \\
&\quad +\frac{1}{72}\frac{\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{b}_{D,w}\widehat{g}_t}{\widehat{b}_D} - \frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{D,\eta\xi}\widehat{g}_\zeta}{\widehat{b}_D} \\
&\quad +\frac{1}{6}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{b}_{D,\eta\xi\omega}\widehat{g}_\zeta}{\widehat{b}_D} - \frac{1}{48}\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_{rs} \\
&\quad +\frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{g}_{tw} - \frac{1}{36}\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta\xi} \\
&\quad +\frac{1}{24}\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta\xi\omega} - \frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi}}{\widehat{b}_D} \\
&\quad +\frac{1}{6}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta\xi}\widehat{b}_{D,\omega}}{\widehat{b}_D} + \frac{1}{4}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D},
\end{aligned}$$

$$B_3 = \left(\frac{1}{2}\sum_{ij}\widehat{\sigma}_{ij}\frac{\widehat{b}_{D,ij}}{\widehat{b}_D} - \frac{1}{6}\sum_{ijkq}\widehat{h}_{ijk}\mu^4_{ijkq}\frac{\widehat{b}_{D,q}}{\widehat{b}_D} + \frac{1}{72}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs} - \frac{1}{24}\sum_{ijkq}\widehat{h}_{ijkq}\mu^4_{ijkq}\right)B_1,$$

with $\sigma_{ij} = h^{ij}$.

Proof. If $\left\{h^N(\boldsymbol{\theta}), b_N\right\}$ and $\left\{h^D(\boldsymbol{\theta}), b_D\right\}$ satisfy the analytical assumptions for the stochastic Laplace method on $\wp_{\boldsymbol{\theta}}$, then, by (44)

$$\frac{\int b_N(\boldsymbol{\theta})\exp\left[-nh^N(\boldsymbol{\theta})\right]d\boldsymbol{\theta}}{\int b_D(\boldsymbol{\theta})\exp\left[-nh^D(\boldsymbol{\theta})\right]d\boldsymbol{\theta}} = \frac{\left|\nabla^2\widehat{h}^N\right|^{-\frac{1}{2}}\exp\left[-nh\left(\widehat{\boldsymbol{\theta}}^N_n\right)\right]}{\left|\nabla^2\widehat{h}^D\right|^{-\frac{1}{2}}\exp\left[-nh\left(\widehat{\boldsymbol{\theta}}^D_n\right)\right]}\frac{b_k\left(\widehat{\boldsymbol{\theta}}^N_n\right) + \frac{1}{n}c_N + \frac{1}{n^2}d_N + O_p\left(n^{-3}\right)}{b_k\left(\widehat{\boldsymbol{\theta}}^D_n\right) + \frac{1}{n}c_D + \frac{1}{n^2}d_D + O_p\left(n^{-3}\right)}.$$

From Tierney and Kadane (1986) and Miyata (2004, 2010), we have

$$\frac{b_N\left(\widehat{\boldsymbol{\theta}}^N_n\right) + \frac{1}{n}c_N + \frac{1}{n^2}d_N + O_p\left(n^{-3}\right)}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right) + \frac{1}{n}c_D + \frac{1}{n^2}d_D + O_p\left(n^{-3}\right)} = \frac{b_N\left(\widehat{\boldsymbol{\theta}}^N_n\right)}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right)}\left[\frac{1 + \frac{1}{n}\frac{c_N}{b_N\left(\widehat{\boldsymbol{\theta}}^N_n\right)} + \frac{1}{n^2}\frac{d_N}{b_N\left(\widehat{\boldsymbol{\theta}}^N_n\right)} + O_p\left(n^{-3}\right)}{1 + \frac{1}{n}\frac{c_D}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right)} + \frac{1}{n^2}\frac{d_D}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right)} + O_p\left(n^{-3}\right)}\right]$$

$$= \frac{b_N\left(\widehat{\boldsymbol{\theta}}^N_n\right)}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right)}\left\{\begin{array}{c} 1 + \frac{1}{n}\left(\frac{c_N}{b_N\left(\widehat{\boldsymbol{\theta}}^N_n\right)} - \frac{c_D}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right)}\right) \\ + \frac{1}{n^2}\left(\frac{d_N}{b_N\left(\widehat{\boldsymbol{\theta}}^N_n\right)} - \frac{d_D}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right)} - \frac{c_D}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right)}\left(\frac{c_N}{b_N\left(\widehat{\boldsymbol{\theta}}^N_n\right)} - \frac{c_D}{b_D\left(\widehat{\boldsymbol{\theta}}^D_n\right)}\right)\right) + O_p\left(n^{-3}\right)\end{array}\right\},$$

where

$$\begin{aligned}
\frac{\widehat{b}_N}{\widehat{b}_D}\left(\frac{c_N}{\widehat{b}_N} - \frac{c_D}{\widehat{b}_D}\right) &= \frac{\widehat{b}_N\left(c_N\widehat{b}_D - c_D\widehat{b}_N\right)}{\widehat{b}^2_D\widehat{b}_N} = \frac{c_N\widehat{b}_D - c_D\widehat{b}_N}{\widehat{b}^2_D} \\
&= \frac{1}{2}\frac{\sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{N,ij}\widehat{b}_D - \sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{D,ij}\widehat{b}_N}{\widehat{b}^2_D} \\
&\quad -\frac{1}{6}\frac{\sum_{ijkq}\widehat{h}^N_{ijk}\mu^4_{ijkq}\widehat{b}_{N,q}\widehat{b}_D - \sum_{ijkq}\widehat{h}^D_{ijk}\mu^4_{ijkq}\widehat{b}_{D,q}\widehat{b}_N}{\widehat{b}^2_D} \\
&\quad +\frac{1}{72}\frac{\widehat{b}_N\sum_{ijkqrs}\widehat{h}^N_{ijk}\widehat{h}^N_{qrs}\mu^6_{ijkqrs} - \sum_{ijkqrs}\widehat{h}^D_{ijk}\widehat{h}^D_{qrs}\mu^6_{ijkqrs}\widehat{b}_N}{\widehat{b}_D}
\end{aligned}$$

$$-\frac{1}{24}\frac{\widehat{b}_N\sum_{ijkq}\widehat{h}^N_{ijkq}\mu^4_{ijkq}-\sum_{ijkq}\widehat{h}^D_{ijkq}\mu^4_{ijkq}\widehat{b}_N}{\widehat{b}_D}.$$

$$\begin{aligned}
&\frac{\widehat{b}_N}{\widehat{b}_D}\left(\frac{d_N}{\widehat{b}_N}-\frac{d_D}{\widehat{b}_D}\right)\\
= \quad & -\frac{1}{720}\frac{\widehat{b}_N\sum_{ijkqrs}\widehat{h}^N_{ijkqrs}\mu^6_{ijkqrs}-\sum_{ijkqrs}\widehat{h}^D_{ijkqrs}\mu^6_{ijkqrs}\widehat{b}_N}{\widehat{b}_D}\\
&+\frac{1}{1152}\frac{\widehat{b}_N\sum_{ijkqrstw}\widehat{h}^N_{ijkq}\widehat{h}^N_{rstw}\mu^8_{ijkqrstw}-\sum_{ijkqrstw}\widehat{h}^D_{ijkq}\widehat{h}^D_{rstw}\mu^8_{ijkqrstw}\widehat{b}_N}{\widehat{b}_D}\\
&+\frac{1}{720}\frac{\widehat{b}_N\sum_{ijkqrstw}\widehat{h}^N_{ijk}\widehat{h}^N_{qrstw}\mu^8_{ijkqrstw}-\sum_{ijkqrstw}\widehat{h}^D_{ijk}\widehat{h}^D_{qrstw}\mu^8_{ijkqrstw}\widehat{b}_N}{\widehat{b}_D}\\
&-\frac{1}{1728}\frac{\widehat{b}_N\sum_{ijkqrstwv\beta}\widehat{h}^N_{ijk}\widehat{h}^N_{qrs}\widehat{h}^N_{twv}\mu^{10}_{ijkqrstwv\beta}-\sum_{ijkqrstwv\beta}\widehat{h}^D_{ijk}\widehat{h}^D_{qrs}\widehat{h}^D_{twv}\mu^{10}_{ijkqrstwv\beta}\widehat{b}_N}{\widehat{b}_D}\\
&+\frac{1}{31104}\frac{\begin{pmatrix}\widehat{b}_N\sum_{ijkqrstwv\beta\tau\phi}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}\mu^{12}_{ijkqrstwv\beta\tau\phi}\\ -\sum_{ijkqrstwv\beta\tau\phi}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{h}_{\beta\tau\phi}\mu^{12}_{ijkqrstwv\beta\tau\phi}\widehat{b}_N\end{pmatrix}}{\widehat{b}_D}\\
&-\frac{1}{120}\frac{\sum_{ijkqrs}\widehat{h}^N_{ijkqr}\mu^6_{ijkqrs}\widehat{b}_{N,s}\widehat{b}_D-\sum_{ijkqrs}\widehat{h}^D_{ijkqr}\mu^6_{ijkqrs}\widehat{b}_{D,s}\widehat{b}_N}{\widehat{b}^2_D}\\
&+\frac{1}{144}\frac{\sum_{ijkqrstw}\widehat{h}^N_{ijk}\widehat{h}^N_{qrst}\mu^8_{ijkqrstw}\widehat{b}_{N,w}\widehat{b}_D-\sum_{ijkqrstw}\widehat{h}^N_{ijk}\widehat{h}^N_{qrst}\mu^8_{ijkqrstw}\widehat{b}_{D,w}\widehat{b}_N}{\widehat{b}^2_D}\\
&-\frac{1}{1296}\frac{\sum_{ijkqrstwv\beta}\widehat{h}^N_{ijk}\widehat{h}^N_{qrs}\widehat{h}^N_{twv}\mu^{10}_{ijkqrstwv\beta}\widehat{b}_{N,\beta}\widehat{b}_D-\sum_{ijkqrstwv\beta}\widehat{h}^D_{ijk}\widehat{h}^D_{qrs}\widehat{h}^D_{twv}\mu^{10}_{ijkqrstwv\beta}\widehat{b}_{D,\beta}\widehat{b}_N}{\widehat{b}^2_D}\\
&-\frac{1}{48}\frac{\sum_{ijkqrs}\widehat{h}^N_{ijkq}\mu^6_{ijkqrs}\widehat{b}_{N,rs}\widehat{b}_D-\sum_{ijkqrs}\widehat{h}^D_{ijkq}\mu^6_{ijkqrs}\widehat{b}_{D,rs}\widehat{b}_N}{\widehat{b}^2_D}\\
&+\frac{1}{144}\frac{\sum_{ijkqrstw}\widehat{h}^N_{ijk}\widehat{h}^N_{qrs}\mu^8_{ijkqrstw}\widehat{b}_{N,tw}\widehat{b}_D-\sum_{ijkqrstw}\widehat{h}^D_{ijk}\widehat{h}^D_{qrs}\mu^8_{ijkqrstw}\widehat{b}_{D,tw}\widehat{b}_N}{\widehat{b}^2_D}\\
&-\frac{1}{36}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}^N_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{N,\zeta\eta\xi}\widehat{b}_D-\sum_{ijk\zeta\eta\xi}\widehat{h}^D_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{D,\zeta\eta\xi}\widehat{b}_N}{\widehat{b}^2_D}\\
&+\frac{1}{24}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{b}_{N,\zeta\eta\xi\omega}\widehat{b}_D-\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{b}_{D,\zeta\eta\xi\omega}\widehat{b}_N}{\widehat{b}^2_D}
\end{aligned}$$

$$
\begin{aligned}
\frac{c_D}{\widehat{b}_D}\frac{\widehat{b}_N}{\widehat{b}_D}\left(\frac{c_N}{\widehat{b}_N}-\frac{c_D}{\widehat{b}_D}\right) &= \frac{1}{2}c_D\frac{\sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{N,ij}\widehat{b}_D-\sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{D,ij}\widehat{b}_N}{\widehat{b}_D^3} \\
&\quad -\frac{1}{6}c_D\frac{\sum_{ijkq}\widehat{h}_{ijk}^N\mu_{ijkq}\widehat{b}_{N,q}\widehat{b}_D-\sum_{ijkq}\widehat{h}_{ijk}^D\widehat{h}_{qrs}^D\mu_{ijkqrs}^6\widehat{b}_{D,q}\widehat{b}_N}{\widehat{b}_D^3} \\
&\quad +\frac{1}{72}c_D\frac{\widehat{b}_N\sum_{ijkqrs}\widehat{h}_{ijk}^N\widehat{h}_{qrs}^N\mu_{ijkqrs}^6-\sum_{ijkq}\widehat{h}_{ijk}^D\widehat{h}_{qrs}^D\mu_{ijkqrs}^6\widehat{b}_N}{\widehat{b}_D^2} \\
&\quad -\frac{1}{24}c_D\frac{\widehat{b}_N\sum_{ijkq}\widehat{h}_{ijkq}^N\mu_{ijkq}^4-\sum_{ijkq}\widehat{h}_{ijkq}^D\mu_{ijkq}^4\widehat{b}_N}{\widehat{b}_D^2}.
\end{aligned}
$$

If we set $b_N(\boldsymbol{\theta}) = g(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta})$ and $h^N(\boldsymbol{\theta}) = h^D(\boldsymbol{\theta}) = h(\boldsymbol{\theta})$, we can show that, for the derivatives of $b_N(\boldsymbol{\theta})$,

$$
b_{N,i}(\boldsymbol{\theta}) = g_i(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) + g(\boldsymbol{\theta})\, b_{D,i}(\boldsymbol{\theta}),
$$

$$
b_{N,ij}(\boldsymbol{\theta}) = g_{ij}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) + g_i(\boldsymbol{\theta})\, b_{D,j}(\boldsymbol{\theta}) + g_j(\boldsymbol{\theta})\, b_{D,i}(\boldsymbol{\theta}) + g(\boldsymbol{\theta})\, b_{D,ij}(\boldsymbol{\theta}),
$$

$$
\begin{aligned}
b_{N,ijk}(\boldsymbol{\theta}) &= g_{ijk}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) + g_{ij}(\boldsymbol{\theta})\, b_{D,k}(\boldsymbol{\theta}) + g_{ik}(\boldsymbol{\theta})\, b_{D,j}(\boldsymbol{\theta}) + g_i(\boldsymbol{\theta})\, b_{D,jk}(\boldsymbol{\theta}) + g_{jk}(\boldsymbol{\theta})\, b_{D,i}(\boldsymbol{\theta}) \\
&\quad + g_j(\boldsymbol{\theta})\, b_{D,ik}(\boldsymbol{\theta}) + g_k(\boldsymbol{\theta})\, b_{D,ij}(\boldsymbol{\theta}) + g(\boldsymbol{\theta})\, b_{D,ijk}(\boldsymbol{\theta}),
\end{aligned}
$$

$$
\begin{aligned}
b_{N,ijkq}(\boldsymbol{\theta}) &= g_{ijkq}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) + g_{ijk}(\boldsymbol{\theta})\, b_{D,q}(\boldsymbol{\theta}) + g_{ijq}(\boldsymbol{\theta})\, b_{D,k}(\boldsymbol{\theta}) + g_{ij}(\boldsymbol{\theta})\, b_{D,kq}(\boldsymbol{\theta}) \\
&\quad + g_{ikq}(\boldsymbol{\theta})\, b_{D,j}(\boldsymbol{\theta}) + g_{ik}(\boldsymbol{\theta})\, b_{D,jq}(\boldsymbol{\theta}) + g_{iq}(\boldsymbol{\theta})\, b_{D,jk}(\boldsymbol{\theta}) + g_i(\boldsymbol{\theta})\, b_{D,jkq}(\theta) \\
&\quad + g_{jkq}(\boldsymbol{\theta})\, b_{D,i}(\boldsymbol{\theta}) + g_{jk}(\boldsymbol{\theta})\, b_{D,iq}(\boldsymbol{\theta}) + g_{jq}(\boldsymbol{\theta})\, b_{D,ik}(\boldsymbol{\theta}) + g_j(\boldsymbol{\theta})\, b_{D,ikq}(\theta) \\
&\quad + g_{kq}(\boldsymbol{\theta})\, b_{D,ij}(\boldsymbol{\theta}) + g_k(\boldsymbol{\theta})\, b_{D,ijq}(\boldsymbol{\theta}) + g_q(\boldsymbol{\theta})\, b_{D,ijk}(\boldsymbol{\theta}) + g(\boldsymbol{\theta})\, b_{D,ijkq}(\theta).
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
& b_{N,ij}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) - b_{D,ij}(\boldsymbol{\theta})\, b_N(\boldsymbol{\theta}) \qquad (45) \\
= & \left[g_{ij}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) + g_i(\boldsymbol{\theta})\, b_{D,j}(\boldsymbol{\theta}) + g_j(\theta)\, b_{D,i}(\boldsymbol{\theta}) + g(\boldsymbol{\theta})\, b_{D,ij}(\boldsymbol{\theta})\right] b_D(\boldsymbol{\theta}) - b_{D,ij}(\boldsymbol{\theta})\, g(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) \\
= & g_{ij}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta})^2 + g_i(\boldsymbol{\theta})\, b_{D,j}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) + g_j(\boldsymbol{\theta})\, b_{D,i}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) \\
& + g(\boldsymbol{\theta})\, b_{D,ij}(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) - b_{D,ij}(\boldsymbol{\theta})\, g(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta})
\end{aligned}
$$

$$= g_{ij}(\boldsymbol{\theta}) b_D(\boldsymbol{\theta})^2 + g_i(\boldsymbol{\theta}) b_{D,j}(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}) + g_j(\boldsymbol{\theta}) b_{D,i}(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}),$$

$$\begin{aligned} & b_{N,i}(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}) - b_{D,i}(\boldsymbol{\theta}) b_N(\boldsymbol{\theta}) \\ = & \left(g_i(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}) + g(\boldsymbol{\theta}) b_{D,i}(\boldsymbol{\theta})\right) b_D(\boldsymbol{\theta}) - b_{D,i}(\boldsymbol{\theta}) g(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}) = g_i(\boldsymbol{\theta}) b_D(\boldsymbol{\theta})^2, \end{aligned} \tag{46}$$

$$\begin{aligned} & b_{N,ijk}(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}) - b_{D,ijk}(\boldsymbol{\theta}) b_N(\boldsymbol{\theta}) \\ = & \left[\begin{array}{c} g_{ijk}(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}) + g_{ij}(\boldsymbol{\theta}) b_{D,k}(\boldsymbol{\theta}) + g_{ik}(\boldsymbol{\theta}) b_{D,j}(\boldsymbol{\theta}) + g_i(\boldsymbol{\theta}) b_{D,jk}(\boldsymbol{\theta}) \\ + g_{jk}(\boldsymbol{\theta}) b_{D,i}(\boldsymbol{\theta}) + g_j(\boldsymbol{\theta}) b_{D,ik}(\boldsymbol{\theta}) + g_k(\boldsymbol{\theta}) b_{D,ij}(\boldsymbol{\theta}) \end{array}\right] b_D(\boldsymbol{\theta}), \end{aligned} \tag{47}$$

$$\begin{aligned} & b_{N,ijkq}(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}) - b_{D,ijkq}(\boldsymbol{\theta}) b_N(\boldsymbol{\theta}) \\ = & \left[\begin{array}{c} g_{ijkq}(\boldsymbol{\theta}) b_D(\boldsymbol{\theta}) + g_{ijk}(\boldsymbol{\theta}) b_{D,q}(\boldsymbol{\theta}) + g_{ijq}(\boldsymbol{\theta}) b_{D,k}(\boldsymbol{\theta}) + g_{ij}(\boldsymbol{\theta}) b_{D,kq}(\theta) \\ + g_{ikq}(\boldsymbol{\theta}) b_{D,j}(\boldsymbol{\theta}) + g_{ik}(\boldsymbol{\theta}) b_{D,jq}(\boldsymbol{\theta}) + g_{iq}(\boldsymbol{\theta}) b_{D,jk}(\boldsymbol{\theta}) + g_i(\boldsymbol{\theta}) b_{D,jkq}(\theta) \end{array}\right] b_D(\boldsymbol{\theta}) \\ & + \left[\begin{array}{c} g_{jkq}(\boldsymbol{\theta}) b_{D,i}(\boldsymbol{\theta}) + g_{jk}(\boldsymbol{\theta}) b_{D,iq}(\boldsymbol{\theta}) + g_{jq}(\boldsymbol{\theta}) b_{D,ik}(\boldsymbol{\theta}) + g_j(\boldsymbol{\theta}) b_{D,ikq}(\theta) \\ + g_{kq}(\boldsymbol{\theta}) b_{D,ij}(\boldsymbol{\theta}) + g_k(\boldsymbol{\theta}) b_{D,ijq}(\boldsymbol{\theta}) + g_q(\boldsymbol{\theta}) b_{D,ijk}(\boldsymbol{\theta}) \end{array}\right] b_D(\boldsymbol{\theta}). \end{aligned} \tag{48}$$

Consequently,

$$\begin{aligned} \frac{\widehat{b}_N}{\widehat{b}_D}\left(\frac{c_N}{\widehat{b}_N} - \frac{c_D}{\widehat{b}_D}\right) &= \frac{1}{2}\frac{\sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{N,ij}\widehat{b}_D - \sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{D,ij}\widehat{b}_N}{\widehat{b}_D^2} \\ & \quad -\frac{1}{6}\frac{\sum_{ijkq}\widehat{h}^N_{ijk}\mu^4_{ijkq}\widehat{b}_{N,q}\widehat{b}_D - \sum_{ijkq}\widehat{h}^D_{ijk}\mu^4_{ijkq}\widehat{b}_{D,q}\widehat{b}_N}{\widehat{b}_D^2} \\ &= \frac{1}{2}\frac{\sum_{ij}\widehat{\sigma}_{ij}\left(\widehat{b}_{N,ij}\widehat{b}_D - \widehat{b}_{D,ij}\widehat{b}_N\right)}{\widehat{b}_D^2} - \frac{1}{6}\frac{\sum_{ijkq}\widehat{h}_{ijk}\mu^4_{ijkq}\left(\widehat{b}_{N,q}\widehat{b}_D - \widehat{b}_{D,q}\widehat{b}_N\right)}{\widehat{b}_D^2}, \end{aligned}$$

where

$$\frac{\sum_{ij}\widehat{\sigma}_{ij}\left(\widehat{b}_{N,ij}\widehat{b}_D - \widehat{b}_{D,ij}\widehat{b}_N\right)}{\widehat{b}_D^2} = \frac{\sum_{ij}\widehat{\sigma}_{ij}\left(\widehat{g}_{ij}\widehat{b}_D^2 + \widehat{g}_i\widehat{b}_{D,j}\widehat{b}_D + \widehat{g}_j\widehat{b}_{D,i}\widehat{b}_D\right)}{\widehat{b}_D^2}$$

$$
\begin{aligned}
&= \frac{\sum_{ij}\widehat{\sigma}_{ij}\widehat{g}_{ij}\widehat{b}_D^2 + 2\sum_{ij}\widehat{\sigma}_{ij}\widehat{g}_i\widehat{b}_{D,j}\widehat{b}_D}{\widehat{b}_D^2} \\
&= \sum_{ij}\widehat{\sigma}_{ij}\widehat{g}_{ij} + \frac{2\sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{D,j}\widehat{g}_i}{\widehat{b}_D},
\end{aligned}
$$

from (45) and

$$
\frac{\widehat{b}_{N,q}\widehat{b}_D - \widehat{b}_{D,q}\widehat{b}_N}{\widehat{b}_D^2} = \frac{\left(\widehat{g}_q\widehat{b}_D + \widehat{g}\widehat{b}_{D,q}\right)\widehat{b}_D - \widehat{b}_{D,q}\widehat{g}\widehat{b}_D}{\widehat{b}_D^2} = \widehat{g}_q
$$

by (46). Hence,

$$
\frac{\widehat{b}_N}{\widehat{b}_D}\left(\frac{c_N}{\widehat{b}_N} - \frac{c_D}{\widehat{b}_D}\right) = \frac{1}{2}\sum_{ij}\widehat{\sigma}_{ij}\widehat{g}_{ij} + \frac{\sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{D,j}\widehat{g}_i}{\widehat{b}_D} - \frac{1}{6}\sum_{ijkq}\widehat{h}_{ijk}\mu^4_{ijkq}\widehat{g}_q.
$$

From (47) and (48), we can get

$$
\begin{aligned}
&\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}^N_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{N,\zeta\eta\xi}\widehat{b}_D - \sum_{ijk\zeta\eta\xi}\widehat{h}^D_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{D,\zeta\eta\xi}\widehat{b}_N}{\widehat{b}_D^2} \\
= &\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\left[\widehat{g}_{\zeta\eta\xi}\widehat{b}_D + \widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi} + \widehat{g}_{\zeta\xi}\widehat{b}_{D,\eta} + \widehat{g}_{\zeta}\widehat{b}_{D,\eta\xi} + \widehat{g}_{\eta\xi}\widehat{b}_{D,\zeta} + \widehat{g}_{\eta}\widehat{b}_{D,\zeta\xi} + \widehat{g}_{\xi}\widehat{b}_{D,\zeta\eta}\right]}{\widehat{b}_D} \\
= &\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta\xi} + \frac{3\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi}}{\widehat{b}_D} + \frac{3\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta}\widehat{b}_{D,\eta\xi}}{\widehat{b}_D}, \quad (49)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{b}_{N,\zeta\eta\xi\omega}\widehat{b}_D - \sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{b}_{D,\zeta\eta\xi\omega}\widehat{b}_N}{\widehat{b}_D^2} \\
= &\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta\xi\omega} + \frac{4\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta\xi}\widehat{b}_{D,\omega}}{\widehat{b}_D} + \frac{6\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D} \\
&+ \frac{4\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta}\widehat{b}_{D,\eta\xi\omega}}{\widehat{b}_D} + \frac{4\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta}\widehat{b}_{D,\eta\xi\omega}}{\widehat{b}_D}. \quad (50)
\end{aligned}
$$

We can also show that

$$
\begin{aligned}
& \frac{\widehat{b}_N}{\widehat{b}_D}\left(\frac{d_N}{\widehat{b}_N}-\frac{d_D}{\widehat{b}_D}\right) \\
= \quad & -\frac{1}{120}\frac{\sum_{ijkqrs}\widehat{h}_{ijkqr}\mu^6_{ijkqrs}\left(\widehat{b}_{N,s}\widehat{b}_D-\widehat{b}_{D,s}\widehat{b}_N\right)}{\widehat{b}_D^2} \\
& +\frac{1}{144}\frac{\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\mu^8_{ijkqrstw}\left(\widehat{b}_{N,w}\widehat{b}_D-\widehat{b}_{D,w}\widehat{b}_N\right)}{\widehat{b}_D^2} \\
& -\frac{1}{1296}\frac{\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\mu^{10}_{ijkqrstwv\beta}\left(\widehat{b}_{N,\beta}\widehat{b}_D-\widehat{b}_{D,\beta}\widehat{b}_N\right)}{\widehat{b}_D^2} \\
& -\frac{1}{48}\frac{\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\left(\widehat{b}_{N,rs}\widehat{b}_D-\widehat{b}_{D,rs}\widehat{b}_N\right)}{\widehat{b}_D^2} \\
& +\frac{1}{144}\frac{\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\left(\widehat{b}_{N,tw}\widehat{b}_D-\widehat{b}_{D,tw}\widehat{b}_N\right)}{\widehat{b}_D^2} \\
& -\frac{1}{36}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\left(\widehat{b}_{N,\zeta\eta\xi}\widehat{b}_D-\widehat{b}_{D,\zeta\eta\xi}\widehat{b}_N\right)}{\widehat{b}_D^2} \\
& +\frac{1}{24}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\left(\widehat{b}_{N,\zeta\eta\xi\omega}\widehat{b}_D-\widehat{b}_{D,\zeta\eta\xi\omega}\widehat{b}_N\right)}{\widehat{b}_D^2},
\end{aligned}
$$

since $h^N(\boldsymbol{\theta}) = h^D(\boldsymbol{\theta}) = h(\boldsymbol{\theta})$. Hence, by (45), (46), (49), and (50), we have

$$
\begin{aligned}
& \frac{\widehat{b}_N}{\widehat{b}_D}\left(\frac{d_N}{\widehat{b}_N}-\frac{d_D}{\widehat{b}_D}\right) \\
= \quad & -\frac{1}{120}\sum_{ijkqrs}\widehat{h}_{ijkqr}\mu^6_{ijkqrs}\widehat{g}_s+\frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\mu^8_{ijkqrstw}\widehat{g}_w \\
& -\frac{1}{1296}\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\mu^{10}_{ijkqrstwv\beta}\widehat{g}_\beta-\frac{1}{24}\frac{\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{b}_{D,s}\widehat{g}_r}{\widehat{b}_D} \\
& +\frac{1}{72}\frac{\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{b}_{D,w}\widehat{g}_t}{\widehat{b}_D}-\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_\zeta\widehat{b}_{D,\eta\xi}}{\widehat{b}_D} \\
& +\frac{1}{6}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_\zeta\widehat{b}_{D,\eta\xi\omega}}{\widehat{b}_D}-\frac{1}{48}\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_{rs}
\end{aligned}
$$

$$+\frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{g}_{tw}-\frac{1}{36}\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta\xi}$$
$$+\frac{1}{24}\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta\xi\omega}-\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi}}{\widehat{b}_D}$$
$$+\frac{1}{6}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta\xi}\widehat{b}_{D,\omega}}{\widehat{b}_D}+\frac{1}{4}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}.$$

■

Using the matrix notation for high order derivatives as in Magnus and Neudecker (1999) (with the exception that the first order derivative of a scalar function in our setting is a column vector), we can write Lemma 5.2 in matrix form. Before we do that, let us first introduce the following Generalized Isserlis theorem.

Theorem 5.1 (Generalized Isserlis Theorem) *If $A = \{\alpha_1, \ldots, \alpha_{2N}\}$ is a set of integers such that $1 \le \alpha_i \le P$, for each $i \in [1, 2N]$ and $X \in R^P$ is a zero mean multivariate normal random vector, then*

$$E(X_A) = \sum_A \prod E(X_i X_j), \tag{51}$$

where $EX_A = E\left(\Pi_{i=1}^{2N} X_{\alpha_i}\right) = \mu_{\alpha_1,\ldots,\alpha_{2N}}$ and the notation $\Sigma\Pi$ means summing over all distinct ways of partitioning $X_{\alpha_1}, \ldots, X_{\alpha_{2N}}$ into pairs (X_i, X_j) and each summand is the product of the N pairs. This yields $(2N)!/\left(2^N N!\right) = (2N-1)!!$ terms in the sum where $(2N-1)!!$ is the double factorial defined by $(2N-1)!! = (2N-1) \times (2N-3) \times \cdots \times 1$.

The Isserlis theorem, first obtained by Isserlis (1918), expresses the higher order moments of a zero mean Gaussian vector $X \in R^P$ in terms of its covariance matrix. The generalized Isserlis theorem is due to Withers (1985) and Vignat (2012). For example, if $2N = 4$, then

$$E(X_1X_2X_3X_4) = E(X_1X_2)E(X_3X_4) + E(X_1X_3)E(X_2X_4) + E(X_1X_4)E(X_2X_3)$$

where $\alpha_i = i$ for $i \in [1, 4]$ and there are three terms in the sum. If $2N = 6$, there are fifteen terms in the sum.

Lemma 5.3 *Let $\nabla^j\widehat{h}$, $\nabla^j\widehat{g}$, and $\nabla^j\widehat{b}_D$ be the jth order derivatives of $h(\boldsymbol{\theta})$, $g(\boldsymbol{\theta})$, and $b_D(\boldsymbol{\theta})$ evaluated at $\widehat{\boldsymbol{\theta}}_n$, respectively. If both $(\{h_n\}, g \times b_D)$ and $(\{h_n\}, b_D)$ satisfy the* *analytical assumptions for the stochastic Laplace method on $\wp_{\boldsymbol{\theta}}$, then,*

$$\frac{\int g(\boldsymbol{\theta})\, b_D(\boldsymbol{\theta}) \exp(-nh_n(\boldsymbol{\theta}))\, d\boldsymbol{\theta}}{\int b_D(\boldsymbol{\theta}) \exp(-nh_n(\boldsymbol{\theta}))\, d\boldsymbol{\theta}} = \widehat{g} + \frac{1}{n}B_1 + \frac{1}{n^2}(B_2 - B_3) + O_p\left(\frac{1}{n^3}\right),$$

where

$$\begin{aligned}
B_1 &= \frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] + (\nabla\widehat{g})'\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} - \frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}, \\
B_2 &= B_{21} + B_{22}, \\
B_3 &= B_1 \times B_4,
\end{aligned}$$

with

$$\begin{aligned}
B_{21} = & -\frac{1}{8}(\nabla\widehat{g})'\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^5\widehat{h}\right)' vec\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right] \\
& +\frac{1}{4}\left[vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)' \otimes \left(vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)' \nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& +\frac{1}{6}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes \left(\nabla^2\widehat{h}\right)^{-1} \otimes \left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& +\frac{1}{16}\mathbf{tr}\left[\left(\left(\nabla^2\widehat{h}\right)^{-1} \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right)'\nabla^4\widehat{h}\right] vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& +\frac{1}{4}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1} \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right)\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& -\frac{3}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes \left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& -\frac{1}{4}vec\left(\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes \left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes \left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& -\frac{1}{16}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}' vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& -\frac{1}{24}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes \left(\nabla^2\widehat{h}\right)^{-1} \otimes \left(\nabla^2\widehat{h}\right)^{-1}\right] vec\left(\nabla^3\widehat{h}\right) \\
& \times vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla g \\
& -\frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]\nabla\widehat{b}_D'\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& -\frac{1}{2}\left[vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)' \otimes \left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{b}_D\right)'\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}
\end{aligned}$$

$$+\frac{1}{8}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\frac{\nabla\widehat{b}_D}{\widehat{b}_D}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\right.$$

$$+\frac{1}{12}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\frac{\nabla\widehat{b}_D}{\widehat{b}_D}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}$$

$$+\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\widehat{b}_D vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\right.$$

$$+\frac{1}{4}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\widehat{b}_D vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\right.\right.$$

$$+\frac{1}{2}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\left(\nabla^2\widehat{h}\right)^{-1}\widehat{b}_D\right)'\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}$$

$$-\frac{1}{2}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\right.$$

$$-\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}$$

$$-\frac{1}{4}\mathbf{tr}\left[\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\right.$$

$$+\frac{1}{2}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\left(\nabla^2\widehat{h}\right)^{-1}\frac{\left(\nabla^3\widehat{b}_D\right)'}{\widehat{b}_D}\nabla\widehat{g},\right.$$

$$\begin{aligned}
B_{22} = & -\frac{1}{16}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
& -\frac{1}{4}\mathbf{tr}\left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right] \\
& +\frac{1}{16}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]\right. \\
& +\frac{1}{24}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
& +\frac{1}{4}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)\right. \\
& +\frac{1}{8}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right. \\
& +\frac{1}{4}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right) \\
& -\frac{1}{4}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{g}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right.
\end{aligned}$$

$$-\frac{1}{6}vec\left(\nabla^3\widehat{g}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)$$
$$+\frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\nabla^4\widehat{g}'\right]$$
$$-\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}$$
$$-\frac{1}{4}\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}$$
$$-\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}$$
$$+\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}$$
$$+\frac{1}{4}\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right]$$
$$+\frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right],$$

$$\begin{aligned}B_4 \;=\;& \frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right]-\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\widehat{h}^{(3)}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\\
&+\frac{1}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\\
&+\frac{1}{12}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\\
&-\frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right].\end{aligned}$$

Proof. From (5.2), we first write each term of B_1 in matrix form based on (51), that is

$$\frac{1}{2}\sum_{ij}\widehat{\sigma}_{ij}\widehat{g}_{ij}=\frac{1}{2}\sum_{ij}\widehat{\sigma}_{ij}\widehat{g}_{ij}=\frac{1}{2}tr\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right],$$

$$\frac{\sum_{ij}\widehat{\sigma}_{ij}\widehat{b}_{D,j}\widehat{g}_i}{\widehat{b}_D}=\sum_{ij}\widehat{g}_i\widehat{\sigma}_{ij}\frac{\widehat{b}_{D,j}}{\widehat{b}_D}=\left(\nabla\widehat{g}\right)'\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D},$$

$$-\frac{1}{6}\sum_{ijkq}\widehat{h}_{ijk}\mu^4_{ijkq}\widehat{g}_q = -\frac{1}{2}\sum_{ijkq}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{g}_q = -\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.$$

Then, we can similarly write each term of B_2 in matrix form based on (51), that is

$$\begin{aligned}
-\frac{1}{120}\sum_{ijkqrs}\widehat{h}_{ijkqr}\mu^6_{ijkqrs}\widehat{g}_s &= -\frac{15}{120}\sum_{ijkqrs}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{h}_{ijkqr}\widehat{\sigma}_{rs}\widehat{g}_s \\
&= -\frac{1}{8}\left(\nabla\widehat{g}\right)'\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^5\widehat{h}\right)' vec\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right].
\end{aligned}$$

By (84) in the next subsection, we have

$$\begin{aligned}
&\frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\mu^8_{ijkqrstw}\widehat{g}_w \\
= &\frac{1}{4}\left[vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\otimes\left(vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
&+\frac{1}{6}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
&+\frac{1}{16}\mathbf{tr}\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right)'\nabla^4\widehat{h}\right]vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
&+\frac{1}{4}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right)\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

By (85) in the next subsection, we have

$$\begin{aligned}
&-\frac{1}{1296}\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\mu^{10}_{ijkqrstwv\beta}\widehat{g}_\beta \\
= &-\frac{3}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
&-\frac{1}{4}vec\left(\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
&-\frac{1}{16}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}' vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
&-\frac{1}{24}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)
\end{aligned}$$

$$\times vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla g,$$

$$\begin{aligned}
&-\frac{1}{24}\sum_{ijkqrs}\frac{\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_r\widehat{b}_{D,s}}{\widehat{b}_D}\\
=\ &-\frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]\frac{\nabla\widehat{b}_D}{\widehat{b}_D}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&-\frac{1}{2}\left[vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\otimes\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right)'\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

Similar to the proof of (79) in the next subsection,, we can get

$$\begin{aligned}
&\frac{1}{72}\frac{\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{b}_{D,w}\widehat{g}_t}{\widehat{b}_D}\\
=\ &\frac{1}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\frac{\nabla\widehat{b}_D}{\widehat{b}_D}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{1}{12}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\frac{\nabla\widehat{b}_D}{\widehat{b}_D}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{1}{4}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{1}{2}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right)'\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

By (57), we can show that

$$\begin{aligned}
&-\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{D,\eta\xi}\widehat{g}_\zeta}{\widehat{b}_D}\\
=\ &-\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&-\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}
\end{aligned}$$

$$-\frac{1}{4}\mathbf{tr}\left[\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g},\right.$$

$$\begin{aligned}
\frac{1}{6}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{b}_{D,\eta\xi\omega}\widehat{g}_\zeta}{\widehat{b}_D} &= \frac{3}{6}\sum_{\zeta\eta\xi\omega}\widehat{g}_\zeta\widehat{\sigma}_{\zeta\eta}\widehat{\sigma}_{\xi\omega}\frac{\widehat{b}_{D,\eta\xi\omega}}{\widehat{b}_D} = \frac{1}{2}\sum_{\zeta\eta\xi\omega}\widehat{g}_\zeta\widehat{\sigma}_{\zeta\eta}\frac{\widehat{b}_{D,\eta\xi\omega}}{\widehat{b}_D}\widehat{\sigma}_{\xi\omega} \\
&= \frac{1}{2}\left(\nabla\widehat{g}\right)'\left(\nabla^2\widehat{h}\right)^{-1}\frac{\left(\nabla^3\widehat{b}_D\right)'}{\widehat{b}_D}\left[vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right].
\end{aligned}$$

By (68) in the next subsection, we have

$$\begin{aligned}
& -\frac{1}{48}\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_{rs} \\
= & -\frac{1}{16}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
& -\frac{1}{4}\mathbf{tr}\left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right].
\end{aligned}$$

By (79) in the next subsection, we have

$$\begin{aligned}
& \frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{g}_{tw} \\
= & \frac{1}{16}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
& +\frac{1}{24}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
& +\frac{1}{4}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right) \\
& +\frac{1}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) \\
& +\frac{1}{4}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right).
\end{aligned}$$

By (73) in the next subsection, we have

$$
\begin{aligned}
& -\frac{1}{36}\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^{6}_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta\xi} \\
= & -\frac{1}{4}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{g}\right)' vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) \\
& -\frac{1}{6}vec\left(\nabla^3\widehat{g}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{24}\sum_{\zeta\eta\xi\omega}\mu^{4}_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta\xi\omega} &= \frac{3}{24}\sum_{\zeta\eta\xi\omega}\widehat{\sigma}_{\zeta\eta}\widehat{\sigma}_{\xi\omega}\widehat{g}_{\zeta\eta\xi\omega} \\
&= \frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{g}\right)'\right].
\end{aligned}
$$

By (57), we can get

$$
\begin{aligned}
& -\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^{6}_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi}}{\widehat{b}_D} \\
= & -\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} \\
& -\frac{1}{4}\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} \\
& -\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D},
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{6}\frac{\sum_{\zeta\eta\xi\omega}\mu^{4}_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta\xi}\widehat{b}_{D,\omega}}{\widehat{b}_D} &= \frac{3}{6}\sum_{\zeta\eta\xi\omega}\widehat{\sigma}_{\zeta\eta}\widehat{g}_{\zeta\eta\xi}\widehat{\sigma}_{\xi\omega}\frac{\widehat{b}_{D,\omega}}{\widehat{b}_D} \\
&= \frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}.
\end{aligned}
$$

From (56),

$$
\frac{1}{4}\frac{\sum_{\zeta\eta\xi\omega}\mu^{4}_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}
$$

$$= \frac{1}{4}\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right]+\frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right].$$

And we have

$$\frac{1}{2}\sum_{ij}\widehat{\sigma}_{ij}\frac{\widehat{b}_{D,ij}}{\widehat{b}_D}=\frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right], \tag{52}$$

$$\begin{aligned} -\frac{1}{6}\sum_{ijkq}\widehat{h}_{ijk}\mu^4_{ijkq}\frac{\widehat{b}_{D,q}}{\widehat{b}_D} &= -\frac{3}{6}\sum_{ijkq}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\frac{\widehat{b}_{D,q}}{\widehat{b}_D} \\ &= -\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}. \end{aligned} \tag{53}$$

From (74) in the next subsection, we have

$$\begin{aligned} &\frac{1}{72}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs} \\ =\ &\frac{1}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) \\ &+\frac{1}{12}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right). \end{aligned} \tag{54}$$

Note that

$$\begin{aligned} -\frac{1}{24}\sum_{ijkq}\widehat{h}_{ijkq}\mu^4_{ijkq} &= -\frac{3}{24}\sum_{ijkq}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq} \\ &= -\frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]. \end{aligned} \tag{55}$$

From (52), (53), (54) and (55), we have

$$\begin{aligned} B_4 &= \frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right]-\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\widehat{h}^{(3)}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} \\ &+\frac{1}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) \end{aligned}$$

$$+\frac{1}{12}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)$$
$$-\frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right].$$

■

5.4.3 Lemma 5.3 in Matrix Form

We show how to express each term of the stochastic expansions reported in Lemma 5.3 in matrix form.

For term $\frac{1}{4}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}$

We can get

$$\frac{1}{4}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}$$
$$=\frac{1}{4}\left(\frac{\sum_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\eta}\widehat{\sigma}_{\xi\omega}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}+\frac{\sum_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\xi}\widehat{\sigma}_{\eta\omega}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}+\frac{\sum_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\xi}\widehat{\sigma}_{\eta\omega}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}\right)$$

where

$$\frac{\sum_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\eta}\widehat{\sigma}_{\xi\omega}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}=\sum_{\zeta\eta}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\eta}\sum_{\xi\omega}\widehat{\sigma}_{\xi\omega}\frac{\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}=\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right],$$

$$\frac{\sum_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\xi}\widehat{\sigma}_{\eta\omega}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}=\sum_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\xi}\widehat{\sigma}_{\eta\omega}\frac{\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}=\sum_{\zeta\eta\xi\omega}\widehat{\sigma}_{\xi\zeta}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\eta\omega}\frac{\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}$$
$$=\sum_{\xi\omega}\frac{\widehat{b}_{D,\xi\omega}}{\widehat{b}_D}\sum_{\zeta\eta}\widehat{\sigma}_{\xi\zeta}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\eta\omega}=\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right],$$

$$\frac{\sum_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\xi}\widehat{\sigma}_{\eta\omega}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D} = \frac{\sum_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{\sigma}_{\zeta\xi}\widehat{\sigma}_{\eta\omega}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D} = \mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right].$$

Then, we have

$$\begin{aligned}
&\frac{1}{4}\frac{\sum_{\zeta\eta\xi\omega}\mu^4_{\zeta\eta\xi\omega}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi\omega}}{\widehat{b}_D} \qquad (56)\\
= \; & \frac{1}{4}\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right] + \frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right]\\
= \; & \frac{1}{4}\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\right] + \frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].
\end{aligned}$$

For term $-\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi}}{\widehat{b}_D}$

Note that

$$-\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi}}{\widehat{b}_D} = -\frac{1}{12}\frac{\sum_{ijkqrs}\widehat{h}_{ijk}\mu^6_{ijkqrs}\widehat{g}_{qr}\widehat{b}_{D,s}}{\widehat{b}_D}$$

where

$$\begin{aligned}
\mu^6_{ijkqrs} \; = \; & \widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\sigma_{rs} + \widehat{\sigma}_{iq}\widehat{\sigma}_{kj}\widehat{\sigma}_{rs} + \widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs} + \widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs} + \widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs} + \widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\\
& + \widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr} + \widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr} + \widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr} + \widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks} + \widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\\
& + \widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr} + \widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr} + \widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq} + \widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}.
\end{aligned}$$

Then, we can decompose it into three groups. The first group has six elements without $\widehat{\sigma}_{qr}$ but with $\widehat{\sigma}_{ij}$, $\widehat{\sigma}_{ik}$ or $\widehat{\sigma}_{jk}$, that is

$$\begin{aligned}
&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{kj}\widehat{\sigma}_{rs}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs}\widehat{g}_{qr}\widehat{b}_{D,s}\\
&+ \sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{g}_{qr}\widehat{b}_{D,s}.
\end{aligned}$$

Note that the six elements are the same since

$$\begin{aligned}
\sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{kj}\widehat{\sigma}_{rs}\widehat{g}_{qr}\widehat{b}_{D,s} &= \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs}\widehat{g}_{qr}\widehat{b}_{D,s} = \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{g}_{qr}\widehat{b}_{D,s} \\
&= \sum_{ijkqrs} \widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kq}\widehat{g}_{qr}\widehat{\sigma}_{rs}\widehat{b}_{D,s} \\
&= vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)' \nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{b}_D,
\end{aligned}$$

and

$$\begin{aligned}
\sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{g}_{qr}\widehat{b}_{D,s} &= \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{g}_{qr}\widehat{b}_{D,s} = \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{qr}\widehat{b}_{D,s} \\
&= \sum_{ijkqrs} \widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{qr}\widehat{b}_{D,s},
\end{aligned}$$

where

$$\begin{aligned}
\sum_{ijkqrs} \widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{qr}\widehat{b}_{D,s} &= \sum_{ijkqrs} \widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kr}\widehat{g}_{rq}\widehat{\sigma}_{qs}\widehat{b}_{D,s} = \sum_{ijkqrs} \widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kq}\widehat{g}_{qr}\widehat{\sigma}_{rs}\widehat{b}_{D,s} \\
&= vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)' \nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{b}_D.
\end{aligned}$$

The second group has six elements without $\widehat{\sigma}_{qr}$, $\widehat{\sigma}_{ij}$, $\widehat{\sigma}_{ik}$ and $\widehat{\sigma}_{jk}$, that is

$$\begin{aligned}
&\sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{g}_{qr}\widehat{b}_{D,s} \\
&+ \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{g}_{qr}\widehat{b}_{D,s}.
\end{aligned}$$

These six elements are the same since

$$\begin{aligned}
&\sum_{ijkqrs} \widehat{h}_{ijk}\left[\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks} + \widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\right]\widehat{g}_{qr}\widehat{b}_{D,s} \\
= &\sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s} + \sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s} \\
= &\ 2\sum_{ijkqrs} \widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s} = 2\sum_{ijkqrs} \widehat{\sigma}_{iq}\widehat{g}_{qr}\widehat{\sigma}_{jr}\widehat{h}_{ijk}\widehat{\sigma}_{ks}\widehat{b}_{D,s}
\end{aligned}$$

$$= \; 2vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{b}_D,$$

$$\begin{aligned}
&\sum_{ijkqrs}\widehat{h}_{ijk}\left[\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}+\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\right]\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{g}_{qr}\widehat{b}_{D,s}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\sum_{ijkqrs}\widehat{h}_{ikj}\widehat{\sigma}_{iq}\widehat{\sigma}_{kr}\widehat{\sigma}_{js}\widehat{g}_{qr}\widehat{b}_{D,s}+\sum_{ijkqrs}\widehat{h}_{kji}\widehat{\sigma}_{kr}\widehat{\sigma}_{jq}\widehat{\sigma}_{is}\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\;2vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{b},
\end{aligned}$$

$$\begin{aligned}
&\sum_{ijkqrs}\widehat{h}_{ijk}\left[\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}+\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\right]\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{g}_{qr}\widehat{b}_{D,s}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\sum_{ijkqrs}\widehat{h}_{kij}\widehat{\sigma}_{kq}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{g}_{qr}\widehat{b}_{D,s}+\sum_{ijkqrs}\widehat{h}_{jki}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{is}\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\;2vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{b}.
\end{aligned}$$

The third group has three elements without $\widehat{\sigma}_{qr}$, that is

$$\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{g}_{qr}\widehat{b}_{D,s}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}\widehat{g}_{qr}\widehat{b}_{D,s}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{g}_{qr}\widehat{b}_{D,s}.$$

We have

$$\begin{aligned}
&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{g}_{qr}\widehat{b}_{D,s}\\
=&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}\widehat{g}_{qr}\widehat{b}_{D,s}=\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{g}_{qr}\widehat{b}_{D,s}=\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{jk}\widehat{\sigma}_{is}\widehat{b}_{D,s}\widehat{\sigma}_{qr}\widehat{g}_{qr}
\end{aligned}$$

$$= \sum_{qr} \widehat{\sigma}_{qr}\widehat{g}_{qr} \sum_{ijks} \widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{is}\widehat{b}_{D,s} = \mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right] vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{b}_D.$$

Then, we have

$$\begin{aligned}
&-\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi}}{\widehat{b}_D} \\
= \; & -\frac{1}{2}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} \\
& -\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} \\
& -\frac{1}{4}\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}.
\end{aligned}$$

For the same reason, we have

$$\begin{aligned}
&-\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{b}_{D,\eta\xi}\widehat{g}_{\zeta}}{\widehat{b}_D} \qquad (57)\\
= \; & -\frac{1}{2}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& -\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& -\frac{1}{4}\mathbf{tr}\left[\frac{\nabla^2\widehat{b}_D}{\widehat{b}_D}\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

Note that

$$\begin{aligned}
& vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} \\
= \; & \mathbf{tr}\left[vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right] \\
= \; & \mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right] \\
= \; & \mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].
\end{aligned}$$

We have

$$\begin{aligned}
& \sum_{ijkqrs} \widehat{\sigma}_{iq}\widehat{g}_{qr}\widehat{\sigma}_{jr}\widehat{h}_{ijk}\widehat{\sigma}_{ks}\frac{\widehat{b}_{D,s}}{\widehat{b}_D} \\
= & \sum_{ijkqrs} \widehat{\sigma}_{iq}\widehat{h}_{ijk}\widehat{\sigma}_{ks}\frac{\widehat{b}_{D,s}}{\widehat{b}_D}\widehat{\sigma}_{jr}\widehat{g}_{qr} = \sum_{qr}\left[\sum_{ijks}\widehat{\sigma}_{iq}\widehat{h}_{ijk}\widehat{\sigma}_{ks}\frac{\widehat{b}_{D,s}}{\widehat{b}_D}\widehat{\sigma}_{jr}\right]\widehat{g}_{qr} \\
= & \sum_{qr}\left[\sum_{ijks}\widehat{\sigma}_{qi}\widehat{\sigma}_{ks}\frac{\widehat{b}_{D,s}}{\widehat{b}_D}\widehat{h}_{ikj}\widehat{\sigma}_{jr}\right]\widehat{g}_{qr} = \sum_{qr}\left[\sum_{j}\left[\sum_{ik}\left[\widehat{\sigma}_{qi}\left[\sum_{s}\widehat{\sigma}_{ks}\frac{\widehat{b}_{D,s}}{\widehat{b}_D}\right]\widehat{h}_{ikj}\right]\widehat{\sigma}_{jr}\right]\right]\widehat{g}_{qr} \\
= & \ \mathbf{tr}\left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right)'\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].
\end{aligned}$$

It can also be shown that

$$\begin{aligned}
& vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right) \\
= & \ vec\left(\nabla^2\widehat{g}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} \\
= & \ vec\left(\nabla^2\widehat{g}\right)' vec\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right] \\
= & \ vec\left(\nabla^2\widehat{g}\right)'\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right)'\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right] vec\left(\nabla^3\widehat{h}\right) \\
= & \ \mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right)\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\right] \\
= & \ \mathbf{tr}\left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right)'\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].
\end{aligned}$$

Then, we have

$$\begin{aligned}
& -\frac{1}{12}\frac{\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta}\widehat{b}_{D,\xi}}{\widehat{b}_D} \\
= & -\frac{1}{2}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D} vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
& -\frac{1}{2}\mathbf{tr}\left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right)'\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]
\end{aligned}$$

$$-\frac{1}{4}\mathbf{tr}\left[\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}.\right.$$

For term $-\frac{1}{48}\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_{rs}$

Note that

$$\begin{aligned}\mu^6_{ijkqrs} &= \widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\sigma_{rs}+\widehat{\sigma}_{iq}\widehat{\sigma}_{kj}\widehat{\sigma}_{rs}+\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs}+\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}+\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}+\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\\ &\quad+\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}+\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}+\widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}+\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}+\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\\ &\quad+\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}+\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}+\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}+\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}.\end{aligned}$$

We can decompose $\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_{rs}$ into two groups. The first group has twelve elements without $\widehat{\sigma}_{rs}$, that is,

$$\begin{aligned}&\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{g}_{rs}+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{g}_{rs}\\ &+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{g}_{rs}+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}\widehat{g}_{rs}+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{g}_{rs}\\ &+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{rs}+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{g}_{rs}+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{g}_{rs}\\ &+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{g}_{rs}+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{g}_{rs}+\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{g}_{rs}.\end{aligned}$$

The twelve elements in this group are the same as

$$\begin{aligned}\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs} &= \sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{g}_{rs}\widehat{\sigma}_{sq} \qquad (58)\\ &= \mathbf{tr}\left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right].\end{aligned}$$

The detailed proof is as follows

$$\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{g}_{rs}=\sum_{ijkqsr}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{g}_{sr}=\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}, \qquad (59)$$

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{g}_{rs} = \sum_{ikjqrs} \widehat{h}_{ikjq}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{g}_{rs} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}, \tag{60}$$

$$\begin{aligned}\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}\widehat{g}_{rs} &= \sum_{ijkqsr} \widehat{h}_{ijkq}\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}\widehat{g}_{sr} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{g}_{rs} \\ &= \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}, \end{aligned} \tag{61}$$

where the last equality is because of (60). And

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{rs} = \sum_{iqjkrs} \widehat{h}_{iqjk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{rs} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}, \tag{62}$$

$$\begin{aligned}\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{g}_{rs} &= \sum_{ijkqsr} \widehat{h}_{ijkq}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{g}_{sr} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{g}_{rs} \\ &= \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}, \end{aligned} \tag{63}$$

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{g}_{rs} = \sum_{jkiqrs} \widehat{h}_{jkiq}\widehat{\sigma}_{jk}\widehat{\sigma}_{ir}\widehat{\sigma}_{qs}\widehat{g}_{rs} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}, \tag{64}$$

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{g}_{rs} = \sum_{jqikrs} \widehat{h}_{jqik}\widehat{\sigma}_{jq}\widehat{\sigma}_{ir}\widehat{\sigma}_{ks}\widehat{g}_{rs} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}, \tag{65}$$

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{g}_{rs} = \sum_{kqijrs} \widehat{h}_{kqij}\widehat{\sigma}_{kq}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{g}_{rs} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs}, \tag{66}$$

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{g}_{rs} = \sum_{jkqirs} \widehat{h}_{jkqi}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{is}\widehat{g}_{rs} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs},$$

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{g}_{rs} = \sum_{jqkirs} \widehat{h}_{jqki}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{is}\widehat{g}_{rs} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs},$$

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{g}_{rs} = \sum_{kqjirs} \widehat{h}_{kqji}\widehat{\sigma}_{kq}\widehat{\sigma}_{jr}\widehat{\sigma}_{is}\widehat{g}_{rs} = \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{rs},$$

$$\begin{aligned} & \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{g}_{rs}\widehat{\sigma}_{qs} \\ = & \sum_{ijkq} \widehat{h}_{ijkq} \left[\widehat{\sigma}_{ij} \sum_{rs} \widehat{\sigma}_{kr}\widehat{g}_{rs}\widehat{\sigma}_{qs} \right] \\ = & \ \mathbf{tr} \left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1} \nabla^2\widehat{g} \left(\nabla^2\widehat{h}\right)^{-1} \right) \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1} \right) \right] \left(\nabla^4\widehat{h}\right)' \right]. \end{aligned}$$

We can illustrate the result by a simple example. Let

$$\hat{h}^{(4)} = \begin{bmatrix} h_{1111} & h_{1112} \\ h_{2111} & h_{2112} \\ h_{1211} & h_{1212} \\ h_{2211} & h_{2212} \\ h_{1121} & h_{1122} \\ h_{2121} & h_{2122} \\ h_{1221} & h_{1222} \\ h_{2221} & h_{2222} \end{bmatrix}, \widehat{e} = \left(\nabla^2\widehat{h}\right)^{-1} \nabla^2\widehat{g} \left(\nabla^2\widehat{h}\right)^{-1},$$

$$\widehat{e} = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix}, \left(\nabla^2\widehat{h}\right)^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, vec\left(\left(\nabla^2\widehat{h}\right)^{-1} \right) = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}.$$

Then,

$$
\widehat{e} \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) = \begin{bmatrix} \sigma_{11}e_{11} & \sigma_{11}e_{12} \\ \sigma_{21}e_{11} & \sigma_{21}e_{12} \\ \sigma_{12}e_{11} & \sigma_{12}e_{12} \\ \sigma_{22}e_{11} & \sigma_{22}e_{12} \\ \sigma_{11}e_{11} & \sigma_{11}e_{12} \\ \sigma_{21}e_{11} & \sigma_{21}e_{12} \\ \sigma_{12}e_{11} & \sigma_{12}e_{12} \\ \sigma_{22}e_{11} & \sigma_{22}e_{12} \end{bmatrix}.
$$

The second group has three elements with $\widehat{\sigma}_{rs}$, that is,

$$
\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{g}_{rs} + \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{iq}\widehat{\sigma}_{kj}\widehat{\sigma}_{rs}\widehat{g}_{rs} + \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{iq}\widehat{\sigma}_{kj}\widehat{\sigma}_{rs}\widehat{g}_{rs}.
$$

These three elements are the same. We can get

$$
\begin{aligned}
& \sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{g}_{rs} \\
= \; & \mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].
\end{aligned} \tag{67}
$$

From (58) and (67), we have

$$
\begin{aligned}
& -\frac{1}{48}\sum_{ijkqrs} \widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_{rs} \\
= \; & -\frac{3}{48}\sum_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{h}_{ijkq}\sum_{rs}\widehat{\sigma}_{rs}\widehat{g}_{rs} - \frac{12}{48}\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{g}_{rs}\widehat{\sigma}_{qs} \\
= \; & -\frac{1}{16}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1} \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
& -\frac{1}{4}\mathbf{tr}\left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right) \otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right],
\end{aligned} \tag{68}
$$

$$\sum_{ijkqrs} \widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{g}_{rs}\widehat{\sigma}_{qs}$$

$$= \sum_{ijkqrs} \widehat{\sigma}_{rk}\widehat{\sigma}_{ij}\widehat{h}_{ijkq}\widehat{\sigma}_{qs}\widehat{g}_{rs} = \sum_{rs}\left[\sum_{ijkq} \widehat{\sigma}_{rk}\widehat{\sigma}_{ij}\widehat{h}_{ijkq}\widehat{\sigma}_{qs}\right]\widehat{g}_{rs}$$

$$= \mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].$$

Then, we have

$$-\frac{1}{48}\sum_{ijkqrs} \widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_{rs} \tag{69}$$

$$= -\frac{3}{48}\sum_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{h}_{ijkq}\sum_{rs}\widehat{\sigma}_{rs}\widehat{g}_{rs} - \frac{12}{48}\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{g}_{rs}\widehat{\sigma}_{qs}$$

$$= -\frac{1}{16}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]$$
$$-\frac{1}{4}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]$$

$$= -\frac{1}{16}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right]\nabla^4\widehat{h}\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]$$
$$-\frac{1}{4}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].$$

For term $-\frac{1}{24}\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_r\widehat{b}_{D,s}$

Similar to the proof of (68), we have

$$-\frac{1}{24}\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^6_{ijkqrs}\widehat{g}_r\widehat{b}_{D,s} \tag{70}$$

$$= -\frac{3}{24}\sum_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{h}_{ijkq}\sum_{rs}\widehat{\sigma}_{rs}\widehat{g}_r\widehat{b}_{D,s} - \frac{12}{24}\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_r\widehat{b}_{D,s}$$

$$= -\frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right]\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\nabla\widehat{b}'_D\right]$$
$$-\frac{1}{2}\mathbf{tr}\left[\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\nabla\widehat{b}'_D\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\left(\nabla^4\widehat{h}\right)'\right].$$

We can also write (70) as

$$
\begin{aligned}
& -\frac{1}{24}\sum_{ijkqrs}\widehat{h}_{ijkq}\mu^{6}_{ijkqrs}\widehat{g}_{r}\widehat{b}_{D,s} \\
= & -\frac{3}{24}\sum_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{h}_{ijkq}\sum_{rs}\widehat{\sigma}_{rs}\widehat{g}_{r}\widehat{b}_{D,s}-\frac{12}{24}\sum_{ijkqrs}\widehat{h}_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{sq}\widehat{g}_{r}\widehat{b}_{D,s} \\
= & -\frac{3}{24}\sum_{ijkq}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{h}_{ijkq}\sum_{rs}\widehat{\sigma}_{rs}\widehat{g}_{r}\widehat{b}_{D,s}-\frac{12}{24}\sum_{ijkqrs}\widehat{\sigma}_{ij}\widehat{\sigma}_{sq}\widehat{b}_{D,s}\widehat{h}_{ijqk}\widehat{\sigma}_{kr}\widehat{g}_{r} \\
= & -\frac{1}{8}\mathbf{tr}\left[\left[\left(\nabla^{2}\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^{2}\widehat{h}\right)^{-1}\right)\right]\left(\nabla^{4}\widehat{h}\right)'\right]\nabla\widehat{b}_{D}'\left(\nabla^{2}\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& -\frac{1}{2}\left[vec\left(\left(\nabla^{2}\widehat{h}\right)^{-1}\right)'\otimes\left(\left(\nabla^{2}\widehat{h}\right)^{-1}\nabla\widehat{b}_{D}\right)'\right]\nabla^{4}\widehat{h}\left(\nabla^{2}\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}
$$

For term $-\frac{1}{36}\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^{6}_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta\xi}$

Note that $\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^{6}_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta\xi}=\sum_{ijkqrs}\widehat{h}_{ijk}\mu^{6}_{ijkqrs}\widehat{g}_{qrs}$. We can decompose $\sum_{ijkqrs}\widehat{h}_{ijk}\mu^{6}_{ijkqrs}\widehat{g}_{qrs}$ into two groups. The first group consists of nine elements, each of which has the term from $(\widehat{\sigma}_{ij},\widehat{\sigma}_{ik},\widehat{\sigma}_{jk})$ and the term from $(\widehat{\sigma}_{qr},\widehat{\sigma}_{rs},\widehat{\sigma}_{qs})$, that is

$$
\begin{aligned}
& \sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{kj}\widehat{\sigma}_{rs}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs}\widehat{g}_{qrs} \\
& +\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{g}_{qrs} \\
& +\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{g}_{qrs}.
\end{aligned}
$$

These nine elements are the same and we have

$$
\begin{aligned}
\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{g}_{qrs} &= \sum_{ijkqrs}\widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kq}\widehat{g}_{qrs}\widehat{\sigma}_{rs} \\
&= vec\left(\left(\nabla^{2}\widehat{h}\right)^{-1}\right)'\nabla^{3}\widehat{h}\left(\nabla^{2}\widehat{h}\right)^{-1}\left(\nabla^{3}\widehat{g}\right)'vec\left(\left(\nabla^{2}\widehat{h}\right)^{-1}\right).
\end{aligned} \tag{71}
$$

The second group consists of six elements that do not include any term from $(\widehat{\sigma}_{ij},\widehat{\sigma}_{ik},\widehat{\sigma}_{jk})$, that is

$$
\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{qs}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{\sigma}_{qs}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{qrs}
$$

$$+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{\sigma}_{qs}\widehat{g}_{qrs}+\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{qs}\widehat{g}_{qrs}.$$

These six elements are the same and we have

$$\begin{aligned}&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{qs}\widehat{g}_{qrs} \\ =\ & vec\left(\nabla^3\widehat{g}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right).\end{aligned} \tag{72}$$

From (71) and (72), we can get

$$\begin{aligned}&-\frac{1}{36}\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\mu^6_{ijk\zeta\eta\xi}\widehat{g}_{\zeta\eta\xi} \\ =\ & -\frac{9}{36}\sum_{ijk\zeta\eta\xi}\widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{k\zeta}\widehat{g}_{\zeta\eta\xi}\widehat{\sigma}_{\eta\xi}-\frac{6}{36}\sum_{ijk\zeta\eta\xi}\widehat{h}_{ijk}\widehat{\sigma}_{i\zeta}\widehat{\sigma}_{j\eta}\widehat{\sigma}_{k\xi}\widehat{g}_{\zeta\eta\xi} \\ =\ & -\frac{1}{4}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{g}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) \\ & -\frac{1}{6}vec\left(\nabla^3\widehat{g}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right).\end{aligned} \tag{73}$$

For term $\frac{1}{72}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs}$

Similar to the proof of (73), we have

$$\frac{1}{72}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs}=\frac{9}{72}\sum_{ijkqrs}\widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kq}\widehat{h}_{qrs}\widehat{\sigma}_{rs}+\frac{6}{72}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{h}_{qrs},$$

where

$$\sum_{ijkqrs}\widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kq}\widehat{h}_{qrs}\widehat{\sigma}_{rs}=vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right),$$

$$\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{h}_{qrs} \quad = \quad \sum_{ijk}\widehat{h}_{ijk}\sum_{qrs}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{h}_{qrs}$$

$$= vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right).$$

Then, we can get

$$\begin{aligned}
&\frac{1}{72}\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs} \qquad (74)\\
=\ &\frac{1}{8}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right.\\
&+\frac{1}{12}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right).
\end{aligned}$$

For term $\frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{g}_{tw}$

We can decompose $\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{g}_{tw}$ into seven groups. One group contains all the terms that involve $\widehat{\sigma}_{tw}$, that is,

$$\begin{aligned}
&\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs}\widehat{\sigma}_{tw}\widehat{g}_{tw} \qquad (75)\\
=\ &\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^6_{ijkqrs}\sum_{tw}\widehat{\sigma}_{tw}\widehat{g}_{tw}\\
=\ &9vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]\right.\\
&+6vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]
\end{aligned}$$

by (68). Note that there are fifteen terms in this group.

The other six groups are the same. One of them, which involves $\widehat{\sigma}_{sw}$, is

$$\begin{aligned}
&\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ij}\widehat{\sigma}_{kt}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ik}\widehat{\sigma}_{jq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ik}\widehat{\sigma}_{jt}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{g}_{tw}
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{iq}\widehat{\sigma}_{jt}\widehat{\sigma}_{kr}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jt}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{g}_{tw}.
\end{aligned}
$$

Note that this group can be further decomposed into three sub-groups. The first sub-group consists of six elements, which involve either $\widehat{\sigma}_{rt}$ or $\widehat{\sigma}_{qt}$, that is

$$
\begin{aligned}
&\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ik}\widehat{\sigma}_{jq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
&+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw}=\sum_{ijkqrstw}\widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{ir}\widehat{h}_{rqs}\widehat{\sigma}_{qt}\widehat{g}_{tw}\widehat{\sigma}_{sw} \qquad (76)\\
&= vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right).
\end{aligned}
$$

The second sub-group consists of three elements, which involve one term from ($\widehat{\sigma}_{kt}$, $\widehat{\sigma}_{rt}$, $\widehat{\sigma}_{qt}$) and one term from ($\widehat{\sigma}_{ij}$, $\widehat{\sigma}_{ik}$, $\widehat{\sigma}_{jk}$), that is

$$
\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ij}\widehat{\sigma}_{kt}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ik}\widehat{\sigma}_{jt}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{g}_{tw}+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{g}_{tw}.
$$

These three elements are the same and we have

$$
\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{g}_{tw} \qquad (77)
$$

$$
\begin{aligned}
&= \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{it}\widehat{g}_{tw}\widehat{\sigma}_{sw} = \sum_{ijkqrstw} \widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{it}\widehat{g}_{tw}\widehat{\sigma}_{sw}\widehat{h}_{sqr}\widehat{\sigma}_{qr} \\
&= vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)' \nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}' vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right).
\end{aligned}
$$

The third sub-group consists of six elements, which involve one term from $(\widehat{\sigma}_{ij}, \widehat{\sigma}_{ik}, \widehat{\sigma}_{jk})$ but do not include any term from $(\widehat{\sigma}_{kt}, \widehat{\sigma}_{rt}, \widehat{\sigma}_{qt})$. Thus,

$$
\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{g}_{tw} + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{iq}\widehat{\sigma}_{jt}\widehat{\sigma}_{kr}\widehat{\sigma}_{sw}\widehat{g}_{tw} \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{g}_{tw} + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jt}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{g}_{tw} \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{sw}\widehat{g}_{tw} + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{g}_{tw}.
\end{aligned}
$$

These six elements are the same and we have

$$
\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{g}_{tw} = \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{it}\widehat{g}_{tw}\widehat{\sigma}_{sw} \qquad (78) \\
&= \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{\sigma}_{it}\widehat{g}_{tw}\widehat{\sigma}_{ws}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{h}_{srq} \\
&= vec\left(\nabla^3\widehat{h}\right)'\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right] vec\left(\nabla^3\widehat{h}\right).
\end{aligned}
$$

Then, from (75), (76), (77), and (78), we have

$$
\begin{aligned}
&\frac{1}{144}\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{g}_{tw} \qquad (79) \\
&= \frac{9}{144} vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)' vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
&+\frac{6}{144} vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right] vec\left(\nabla^3\widehat{h}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right] \\
&+\frac{36}{144} vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}' vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right) \\
&+\frac{18}{144} vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}' vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) \\
&+\frac{36}{144} vec\left(\nabla^3\widehat{h}\right)'\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right] vec\left(\nabla^3\widehat{h}\right).
\end{aligned}
$$

We can write (76) as

$$
\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw} \qquad (80)\\
=& \sum_{ijkqrstw} \widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{ir}\widehat{h}_{rqs}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{g}_{tw} = \sum_{ijkqrstw} \widehat{\sigma}_{tq}\widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{ir}\widehat{h}_{qrs}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
=& \sum_{tw}\left[\sum_{ijkqrs} \widehat{\sigma}_{tq}\widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{ir}\widehat{h}_{qrs}\widehat{\sigma}_{sw}\right]\widehat{g}_{tw} = \sum_{tw}\left[\sum_{ijkqrs} \widehat{\sigma}_{tq}\widehat{h}_{ijk}\widehat{\sigma}_{jk}\widehat{\sigma}_{ir}\widehat{h}_{qrs}\widehat{\sigma}_{sw}\right]\widehat{g}_{tw}\\
=& \mathbf{tr}\left[\left[\left(vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right],
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{g}_{tw} \qquad (81)\\
=& \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{it}\widehat{g}_{tw}\widehat{\sigma}_{sw} = \sum_{ijkqrstw} \widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{it}\widehat{g}_{tw}\widehat{\sigma}_{sw}\widehat{h}_{sqr}\widehat{\sigma}_{qr}\\
=& \sum_{ijkqrstw} \widehat{\sigma}_{ti}\widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{qr}\widehat{h}_{qrs}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
=& \mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^2\widehat{g}\right],
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{g}_{tw} = \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{it}\widehat{g}_{tw}\widehat{\sigma}_{sw} \qquad (82)\\
=& \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{\sigma}_{it}\widehat{g}_{tw}\widehat{\sigma}_{ws}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{h}_{srq} = \sum_{ijkqrstw} \widehat{\sigma}_{ti}\widehat{h}_{ikj}\widehat{\sigma}_{kq}\widehat{\sigma}_{jr}\widehat{h}_{qrs}\widehat{\sigma}_{sw}\widehat{g}_{tw}\\
=& \mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\frac{1}{144}\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{g}_{tw} \qquad (83)\\
=& \frac{9}{144}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]
\end{aligned}
$$

$$+\frac{6}{144}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]$$
$$+\frac{36}{144}\mathbf{tr}\left[\left[\left(vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right]$$
$$+\frac{18}{144}\mathbf{tr}\left[\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^2\widehat{g}\right]$$
$$+\frac{36}{144}\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla^2\widehat{g}\right].$$

For term $\frac{1}{72}\frac{\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{b}_{D,w}\widehat{g}_t}{\widehat{b}_D}$ Note that

$$\begin{aligned}&\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{b}_{D,w}\widehat{g}_t\\ =&\sum_{ijkqrstw}\widehat{\sigma}_{sw}\widehat{b}_{D,w}\widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{ir}\widehat{h}_{srq}\widehat{\sigma}_{qt}\widehat{g}_t\\ =&\ vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\widehat{b}_D vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g},\end{aligned}$$

$$\begin{aligned}&\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{b}_{D,w}\widehat{g}_t\\ =&\sum_{ijkqrstw}\widehat{\sigma}_{qr}\widehat{h}_{qrs}\widehat{b}_{D,w}\widehat{\sigma}_{sw}\widehat{\sigma}_{jk}\widehat{h}_{jki}\widehat{\sigma}_{it}\widehat{g}_t\\ =&\ vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\widehat{b}_D vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g},\end{aligned}$$

$$\begin{aligned}&\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{it}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{b}_{D,w}\widehat{g}_t\\ =&\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{it}\widehat{b}_{D,w}\widehat{g}_t\widehat{\sigma}_{sw}=\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{\sigma}_{it}\widehat{b}_{D,w}\widehat{g}_t\widehat{\sigma}_{sw}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{h}_{srq}\\ =&\sum_{ijkqrstw}\widehat{\sigma}_{jr}\widehat{\sigma}_{sw}\widehat{b}_{D,w}\widehat{h}_{srq}\widehat{\sigma}_{qk}\widehat{h}_{jki}\widehat{\sigma}_{it}\widehat{g}_t\\ =&\ vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\left(\nabla^2\widehat{h}\right)^{-1}\widehat{b}_D\right)'\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.\end{aligned}$$

$$\frac{1}{72}\frac{\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrs}\mu^8_{ijkqrstw}\widehat{b}_{D,w}\widehat{g}_t}{\widehat{b}_D}$$

$$\begin{aligned}
= & \frac{1}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\nabla\widehat{g}'\right] \\
& +\frac{1}{12}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\mathbf{tr}\left[\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\nabla\widehat{g}'\right] \\
& +\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\nabla\widehat{g}'\left(\nabla^2\widehat{h}\right)^{-1}\right) \\
& +\frac{1}{4}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\nabla\widehat{g}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right) \\
& +\frac{1}{2}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\nabla\widehat{g}'\left(\nabla^2\widehat{h}\right)^{-1}\right)\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right) \\
= & \frac{1}{8}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\left(\nabla^3\widehat{h}\right)'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\frac{\nabla\widehat{b}_D}{\widehat{b}_D}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& +\frac{1}{12}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\frac{\nabla\widehat{b}_D}{\widehat{b}_D}'\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& +\frac{1}{2}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& +\frac{1}{4}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g} \\
& +\frac{1}{2}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\left(\nabla^2\widehat{h}\right)^{-1}\frac{\nabla\widehat{b}_D}{\widehat{b}_D}\right)'\right)\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

For term $\frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\mu^8_{ijkqrstw}\widehat{g}_w$

In this case, we can decompose $\mu^8_{ijkqrstw}$ into seven groups, each containing fifteen terms. The first four groups contain $\widehat{\sigma}_{qw}$, $\widehat{\sigma}_{rw}$, $\widehat{\sigma}_{sw}$, $\widehat{\sigma}_{tw}$. Out of these four groups, the group involving $\widehat{\sigma}_{tw}$ is

$$\begin{aligned}
& \sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{g}_w+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{g}_w \\
& +\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{g}_w+\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ik}\widehat{\sigma}_{jq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{g}_w
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{tw}\widehat{g}_w.
\end{aligned}
$$

The above fifteen terms can be further decomposed into two different sub-groups by whether two of (i, j, k) are in the same combination of subscripts (i.e., whether there involves $\widehat{\sigma}_{ij}$ or $\widehat{\sigma}_{jk}$ or $\widehat{\sigma}_{ik}$). One sub-group is

$$
\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ik}\widehat{\sigma}_{jq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ik}\widehat{\sigma}_{jr}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ik}\widehat{\sigma}_{js}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{is}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{g}_w.
\end{aligned}
$$

Note that each term in this sub-group are the same and we have

$$
\begin{aligned}
& \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{g}_w \\
= & \sum_{ijkqrstw} \widehat{\sigma}_{rs}\widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kq}\widehat{h}_{rsqt}\widehat{\sigma}_{tw}\widehat{g}_w \\
= & \left[vec\left(\left(\nabla^2\widehat{h}\right)^{-1} \right)' \otimes \left(vec\left(\left(\nabla^2\widehat{h}\right)^{-1} \right)' \nabla^3\widehat{h} \left(\nabla^2\widehat{h}\right)^{-1} \right) \right] \nabla^4\widehat{h} \left(\nabla^2\widehat{h}\right)^{-1} \nabla\widehat{g}.
\end{aligned}
$$

Another sub-group is

$$\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{ir}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{\sigma}_{tw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{is}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{tw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{is}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{tw}\widehat{g}_w.
\end{aligned}$$

Note that each term in this subgroup are the same and we have

$$\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{g}_w \\
= &\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{h}_{qrst}\widehat{\sigma}_{tw}\widehat{g}_w \\
= &\ vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

The second three groups involve $\widehat{\sigma}_{iw}$, $\widehat{\sigma}_{jw}$, $\widehat{\sigma}_{kw}$ in each term. Out of these three groups, the group involving $\widehat{\sigma}_{iw}$ is

$$\begin{aligned}
&\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{st}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{\sigma}_{rt}\widehat{\sigma}_{iw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{rs}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{st}\widehat{\sigma}_{iw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{\sigma}_{rt}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jq}\widehat{\sigma}_{kt}\widehat{\sigma}_{rs}\widehat{\sigma}_{iw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{st}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{qt}\widehat{\sigma}_{iw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jr}\widehat{\sigma}_{kt}\widehat{\sigma}_{qs}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{\sigma}_{rt}\widehat{\sigma}_{iw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{\sigma}_{qt}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{js}\widehat{\sigma}_{kt}\widehat{\sigma}_{qr}\widehat{\sigma}_{iw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jt}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jt}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{\sigma}_{iw}\widehat{g}_w \\
&+ \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jt}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{\sigma}_{iw}\widehat{g}_w.
\end{aligned}$$

This group can be further decomposed into two sub-groups. The first sub-group is as follows

$$\sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{st}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{\sigma}_{rt}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{rs}\widehat{\sigma}_{iw}\widehat{g}_w.$$

Note that each element in this group contains $\widehat{\sigma}_{jk}$. The three terms in the first sub-group are the same and we have

$$\begin{aligned}
& \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jk}\widehat{\sigma}_{qr}\widehat{\sigma}_{st}\widehat{\sigma}_{iw}\widehat{g}_w \\
= & \sum_{ijkqrstw} \widehat{g}_w\widehat{\sigma}_{iw}\widehat{\sigma}_{jk}\widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{qr}\widehat{\sigma}_{st} = \sum_{ijkw} \widehat{g}_w\widehat{\sigma}_{iw}\widehat{\sigma}_{jk}\widehat{h}_{ijk} \sum_{qrst} \widehat{h}_{qrst}\widehat{\sigma}_{qr}\widehat{\sigma}_{st} \\
= & \, vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)' \nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\,\mathbf{tr}\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right)'\nabla^4\widehat{h}\right].
\end{aligned}$$

The second sub-group consists of the remaining terms

$$\begin{aligned}
& \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{st}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{\sigma}_{rt}\widehat{\sigma}_{iw}\widehat{g}_w \\
& + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jq}\widehat{\sigma}_{kt}\widehat{\sigma}_{rs}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jr}\widehat{\sigma}_{kq}\widehat{\sigma}_{st}\widehat{\sigma}_{iw}\widehat{g}_w \\
& + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{qt}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jr}\widehat{\sigma}_{kt}\widehat{\sigma}_{qs}\widehat{\sigma}_{iw}\widehat{g}_w \\
& + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{\sigma}_{rt}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{\sigma}_{qt}\widehat{\sigma}_{iw}\widehat{g}_w \\
& + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{js}\widehat{\sigma}_{kt}\widehat{\sigma}_{qr}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jt}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{\sigma}_{iw}\widehat{g}_w \\
& + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jt}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{\sigma}_{iw}\widehat{g}_w + \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jt}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{\sigma}_{iw}\widehat{g}_w.
\end{aligned}$$

The twelve terms in the second sub-group are the same and we have

$$\begin{aligned}
& \sum_{ijkqrstw} \widehat{h}_{ijk}\widehat{h}_{qrst}\widehat{\sigma}_{jq}\widehat{\sigma}_{kr}\widehat{\sigma}_{st}\widehat{\sigma}_{iw}\widehat{g}_w = \sum_{ijkqrstw} \widehat{\sigma}_{jq}\widehat{\sigma}_{st}\widehat{h}_{qstr}\widehat{\sigma}_{rk}\widehat{h}_{jki}\widehat{\sigma}_{iw}\widehat{g}_w \\
= & \, vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right)\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

Then, we have

$$
\begin{aligned}
&\frac{1}{144}\sum_{ijkqrstw}\widehat{h}_{ijk}\widehat{h}_{qrst}\mu^{8}_{ijkqrstw}\widehat{g}_{w} \qquad (84)\\
=\;&\frac{36}{144}\left[vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\otimes\left(vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{24}{144}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{9}{144}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\,\mathbf{tr}\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right)'\nabla^4\widehat{h}\right]\\
&+\frac{36}{144}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right)\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
=\;&\frac{1}{4}\left[vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\otimes\left(vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{1}{6}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{1}{16}\mathbf{tr}\left[\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\right)'\nabla^4\widehat{h}\right]vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
&+\frac{1}{4}vec\left(\left(\left(\nabla^2\widehat{h}\right)^{-1}\otimes vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\right)\nabla^4\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}
$$

For term $-\frac{1}{1296}\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\mu^{10}_{ijkqrstwv\beta}\widehat{g}_{\beta}$

Note that $\mu^{10}_{ijkqrstwv\beta}$ can be decomposed into nine groups that are the same. Each group has one hundred and five elements that involve $\widehat{\sigma}_{i\beta},\widehat{\sigma}_{j\beta},\widehat{\sigma}_{k\beta},\widehat{\sigma}_{q\beta},\widehat{\sigma}_{r\beta},\widehat{\sigma}_{s\beta},\widehat{\sigma}_{t\beta},\widehat{\sigma}_{w\beta}$ and $\widehat{\sigma}_{v\beta}$. We take the group involving $\widehat{\sigma}_{v\beta}$ as an example. This group is further decomposed into seven groups with fifteen elements in each group. There are six groups out of seven with the following structure:

$$
\begin{aligned}
&\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kt}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{jq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{kr}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{kt}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jt}\widehat{\sigma}_{kr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{jt}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{kj}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{kq}\widehat{\sigma}_{jr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{kr}\widehat{\sigma}_{jq}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}.
\end{aligned}
$$

The elements in these group do not include $\widehat{\sigma}_{tw}$. We can further decompose the above group into two sub-groups. The first sub-group has elements involving $\widehat{\sigma}_{ij}$, $\widehat{\sigma}_{ik}$ or $\widehat{\sigma}_{jk}$ and is expressed as

$$
\begin{aligned}
&\sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kt}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{jq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{kr}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{kt}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{kj}\widehat{\sigma}_{qr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}.
\end{aligned}
$$

The nine terms in the first sub-group are the same and we have

$$
\begin{aligned}
&\sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
= &\sum_{ijkqrstwv\beta} \widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kq}\widehat{h}_{qrs}\widehat{\sigma}_{rt}\widehat{\sigma}_{sw}\widehat{h}_{twv}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
= &\ vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)' \nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}
$$

The second sub-group is

$$
\begin{aligned}
&\sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jt}\widehat{\sigma}_{kr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} \\
&+ \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta} + \sum_{ijkqrstwv\beta} \widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{jt}\widehat{\sigma}_{kq}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}
\end{aligned}
$$

$$+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{kq}\widehat{\sigma}_{jr}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{kr}\widehat{\sigma}_{jq}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}$$

The six terms in the second sub-group are the same and we have

$$\begin{aligned}
&\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
=&\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{h}_{qrs}\widehat{\sigma}_{kt}\widehat{\sigma}_{sw}\widehat{h}_{twv}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
=&\ vec\left(\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

Note that out of the seven groups, we have one group remained, which is expressed as

$$\begin{aligned}
&\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{jq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{is}\widehat{\sigma}_{kj}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{is}\widehat{\sigma}_{kq}\widehat{\sigma}_{jr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{is}\widehat{\sigma}_{kr}\widehat{\sigma}_{jq}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}.
\end{aligned}$$

We can further decompose the above group into two sub-groups. The first sub-group is

$$\begin{aligned}
&\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{jq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
&+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{kr}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ik}\widehat{\sigma}_{ks}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}
\end{aligned}$$

$$+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jk}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ir}\widehat{\sigma}_{jk}\widehat{\sigma}_{qs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}$$
$$+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{is}\widehat{\sigma}_{kj}\widehat{\sigma}_{qr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}.$$

The elements of this sub-group involve $\widehat{\sigma}_{ij}$, $\widehat{\sigma}_{ik}$ or $\widehat{\sigma}_{jk}$. The nine terms in the first sub-group are the same and we have

$$\begin{aligned}
&\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ij}\widehat{\sigma}_{kq}\widehat{\sigma}_{rs}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
=\;&\sum_{ijkqrs}\widehat{\sigma}_{ij}\widehat{h}_{ijk}\widehat{\sigma}_{kq}\widehat{h}_{qrs}\widehat{\sigma}_{rs}\sum_{twv\beta}\widehat{\sigma}_{tw}\widehat{h}_{twv}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
=\;&vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\times vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

The second sub-group is

$$\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{js}\widehat{\sigma}_{kr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}$$
$$+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{ir}\widehat{\sigma}_{jq}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{it}\widehat{\sigma}_{js}\widehat{\sigma}_{kq}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}$$
$$+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{is}\widehat{\sigma}_{kq}\widehat{\sigma}_{jr}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}+\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{is}\widehat{\sigma}_{kr}\widehat{\sigma}_{jq}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}.$$

The six terms in the second sub-group are the same and we have

$$\begin{aligned}
&\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{\sigma}_{tw}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
=\;&\sum_{ijkqrs}\widehat{h}_{ijk}\widehat{\sigma}_{iq}\widehat{\sigma}_{jr}\widehat{\sigma}_{ks}\widehat{h}_{qrs}\sum_{twv\beta}\widehat{\sigma}_{tw}\widehat{h}_{twv}\widehat{\sigma}_{v\beta}\widehat{g}_{\beta}\\
=\;&vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\\
&\times vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}.
\end{aligned}$$

Then, we have

$$
\begin{aligned}
& -\frac{1}{1296}\sum_{ijkqrstwv\beta}\widehat{h}_{ijk}\widehat{h}_{qrs}\widehat{h}_{twv}\mu^{10}_{ijkqrstwv\beta}\widehat{g}_{\beta} \qquad (85)\\
= & -\frac{9\times 6\times 9}{1296}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'\\
& \times\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
& -\frac{9\times 6\times 6}{1296}vec\left(\nabla^3\widehat{h}'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\right)'\\
& \times\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
& -\frac{9\times 1\times 9}{1296}vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla^3\widehat{h}'vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)\\
& \times vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla\widehat{g}\\
& -\frac{9\times 1\times 6}{1296}vec\left(\nabla^3\widehat{h}\right)'\left[\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\otimes\left(\nabla^2\widehat{h}\right)^{-1}\right]vec\left(\nabla^3\widehat{h}\right)\\
& \times vec\left(\left(\nabla^2\widehat{h}\right)^{-1}\right)'\nabla^3\widehat{h}\left(\nabla^2\widehat{h}\right)^{-1}\nabla g
\end{aligned}
$$

Note that

$$
\frac{9\times 6\times 9}{1296}=\frac{3}{8},\frac{9\times 6\times 6}{1296}=\frac{1}{4},\frac{9\times 1\times 9}{1296}=\frac{1}{16},\frac{9\times 1\times 6}{1296}=\frac{1}{24}.
$$

5.4.4 Proof of Lemma 2.2

Proof of Lemma 2.2 can be obtained by directly applying Lemma 5.3 by setting $b_D(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$, $g(\boldsymbol{\theta}) = l_t(\boldsymbol{\theta})$, $-nh^N(\boldsymbol{\theta}) = -nh^D(\boldsymbol{\theta}) = \ln p(\mathbf{y}|\boldsymbol{\theta})$. Then, under Assumptions 1-10, we have

$$
\begin{aligned}
& \frac{\int l_t(\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, p(\mathbf{y}|\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int p(\boldsymbol{\theta})\, p(\mathbf{y}|\boldsymbol{\theta})\, d\boldsymbol{\theta}}\\
= & \; l_t\left(\widehat{\boldsymbol{\theta}}_n\right)+\frac{1}{n}B_{t,1}+\frac{1}{n^2}\left(B^1_{t,21}+B^2_{t,21}+B_{t,22}-B_4B_{t,1}\right)+O_p\left(n^{-3}\right),
\end{aligned}
$$

where the expressions for are given in Appendix 5.2.

5.5 Proof of Lemma 2.3

By definition, we have

$$
\begin{aligned}
P_D &= \int -2\left[\ln p\left(\mathbf{y}|\boldsymbol{\theta}\right)-\ln p\left(\mathbf{y}|\bar{\boldsymbol{\theta}}\right)\right]p\left(\boldsymbol{\theta}|\mathbf{y}\right)d\boldsymbol{\theta}\\
&= -2\int \ln p\left(\mathbf{y}|\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}|\mathbf{y}\right)d\boldsymbol{\theta}+2\ln p\left(\mathbf{y}|\bar{\boldsymbol{\theta}}\right)=-2\sum_{t=1}^{n}\int l_t\left(\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}|\mathbf{y}\right)d\boldsymbol{\theta}+2\ln p\left(\mathbf{y}|\bar{\boldsymbol{\theta}}\right).
\end{aligned}
$$

By Lemma 2.2, we have

$$
\int l_t\left(\boldsymbol{\theta}\right)p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}=l_t\left(\widehat{\boldsymbol{\theta}}_n\right)+\frac{1}{n}B_{t,1}+\frac{1}{n^2}\left(B_{t,2}-B_{t,3}\right)+O_p\left(\frac{1}{n^3}\right).
$$

Note that

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}B_{t,1} &= -\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right]-\frac{1}{n}\sum_{t=1}^{n}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)'\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}\\
&\quad+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\\
&= -\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)\right]=-\frac{1}{2}P,
\end{aligned}
$$

since $\sum_{t=1}^{n}\nabla l_t\left(\widehat{\boldsymbol{\theta}}_n\right)=0$. For the same reason,

$$
\frac{1}{n}\sum_{t=1}^{n}B_{t,21}=0.
$$

Moreover,

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}B_{t,22}\\
&= -\frac{1}{16}\mathbf{tr}\left[A_2\right]\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right]\\
&\quad-\frac{1}{4}\mathbf{tr}\left[\left[\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\right]\bar{\mathbf{H}}_n^{(4)}\left(\widehat{\boldsymbol{\theta}}_n\right)'\right]
\end{aligned}
$$

$$+\frac{1}{16}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)$$
$$\times\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right]$$
$$+\frac{1}{24}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)$$
$$\times\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right]$$
$$+\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)'$$
$$\times vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)$$
$$+\frac{1}{8}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}$$
$$\times\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)'vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)$$
$$+\frac{1}{4}vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)'\left[\begin{array}{c}\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\\ \otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\end{array}\right]$$
$$\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)$$
$$-\frac{1}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)'vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)$$
$$-\frac{1}{6}vec\left(\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right)\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right]$$
$$\times vec\left(\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\right)$$
$$+\frac{1}{8}\mathbf{tr}\left[\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\otimes vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)\right]\frac{1}{n}\sum_{t=1}^{n}\nabla^4 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)'\right]$$
$$-\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}$$
$$-\frac{1}{4}\mathbf{tr}\left[\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right]vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}$$
$$-\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}$$

$$+\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown\widehat{p}}{\widehat{p}}$$
$$+\frac{1}{4}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\right]\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown^2\widehat{p}}{\widehat{p}}\right]$$
$$+\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\nabla^2 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown^2\widehat{p}}{\widehat{p}}\right].$$

Then, we have

$$\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n}B_{t,22} \\
= \;& -\frac{P}{16}\mathbf{tr}\left[A_2\right]-\frac{1}{4}\mathbf{tr}\left[A_2\right]+\frac{P}{16}A_1+\frac{P}{24}A_3+\frac{1}{4}A_1+\frac{1}{8}A_1+\frac{1}{4}A_3-\frac{1}{4}A_1-\frac{1}{6}A_3+\frac{1}{8}\mathbf{tr}\left[A_2\right] \\
& -\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown\widehat{p}}{\widehat{p}} \\
& -\frac{P}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown\widehat{p}}{\widehat{p}} \\
& -\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown\widehat{p}}{\widehat{p}} \\
& +\frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown\widehat{p}}{\widehat{p}} \\
& +\frac{P}{4}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown^2\widehat{p}}{\widehat{p}}\right]+\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown^2\widehat{p}}{\widehat{p}}\right] \\
= \;& \frac{P+2}{16}A_1-\frac{P+2}{16}\mathbf{tr}\left[A_2\right]+\frac{P+2}{24}A_3-\frac{P+2}{4}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown\widehat{p}}{\widehat{p}} \\
& +\frac{P+2}{4}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\bigtriangledown^2\widehat{p}}{\widehat{p}}\right] \\
= \;& \frac{P+2}{16}A_1-\frac{P+2}{16}\mathbf{tr}\left[A_2\right]+\frac{P+2}{24}A_3-\frac{P+2}{4}C_{21}+\frac{P+2}{4}C_2+\frac{P+2}{4}C_{23},
\end{aligned}$$

where C_{21}, C_2 and C_{23} are defined in Lemma 2.3. Clearly the first three terms are not related to the prior while the last three terms are related to the prior. We can also show that

$$-\frac{1}{n}\sum_{t=1}^{n}B_{t,3}$$

$$
\begin{aligned}
&= -\left(\frac{1}{n}\sum_{t=1}^{n} B_{t,1}\right) B_4 \\
&= \frac{P}{2}\left(-\frac{1}{2}\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\right] + \frac{1}{2}vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)'\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla\widehat{p}}{\widehat{p}}\right) \\
&\quad +\frac{P}{2}\left(-\frac{1}{8}A_1 - \frac{1}{12}A_3 + \frac{1}{8}\mathbf{tr}\left[A_2\right]\right) \\
&= -\frac{P}{4}C_2 - \frac{P}{4}C_{23} + \frac{P}{4}C_{21} - \frac{P}{16}A_1 - \frac{P}{24}A_3 + \frac{P}{16}\mathbf{tr}\left[A_2\right]
\end{aligned}
$$

where

$$
\mathbf{tr}\left[\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\frac{\nabla^2\widehat{p}}{\widehat{p}}\right] = C_2 + C_{23}.
$$

Hence,

$$
\begin{aligned}
&\sum_{t=1}^{n}\int l_t\left(\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}|\mathbf{y}\right)\mathrm{d}\boldsymbol{\theta} \\
&= \sum_{t=1}^{n} l_t\left(\widehat{\boldsymbol{\theta}}_n\right) - \frac{P}{2} \\
&\quad +\frac{1}{n}\left(\frac{P+2}{16}A_1 - \frac{P+2}{16}\mathbf{tr}\left[A_2\right] + \frac{P+2}{24}A_3 - \frac{P+2}{4}C_{21} + \frac{P+2}{4}C_2 + \frac{P+2}{4}C_{23}\right) \\
&\quad +\frac{1}{n}\left[-\frac{P}{4}C_2 - \frac{P}{4}C_{23} + \frac{P}{4}C_{21} - \frac{P}{16}A_1 - \frac{P}{24}A_3 + \frac{P}{16}\mathbf{tr}\left[A_2\right]\right] + O_p\left(n^{-2}\right) \\
&= \sum_{t=1}^{n} l_t\left(\widehat{\boldsymbol{\theta}}_n\right) - \frac{P}{2} \\
&\quad +\frac{1}{n}\left(\frac{1}{8}A_1 - \frac{1}{8}\mathbf{tr}\left[A_2\right] + \frac{1}{12}A_3 - \frac{1}{2}C_{21} + \frac{1}{2}C_2 + \frac{1}{2}C_{23}\right) + O_p\left(n^{-2}\right).
\end{aligned}
$$

From the stochastic expansion, we have

$$
\begin{aligned}
\overline{\boldsymbol{\theta}}_n &= \widehat{\boldsymbol{\theta}}_n - \frac{1}{n}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla\widehat{\pi} \\
&\quad +\frac{1}{2n}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right) + O_p\left(\frac{1}{n^2}\right),
\end{aligned}
$$

and

$$\widehat{\boldsymbol{\theta}}_n - \overline{\boldsymbol{\theta}}_n = \frac{1}{n}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla\widehat{\pi} - \frac{1}{2n}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n^{(3)}\left(\widehat{\boldsymbol{\theta}}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right) + O_p\left(\frac{1}{n^2}\right). \quad (86)$$

By the Taylor expansion, we get

$$\begin{aligned}
& \ln p\left(\mathbf{y}|\bar{\boldsymbol{\theta}}_n\right) \\
= & \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right) + \frac{\partial \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}'}\left(\bar{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n\right) \\
& + \frac{1}{2}\left(\bar{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n\right)' \frac{\partial^2 \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\left(\bar{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n\right) + O_p\left(\frac{1}{n^2}\right) \\
= & \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right) + \frac{1}{2n}\nabla\widehat{\pi}'\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\nabla\widehat{\pi} \\
& + \frac{1}{8n} vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right)' \frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1} \\
& \times \frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right) \\
& - \frac{1}{2n}\nabla\widehat{\pi}'\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1} \\
& \times \frac{1}{n}\sum_{t=1}^{n}\nabla^3 l_t\left(\widehat{\boldsymbol{\theta}}_n\right)' vec\left(\bar{\mathbf{H}}_n\left(\widehat{\boldsymbol{\theta}}_n\right)^{-1}\right) + O_p(n^{-2}) \\
= & \ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right) - \frac{1}{2n}C_{21} + \frac{1}{2n}C_{23} + \frac{1}{8n}A_1 + O_p(n^{-2}). \quad (87)
\end{aligned}$$

Hence, from (86) and (87) we have,

$$\begin{aligned}
P_D = & -2\sum_{t=1}^{n}\int l_t(\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{y})\, \mathrm{d}\boldsymbol{\theta} + 2\ln p\left(\mathbf{y}|\bar{\boldsymbol{\theta}}_n\right) \\
= & -2\ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right) + P \\
& + \frac{1}{n}\left(-\frac{1}{4}A_1 + \frac{1}{4}\mathbf{tr}\left[A_2\right] - \frac{1}{6}A_3 + C_{21} - C_2 - C_{23}\right) \\
& + 2\ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right) - \frac{1}{n}C_{21} + \frac{1}{n}C_{23} + \frac{1}{4n}A_1 + O_p(n^{-2}) \\
= & P + \frac{1}{n}\left(\frac{1}{4}\mathbf{tr}\left[A_2\right] - \frac{1}{6}A_3 - C_2\right) + O_p\left(n^{-2}\right). \quad (88)
\end{aligned}$$

From the definition of DIC, (87), and (88), we have

$$
\begin{aligned}
\text{DIC} &= -4\sum_{t=1}^{n}\int l_t(\boldsymbol{\theta})\,p(\boldsymbol{\theta}|\mathbf{y})\,d\boldsymbol{\theta}+2\ln p\left(\mathbf{y}|\bar{\boldsymbol{\theta}}\right) \\
&= -2\sum_{t=1}^{n}\int l_t(\boldsymbol{\theta})\,p(\boldsymbol{\theta}|\mathbf{y})\,d\boldsymbol{\theta}+P_D \\
&= -2\ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right)+P+\frac{1}{n}\left(-\frac{1}{4}A_1+\frac{1}{4}\mathbf{tr}\left[A_2\right]-\frac{1}{6}A_3+C_{21}-C_2-C_{23}\right) \\
&\quad +P+\frac{1}{n}\left(\frac{1}{4}\mathbf{tr}\left[A_2\right]-\frac{1}{6}A_3-C_2\right)+O_p\left(n^{-2}\right) \\
&= -2\ln p\left(\mathbf{y}|\widehat{\boldsymbol{\theta}}_n\right)+2P \\
&\quad +\frac{1}{n}\left[-\frac{1}{4}A_1+\frac{1}{2}\mathbf{tr}\left[A_2\right]-\frac{1}{3}A_3+C_{21}-2C_2-C_{23}\right]+O_p\left(n^{-2}\right) \\
&= \text{AIC}+\frac{1}{n}D_1+\frac{1}{n}D_2+O_p\left(n^{-2}\right).
\end{aligned}
$$

References

Chung, K. L. (2001). *A course in probability theory.* Academic press.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability.* New York: Springer.

Kamat, A. R. (1953). Incomplete and Absolute Moments of the Multivariate Normal Distribution with Some Applications, *Biometrika*, **40**, 20–34.

Kan, R. and Robotti, C. (2017). On Moments of Folded and Truncated Multivariate Normal Distributions, *Journal of Computational and Graphical Statistics*, **26(4)**, 930–934.

Kass, R., Tierney, L., and Kadane, J.B. (1990). The validity of posterior expansions based on Laplace's Method. in *Bayesian and Likelihood Methods in Statistics and Econometrics*, ed. by S. Geisser, J.S. Hodges, S.J. Press and A. Zellner. Elsevier Science Publishers B.V.: North-Holland.

Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition. John Wiley and Sons.

Miyata, Y. (2004). Fully exponential Laplace approximations using asymptotic modes. *Journal of the American Statistical Association*, **99,** 1037–1049.

Miyata, Y. (2010). Laplace approximations to means and variances with asymptotic modes. *Journal of Statistical Planning and Inference*, **140**, 382–392.

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association,* **81**, 82–86.

Vignat, C. (2012). A generalized Isserlis theorem for location mixtures of Gaussian random vectors. *Statistics & Probability Letters*, **82(1)**, 67-71.

Withers, C.S. (1985). The moments of the multivariate normal, *Bulletin of the Australian Mathematical Society*, **32**, 103-107.