No.174

Asymptotic cumulants of some information criteria (2nd version)

Haruhiko Ogasawara

November 2015

Department of Information and Management Science
Otaru University of Commerce

Discussion Paper, No.174, November, 2015, Center for Business Creation,
Otaru University of Commerce, Otaru, Japan.

Asymptotic cumulants of some information criteria
(2nd version)

Haruhiko Ogasawara
Otaru University of Commerce

This work was partially supported by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology (JSPS KAKENHI, Grant No.26330031). Author's address: Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501, Japan. Email: hogasa@res.otaru-uc.ac.jp

Asymptotic cumulants of some information criteria

Asymptotic cumulants of the Akaike and Takeuchi information criteria are given under possible model misspecification up to the fourth order with the higher-order asymptotic variances, where two versions of the latter information criterion are defined using observed and estimated expected information matrices. The asymptotic cumulants are provided before and after studentization using the parameter estimates by the weighted score method, which include the maximum likelihood and Bayes modal estimators as special cases. Higher-order bias corrections of the criteria are derived using log-likelihood derivatives, which yields simple results for cases under canonical parametrization in the exponential family. The results are illustrated by three examples. Simulations for model selection in regression and interval estimation are also given.

Keywords: Akaike information criterion; Takeuchi information criterion; Kullback-Leibler distance; canonical parameters; higher-order bias correction.

1. Introduction

Typical information criteria are given by Akaike (1973) and Takeuchi (1976), which are called the Akaike information criterion (AIC) and Takeuchi information criterion (TIC), respectively. The criteria are used to assess the goodness of statistical models based on the Kullback-Leibler (1951) distance using the maximum likelihood estimators (MLEs) of associated parameters. In the AIC, it is assumed that a posited model holds or that a true model is a special case of the model employed. On the other hand in the TIC, possible model misspecification is considered. Stone (1977) derived the TIC in the context of cross validation. Linhart and Zucchini (1986, Proposition 2, Appendix A.2.1) also derived the TIC. For properties of the TIC, see Shibata (1989).

After the AIC and TIC were coined, information criteria with similar purposes have been introduced by e.g., Schwarz (1978; the Bayesian information criterion, BIC); Kishino and Hasegawa (1989), Ishiguro, Sakamoto and Kitagawa (1997; the extended information criterion, EIC), Shimodaira and Hasegawa (1999) for the methods using the bootstrap; Shibata (1989; the regularization information criterion, RIC) and, Konishi and Kitagawa (1996; the generalized information criterion, GIC; see also Konishi & Kitagawa, 2003; 2008, Chapters 5 to 8). In the RIC and GIC, the exclusive usage of the MLEs by the AIC and TIC was relaxed to cover e.g., robust and ridge-type estimators. For other information criteria, see Konishi and Kitagawa (2008) and Burnham and Anderson (2010).

The above information criteria are seen as point estimators of a corresponding population quantity with bias correction under correct model specification for the AIC and under possible model misspecification for the TIC, RIC and GIC. The population quantity is the so-called mean expected log-likelihood (Sakamoto, Ishiguro & Kitagawa, 1986, Equation (4.9)) associated with the Kullback-Leibler distance, where independent two-fold expectation is used one for data in the future for prediction and the other for current data for estimation with the same sample size denoted by n. When n increases, the population value increases proportionately in an asymptotic sense. On the other hand, the terms of bias correction are of order $O(1)$ for the AIC and $O_p(1)$ for the TIC, RIC and GIC. For tractability, divide the information criteria by n yielding quantities per observation as $n^{-1}\mathrm{AIC}$ and $n^{-1}\mathrm{TIC}$. Then, the population value mentioned above is written symbolically as $O(1)+O(n^{-1})$ depending on *n*. The situation is somewhat different from that of typical parameter estimators as MLEs, where the population parameters usually do not depend on *n*. When *n* becomes infinitely large, the population value $O(1)+O(n^{-1})$ for e.g., $n^{-1}\mathrm{AIC}$ becomes $O(1)$, which is the expected log-likelihood averaged over observations, where the parameters are evaluated by their population values followed by expectation. The last population value of order $O(1)$ is also of interest as well as that of $O(1)+O(n^{-1})$.

The bias correction of the TIC was extended to the higher-order version by Konishi and Kitagawa (2003), which gives a refined point estimator of the population counterpart. On the other hand, statistical testing of the difference of the information criteria for different models have been developed by Steiger, Shapiro and Browne (1985) and Shimodaira (1997) under local alternatives and by Linhart (1988), and Kishino and Hasegawa (1989) under fixed alternatives. Interval estimation of the corresponding population quantities can also be done in similar manners. While the above methods of testing and estimation is for general models, the results for special models are available for the higher-order bias correction by Sugiura (1978), Yanagihara, Sekiguchi and Fujikoshi (2003) and Kamo, Yanagihara and Satoh (2013), and the asymptotic cumulants for standardized estimators by Yanagihara and Ohmoto (2005) among others.

One of the purposes of this study is to derive general expressions of the higher-order bias corrections of $n^{-1}\mathrm{AIC}$ and $n^{-1}\mathrm{TIC}$ based on the parameter estimators by the weighted score method under possible model misspecification, where the expression is different from that of Konishi and Kitagawa (2003). The expression is given by the log-likelihood derivatives, which yields some transparent results for e.g., the cases of the natural exponential family. Note that Konishi and Kitagawa (2003) used the von Mises calculus (von Mises, 1947; Withers, 1983).

The second purpose is to give general formulas for the asymptotic cumulants of $n^{-1}\mathrm{AIC}$ and $n^{-1}\mathrm{TIC}$ up to the fourth order and the higher-order asymptotic variances before and after studentization for testing and interval estimation of the population quantities of interest. Three examples using basic distributions in statistics are shown. The first two examples of the exponential and normal distributions use MLEs under model misspecification, while the third example of the Bernoulli distribution uses the parameter estimators by the weighted score under correct model specification. Simulations for model selection in regression and interval estimation are also given, where the higher-order bias correction of $n^{-1}\mathrm{AIC}$ are used for model selection.

2. The higher-order asymptotic biases

Let $\boldsymbol{\theta}$ be a $q \times 1$ vector of parameters in a statistical model with a $p \times 1$ vector $\mathbf{x}^*$ of observable variables. Then, the log-likelihood of $\boldsymbol{\theta}$ based on *n* i.i.d. observations is denoted by

$$l \equiv l(\boldsymbol{\theta} \,|\, \mathbf{X}^*) \equiv \sum_{j=1}^{n} l_j \equiv \sum_{j=1}^{n} \log f(\mathbf{x}_j^* \,|\boldsymbol{\theta}) \equiv f(\mathbf{X}^* \,|\, \boldsymbol{\theta}), \tag{2.1}$$

where $\mathbf{X}^*$ is a $n \times p$ matrix whose rows $(\mathbf{x}_j^*{}', \; j = 1,..,n)$ are independent copies of $\mathbf{x}^*$ or their realizations for simplicity of notation, and $f(\mathbf{x}_j^* \,|\, \boldsymbol{\theta})$ is the probability density/mass function for a posited statistical model. The log-likelihood averaged over observations is denoted by $\bar{l} = n^{-1} l$. Define

$$\hat{\bar{l}}_{\mathrm{ML}} = \bar{l}\,(\hat{\boldsymbol{\theta}}_{\mathrm{ML}} \,|\, \mathbf{X}^*) = \bar{l}\,\{\boldsymbol{\theta}_{\mathrm{ML}}(\mathbf{X}^*) \,|\, \mathbf{X}^*\}, \tag{2.2}$$

where $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ is the MLE of the corresponding population quantity $\boldsymbol{\theta}_0$. Let $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ be the vector of the parameter estimators by the weighted score method (WSEs) or the solution of $\boldsymbol{\theta}$ satisfying

$$\frac{\partial \bar{l}\,(\boldsymbol{\theta} \,|\, \mathbf{X}^*)}{\partial \boldsymbol{\theta}} + n^{-1}\mathbf{q}^* = \mathbf{0}, \tag{2.3}$$

where $\mathbf{q}^* = \mathbf{q}^*(\boldsymbol{\theta})$, a function of $\boldsymbol{\theta}$, is a $q \times 1$ weight vector, which becomes the log-prior derivatives in the case of Bayesian estimation but can be other general weights. Define

$$\hat{\bar{l}}_{\mathrm{W}} = \bar{l}\,(\hat{\boldsymbol{\theta}}_{\mathrm{W}} \,|\, \mathbf{X}^*) = \bar{l}\,\{\boldsymbol{\theta}_{\mathrm{W}}(\mathbf{X}^*) \,|\, \mathbf{X}^*\}, \tag{2.4}$$

whose special case is $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ in (2.2) when $\mathbf{q}^* = \mathbf{0}$. Let $\mathbf{Z}^*$ be an independent copy of $\mathbf{X}^*$, where $\mathbf{Z}^*$ is interpreted as an independent data set in the future with the same sample size as *n* from the viewpoint of prediction. Define

$$\bar{l}_0^* = \mathrm{E}_g\{\bar{l}(\boldsymbol{\theta}_0 \,|\, \mathbf{Z}^*)\} = \int_{R(\mathbf{Z})} \bar{l}(\boldsymbol{\theta}_0 \,|\, \mathbf{Z}) g(\mathbf{Z} \,|\, \boldsymbol{\zeta}_0) \,\mathrm{d}\mathbf{Z}, \tag{2.5}$$

where $g(\mathbf{Z} \,|\, \boldsymbol{\zeta}_0)$ is the true density of $\mathbf{Z}^*$ determined by the parameter vector $\boldsymbol{\zeta}_0$ of an appropriate size, and is possibly different from $f(\mathbf{Z} \,|\, \boldsymbol{\theta}_0)$. Equation (2.5) is to be interpreted as the corresponding summation when $g(\mathbf{Z} \,|\, \boldsymbol{\zeta}_0)$ is a probability mass. Similarly, define

$$\bar{l}_0 = \bar{l}(\boldsymbol{\theta}_0 \,|\, \mathbf{X}^*) = O_p(1) \;\text{ with }\; \mathrm{E}_g(\bar{l}_0) = \bar{l}_0^* \tag{2.6}$$

and $$\hat{\bar{l}}_{\mathrm{W}}^* = \int_{R(\mathbf{Z})} \bar{l}(\hat{\boldsymbol{\theta}}_{\mathrm{W}} \,|\, \mathbf{Z}) g(\mathbf{Z} \,|\, \boldsymbol{\zeta}_0) \,\mathrm{d}\mathbf{Z} = \int_{R(\mathbf{Z})} \bar{l}\{\boldsymbol{\theta}_{\mathrm{W}}(\mathbf{X}^*) \,|\, \mathbf{Z}\} g(\mathbf{Z} \,|\, \boldsymbol{\zeta}_0) \,\mathrm{d}\mathbf{Z} = O_p(1). \tag{2.7}$$

It is assumed that

$$-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}} - \hat{\bar{l}}_{\mathrm{W}}^*) = n^{-1}b_1 + n^{-2}b_2 + O(n^{-3}) \tag{2.8}$$

holds, where $n^{-1}b_1$ and $n^{-2}b_2$ are defined as the asymptotic biases up to order $O(n^{-2})$ of $-2\hat{\bar{l}}_{\mathrm{W}}$ whose population counterpart is $-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*) = O(1)$ for the AIC and TIC with $n^{-2}b_2$ being the higher-order added asymptotic bias.

Theorem 1. *Under (2.8) with regularity conditions for (A1.1) and (A1.2) in Subsection A1 of the appendix, the asymptotic biases* $n^{-1}b_1$ *and* $n^{-2}b_2$ *of* $-2\hat{\bar{l}}_{\mathrm{W}}^*$ *up to order* $O(n^{-2})$, *based on the WSE* $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ *derived by the estimation equation of (2.3), are given by*

$$\begin{aligned} &-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}} - \hat{\bar{l}}_{\mathrm{W}}^*) \\ &= n^{-1}2\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}) + n^{-2}(c_1 + c_2 + c_3) + O(n^{-3}) = n^{-1}b_1 + n^{-2}b_2 + O(n^{-3}), \end{aligned} \tag{2.9}$$

where c_1, c_2 *and* c_3 *are obtained by (A1.5) to (A1.7), respectively.*

For the proof of Theorem 1, see Subsection A1 of the appendix.

From (A1.5) to (A1.7), we find that b_1 and c_3 do not depend on $\mathbf{q}_0^*$ and are common to the results by the MLE $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ and the WSE $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ while c_1 and c_2 depend on $\mathbf{q}_0^*$. A considerably simplified result is obtained in the following case.

Corollary 1. *When the vector of canonical parameters in the exponential family of distributions is used under possible model misspecification,*

$$-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}} - \hat{\bar{l}}_{\mathrm{W}}^*) = n^{-1}b_1 + n^{-2}c_1 + O(n^{-3}) \quad \textit{with} \quad b_2 = c_1 \quad \textit{and} \quad c_2 = c_3 = 0\,, \tag{2.10}$$

where c_1 *is simplified as*

$$\begin{aligned}
&-2\mathrm{E}_g\left\{\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'}(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)\right\} = 2n^{-1}\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}) \\
&\quad - 2n^{-2}\Bigg[\underset{(\mathrm{A})}{\;}\sum_{a,b,c=1}^{q} (\boldsymbol{\Lambda}^{(2-2)})_{(c:a,b)} n^2 \mathrm{E}_g\left(\frac{\partial \bar{l}}{\partial \theta_{0a}}\frac{\partial \bar{l}}{\partial \theta_{0b}}\frac{\partial \bar{l}}{\partial \theta_{0c}}\right) \\
&\qquad + \sum_{a,b,c,d=1}^{q} (\boldsymbol{\Lambda}^{(3-4)})_{(d:a,b,c)} (\gamma_{ab}\gamma_{cd} + \gamma_{ac}\gamma_{bd} + \gamma_{ad}\gamma_{bc}) \\
&\quad + \mathrm{tr}\left(\frac{\partial \mathbf{q}^*}{\partial \boldsymbol{\theta}_0{}'}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\right) - \mathrm{tr}[\mathrm{E}_g(\mathbf{J}_0^{(3)})\{(\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^*) \otimes (\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1})\}] \;\Bigg]_{(\mathrm{A})} + O(n^{-3}) \\
&= n^{-1}b_1 + n^{-2}c_1 + O(n^{-3}).
\end{aligned} \tag{2.11}$$

For the definitions of the undefined quantities and the proof of Corollary 1, see Subsection A2 of the appendix.

The above result becomes further simplified in the following case.

Corollary 2. *Under correct model specification and canonical parametrization*,

$$\begin{aligned}
c_1 &= -\boldsymbol{\kappa}_{f3}{}'(\tilde{\mathbf{x}}^*)\boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*) - \boldsymbol{\kappa}_{f3}{}'(\tilde{\mathbf{x}}^*)[\mathbf{I}_{(q)} \otimes \{\mathrm{vec}(\mathbf{I}_{(q)})\mathrm{vec}'(\mathbf{I}_{(q)})\}]\boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*) \\
&\quad + \boldsymbol{\kappa}_{f4}{}'(\tilde{\mathbf{x}}^*)\mathrm{vec}(\mathbf{I}_{(q^2)}),
\end{aligned} \tag{2.12}$$

where $\tilde{\mathbf{x}}^* = \mathbf{I}_0^{-1/2}\mathbf{x}^*$; $\mathbf{x}^*$ *is the* $q \times 1$ *vector of observable variables associated with the minimum sufficient statistics* (*p* = *q*); $\boldsymbol{\kappa}_{fj}(\cdot)$ *is the* $q^j \times 1$ *vector of the j-th multivariate cumulants of a* $q \times 1$ *random vector in parentheses using the distribution* $f(\mathbf{x}^* \mid \boldsymbol{\theta}_0)$ *for* $\mathbf{x}^*$; $\mathbf{I}_0^{1/2}$ *is a non-negative definite symmetric matrix-square-root of* $\mathbf{I}_0$ *(the information matrix per observation) with* $\mathbf{I}_0^{-1/2} = (\mathbf{I}_0^{1/2})^{-1}$ *under the assumption of its existence; and* $\mathbf{I}_{(q)}$ *is the* $q \times q$ *identity matrix*.

For the proof of Corollary 2, see Subsection A3 of the appendix.

Under correct model specification, since $\mathrm{cov}_f(\mathbf{x}^*) = \mathbf{I}_0$ due to canonical parametrization, $\tilde{\mathbf{x}}^*$ is the vector of standardized variables with

$\mathrm{cov}_f(\tilde{\mathbf{x}}^*) = \mathrm{cov}_f\left(\mathbf{I}_0^{-1/2}\frac{\partial l_j}{\partial \boldsymbol{\theta}_0}\right) \equiv \mathrm{cov}_f\left(\frac{\partial \tilde{l}_j}{\partial \boldsymbol{\theta}_0}\right) = \mathbf{I}_{(q)}$, where $\mathrm{cov}_f(\cdot)$ is the exact covariance matrix using $f(\mathbf{x}^* \mid \boldsymbol{\theta}_0)$. Then, $\boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*)$ and $\boldsymbol{\kappa}_{f3}(\partial \tilde{l}_j / \partial \boldsymbol{\theta}_0)(= \boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*))$ are seen as $q^3 \times 1$ vectors of the multivariate skewnesses of $\tilde{\mathbf{x}}^*$ and $\partial \tilde{l}_j / \partial \boldsymbol{\theta}_0$, respectively. Similarly, $\boldsymbol{\kappa}_{f4}(\tilde{\mathbf{x}}^*)$ is seen as a $q^4 \times 1$ vector of the multivariate kurtoses of $\tilde{\mathbf{x}}^*$. In the univariate case, (2.12) becomes the sum of -2 times the squared skewness and the excess kurtosis. A special case of the expression of (2.12) is shown in Poisson regression by Kamo et al. (2013, Equation (8)). Other expressions and that for Poisson regression by a unified formula will be shown in a later section.

Similarly, under correct model specification, b_1 in the asymptotic bias of order $O(n^{-1})$ in (2.18) is also written as

$$\begin{aligned} b_1 &= 2\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}) = -2q = -2\mathrm{vec}'(\mathbf{I}_0)\mathrm{vec}(\mathbf{I}_0^{-1}) = -2\boldsymbol{\kappa}_{f2}{}'(\mathbf{x}^*)\boldsymbol{\kappa}_{f2}\left(\mathbf{I}_0^{-1}\frac{\partial l_j}{\partial \boldsymbol{\theta}_0}\right) \\ &= -2\boldsymbol{\kappa}_{f2}{}'(\tilde{\mathbf{x}}^*)\boldsymbol{\kappa}_{f2}\left(\frac{\partial \tilde{l}_j}{\partial \boldsymbol{\theta}_0}\right) = -2\boldsymbol{\kappa}_{f2}{}'(\tilde{\mathbf{x}}^*)\boldsymbol{\kappa}_{f2}(\tilde{\mathbf{x}}^*). \end{aligned} \tag{2.13}$$

The above results give

Corollary 3. *Under correct model specification and canonical parametrization in the exponential family, when the multivariate skewnesses and kurtoses of the associated observable variables are zero, the MLE gives*

$$-2\mathrm{E}_f(\hat{\bar{l}}_{\mathrm{ML}} - \hat{\bar{l}}_{\mathrm{ML}}^*) = -n^{-1}2q + O(n^{-3}) \quad (b_1 = -2q, b_2 = c_1 = c_2 = c_3 = 0) \tag{2.14}$$

where $\mathrm{E}_f(\cdot)$ *is defined using* $f(\mathbf{x}^* \mid \boldsymbol{\theta}_0)$ *similarly to* $\mathrm{E}_g(\cdot)$.

This can happen, for example, when the covariance matrix in the multivariate normal distribution is known, where the vector of canonical parameters is the mean vector.

Corollary 4. *When the covariance matrix* $\boldsymbol{\Sigma}$ *of the q-variate normal distribution is known, the MLE (the usual sample mean vector* $\bar{\mathbf{x}}$*) of the population mean vector* $\boldsymbol{\mu}_0$ *under possible model misspecification gives*

$$-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{ML}} - \hat{\bar{l}}^*_{\mathrm{ML}}) = -n^{-1}2q \tag{2.15}$$

For the proof of Corollary 4, see Subsection A4 of the appendix. Note that there is no remainder term in (2.15).

3. Bias correction for the AIC and TIC

Define

$$\begin{aligned} n^{-1}\mathrm{AIC}_{\mathrm{W}} &= -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}2q, \\ n^{-1}\mathrm{TIC}^{(1)}_{\mathrm{W}} &= -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}\mathrm{tr}(-\hat{\mathbf{L}}^{-1}_{\mathrm{W}}\hat{\boldsymbol{\Gamma}}_{\mathrm{W}}) \end{aligned} \tag{3.1}$$

$$\text{and } n^{-1}\mathrm{TIC}^{(2)}_{\mathrm{W}} = -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}\mathrm{tr}(\hat{\mathbf{I}}^{(-\boldsymbol{\Lambda})-1}_{\mathrm{W}}\hat{\mathbf{I}}^{(\boldsymbol{\Gamma})}_{\mathrm{W}}) \text{ with } \hat{\mathbf{I}}^{(-\boldsymbol{\Lambda})-1}_{\mathrm{W}} = (\hat{\mathbf{I}}^{(-\boldsymbol{\Lambda})}_{\mathrm{W}})^{-1},$$

where

$$\hat{\mathbf{L}}_{\mathrm{W}} = \frac{\partial^2 \bar{l}}{\partial\hat{\boldsymbol{\theta}}_{\mathrm{W}}\partial\hat{\boldsymbol{\theta}}_{\mathrm{W}}'},\ \hat{\boldsymbol{\Gamma}}_{\mathrm{W}} = n^{-1}\sum_{j=1}^{n}\frac{\partial l_j}{\partial\hat{\boldsymbol{\theta}}_{\mathrm{W}}}\frac{\partial l_j}{\partial\hat{\boldsymbol{\theta}}_{\mathrm{W}}'},\ \hat{\mathbf{I}}^{(-\boldsymbol{\Lambda})}_{\mathrm{W}} = \left\{-\mathrm{E}_g\left(\frac{\partial^2\bar{l}}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right)\right\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{W}}} \text{ and}$$

$$\hat{\mathbf{I}}^{(\boldsymbol{\Gamma})}_{\mathrm{W}} = \left\{\mathrm{E}_g\left(\frac{\partial l_j}{\partial\boldsymbol{\theta}}\frac{\partial l_j}{\partial\boldsymbol{\theta}'}\right)\right\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{W}}}. \tag{3.2}$$

When the MLE is used, the subscript W in (3.1) becomes ML with $\mathrm{AIC}_{\mathrm{ML}}$ = AIC (the usual AIC), $\mathrm{TIC}^{(j)}_{\mathrm{ML}} = \mathrm{TIC}^{(j)} (j = 1,2)$. The original definition of the Takeuchi information criterion (Takeuchi, 1976, Equation (15)) denoted by $\mathrm{TIC}_{\mathrm{ML}}$ = TIC seems to be $\mathrm{TIC}^{(2)}_{\mathrm{ML}} = \mathrm{TIC}^{(2)}$ in (3.1), while the definition of the TIC by Linhart and Zucchini (1986, p.245), Konishi and Kitagawa (2008, p.60) and Burnham and Anderson (2010, Subsection 7.3.1) is $\mathrm{TIC}^{(1)}_{\mathrm{ML}} = \mathrm{TIC}^{(1)}$ in (3.1). The two matrices $-\hat{\mathbf{L}}_{\mathrm{W}}$ and $\hat{\boldsymbol{\Gamma}}_{\mathrm{W}}$ are observed information matrices given by $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ and $\mathbf{X}^*$, which are estimators of $-\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$, respectively, and become the estimators of $\mathbf{I}_0$ under correct model specification. The two matrices $\hat{\mathbf{I}}^{(-\boldsymbol{\Lambda})}_{\mathrm{W}}$ and $\hat{\mathbf{I}}^{(\boldsymbol{\Gamma})}_{\mathrm{W}}$ are also estimators of $-\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$, respectively, and are the expected information matrices followed by estimation using $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ without $\mathbf{X}^*$ except in $\boldsymbol{\theta}_{\mathrm{W}}(\mathbf{X}^*)$. Since it is often difficult to derive the expectation $\mathrm{E}_g(\cdot)$ in (3.2) when

$g(\mathbf{x}^* \mid \boldsymbol{\zeta}_0)$ is unknown, $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(1)}$ is of practical use though $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(1)}$ is more complicated than $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(2)}$. The remaining combinations $n^{-1}\mathrm{tr}(-\hat{\mathbf{L}}_{\mathrm{W}}^{-1}\hat{\mathbf{I}}_{\mathrm{W}}^{(\boldsymbol{\Gamma})})$ and $n^{-1}\mathrm{tr}(\hat{\mathbf{I}}_{\mathrm{W}}^{(-\boldsymbol{\Lambda})-1}\hat{\boldsymbol{\Gamma}}_{\mathrm{W}})$ for the correction term are not dealt with in this paper.

The higher-order bias correction of $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ is meaningless under model misspecification since the term $n^{-1}2q$ for bias correction is incorrect and should be replaced by that of $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(\bullet)}$ which stands generically for $\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$. Consequently, this reduces to the higher-order bias correction of $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(\bullet)}$ and will be dealt with later.

Theorem 2. *Assume that a statistical model holds. Then, under regularity conditions, define*

$$n^{-1}\mathrm{CAIC}_{\mathrm{W}} \equiv n^{-1}\mathrm{AIC}_{\mathrm{W}} - n^{-2}\hat{b}_2 = -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}2q - n^{-2}(\hat{c}_1 + \hat{c}_2 + \hat{c}_3). \tag{3.3}$$

Then, $\mathrm{E}_f(n^{-1}\mathrm{CAIC}_{\mathrm{W}} + 2\hat{\bar{l}}_{\mathrm{W}}^*) = O(n^{-3})$, *where* $\hat{c}_1$, $\hat{c}_2$ *and* $\hat{c}_3$ *are consistent estimators of* c_1, c_2 *and* c_3, *respectively.*

In some special cases, $n^{-1}\mathrm{AIC}_{\mathrm{ML}} (= n^{-1}\mathrm{AIC})$ gives the same result as that of Theorem 2 i.e., $\mathrm{E}_f(n^{-1}\mathrm{AIC} + 2\hat{\bar{l}}_{\mathrm{ML}}^*) = O(n^{-3})$. When the multivariate skewnesses and kurtoses of the associated observable variables are zero, from Corollary 3 we have this result. Similarly, when the covariance matrix of the multivariate normal distribution is known, Corollary 4 using the MLE of the mean vector gives the exact result $\mathrm{E}_g(n^{-1}\mathrm{AIC} + 2\hat{\bar{l}}_{\mathrm{ML}}^*) = 0$ even under non-normality.

For $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(\bullet)}$, under possible model misspecification, define stochastic $\mathrm{tr}_{\Delta}^{(\mathrm{T}j)}$ and $\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T}j)}$ in the expansions of $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ as follows.

Definition 1.

$$\begin{aligned} n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(1)} &= -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}2\mathrm{tr}(-\hat{\mathbf{L}}_{\mathrm{W}}^{-1}\hat{\boldsymbol{\Gamma}}_{\mathrm{W}}) \\ &= -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}2\mathrm{tr}(-\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}) + 2(n^{-1}\mathrm{tr}_{\Delta}^{(\mathrm{T1})})_{O_p(n^{-3/2})} + 2(n^{-1}\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T1})})_{O_p(n^{-2})} + O_p(n^{-5/2}) \end{aligned} \tag{3.4}$$

and

$$
\begin{aligned}
n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(2)} &= -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}2\mathrm{tr}(\hat{\mathbf{I}}_{\mathrm{W}}^{(-\mathbf{\Lambda})-1}\hat{\mathbf{I}}_{\mathrm{W}}^{(\mathbf{\Gamma})}) \\
&= -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}2\mathrm{tr}(-\mathbf{\Lambda}^{-1}\mathbf{\Gamma}) + 2(n^{-1}\mathrm{tr}_{\Delta}^{(\mathrm{T2})})_{O_p(n^{-3/2})} + 2(n^{-1}\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T2})})_{O_p(n^{-2})} + O_p(n^{-5/2}).
\end{aligned} \tag{3.5}
$$

For (3.4), define stochastic $-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)}$ and $-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta\Delta)}$ as

$$
-\hat{\mathbf{L}}_{\mathrm{W}}^{-1} = -\mathbf{\Lambda}^{-1} + (-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)})_{O_p(n^{-1/2})} + (-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2}). \tag{3.6}
$$

Similarly, define stochastic $\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta)}$ and $\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta\Delta)}$ as

$$
\hat{\mathbf{\Gamma}}_{\mathrm{W}} = \mathbf{\Gamma} + (\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta)})_{O_p(n^{-1/2})} + (\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2})\,. \tag{3.7}
$$

The actual expressions of $-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)}$, $-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta\Delta)}$, $\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta)}$ and $\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta\Delta)}$ in (3.6) and (3.7) are given in Subsection A5 of the appendix.

From (3.6) and (3.7), we have

Lemma 1. *The stochastic correction term in (3.4) of* $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(1)}$ *in Definition 1 is expanded as*

$$
\begin{aligned}
& n^{-1}2\mathrm{tr}(-\hat{\mathbf{L}}_{\mathrm{W}}^{-1}\hat{\mathbf{\Gamma}}_{\mathrm{W}}) \\
&= n^{-1}2\mathrm{tr}(-\mathbf{\Lambda}^{-1}\mathbf{\Gamma}) + 2(n^{-1}\mathrm{tr}_{\Delta}^{(\mathrm{T1})})_{O_p(n^{-3/2})} + 2(n^{-1}\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T1})})_{O_p(n^{-2})} + O_p(n^{-5/2}) \\
&\equiv n^{-1}2\mathrm{tr}(-\mathbf{\Lambda}^{-1}\mathbf{\Gamma}) + 2\{n^{-1}\mathrm{tr}(-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)}\mathbf{\Gamma} - \mathbf{\Lambda}^{-1}\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta)})\}_{O_p(n^{-3/2})} \\
&\quad + 2\{n^{-1}\mathrm{tr}(-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)}\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta)} - \mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta\Delta)}\mathbf{\Gamma} - \mathbf{\Lambda}^{-1}\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta\Delta)})\}_{O_p(n^{-2})} + O_p(n^{-5/2}),
\end{aligned} \tag{3.8}
$$

where the stochastic quantities are given by (3.6) and (3.7).

For (3.5) of $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(2)}$, define stochastic $-\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta)}, -\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta\Delta)}, \mathbf{\Gamma}_{\mathbf{I}}^{(\Delta)}$ and $\mathbf{\Gamma}_{\mathbf{I}}^{(\Delta\Delta)}$ as

$$
\hat{\mathbf{I}}_{\mathrm{W}}^{(-\mathbf{\Lambda})-1} = -\mathbf{\Lambda}^{-1} + (-\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta)})_{O_p(n^{-1/2})} + (-\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2}), \tag{3.9}
$$

$$
\hat{\mathbf{I}}_{\mathrm{W}}^{(\mathbf{\Gamma})} = \mathbf{\Gamma} + (\mathbf{\Gamma}_{\mathbf{I}}^{(\Delta)})_{O_p(n^{-1/2})} + (\mathbf{\Gamma}_{\mathbf{I}}^{(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2}),
$$

where the actual expressions of $-\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta)}, -\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta\Delta)}, \mathbf{\Gamma}_{\mathbf{I}}^{(\Delta)}$ and $\mathbf{\Gamma}_{\mathbf{I}}^{(\Delta\Delta)}$ are given in Subsection A6 of the appendix.

Then, we have

Lemma 2. *The stochastic correction term in (3.5) of* $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(2)}$ *in Definition 1 is expanded as*

$$
\begin{aligned}
& n^{-1} 2\mathrm{tr}(\hat{\mathbf{I}}_{\mathrm{W}}^{(-\mathbf{\Lambda})-1}\hat{\mathbf{I}}_{\mathrm{W}}^{(\mathbf{\Gamma})}) \\
& = n^{-1}2\mathrm{tr}(-\mathbf{\Lambda}^{-1}\mathbf{\Gamma}) + 2(n^{-1}\mathrm{tr}_{\Delta}^{(\mathrm{T2})})_{O_p(n^{-3/2})} + 2(n^{-1}\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T2})})_{O_p(n^{-2})} + O_p(n^{-5/2}) \\
& \equiv n^{-1}2\mathrm{tr}(-\mathbf{\Lambda}^{-1}\mathbf{\Gamma}) + 2\{n^{-1}\mathrm{tr}(-\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta)}\mathbf{\Gamma} - \mathbf{\Lambda}^{-1}\mathbf{\Gamma}_{\mathbf{I}}^{(\Delta)})\}_{O_p(n^{-3/2})} \\
& \quad + 2\{n^{-1}\mathrm{tr}(-\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta)}\mathbf{\Gamma}_{\mathbf{I}}^{(\Delta)} - \mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta\Delta)}\mathbf{\Gamma} - \mathbf{\Lambda}^{-1}\mathbf{\Gamma}_{\mathbf{I}}^{(\Delta\Delta)})\}_{O_p(n^{-2})} + O_p(n^{-5/2}),
\end{aligned}
\tag{3.10}
$$

where the stochastic quantities are given by (3.9).

For the bias correction of $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(1)}$ (see (3.4) of Definition 1), we derive the expectations of the two stochastic terms $2(\mathrm{tr}_{\Delta}^{(\mathrm{T1})})_{O_p(n^{-1/2})}$ and $2(\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T1})})_{O_p(n^{-1})}$, where the former expectation becomes $2\mathrm{E}_g(\mathrm{tr}_{\Delta}^{(\mathrm{T1})}) = 2\mathrm{E}_g\{\mathrm{tr}(-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)}\mathbf{\Gamma} - \mathbf{\Lambda}^{-1}\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta)})\} = 0$ (see (3.8) of Lemma 1) since $\mathrm{E}_g(-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)}) = \mathrm{E}_g(\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta)}) = \mathbf{O}$ by construction. The latter expectation (see (3.8)) is denoted by

$$
\mathrm{E}_g\{2(\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T1})})\} = n^{-1}d^{(\mathrm{T1})}, \tag{3.11}
$$

where the actual expression of $d^{(\mathrm{T1})}$ is given in Subsection A7 of the appendix.

For $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(2)}$ (see (3.5) of Definition 1), similarly we have $2\mathrm{E}_g(\mathrm{tr}_{\Delta}^{(\mathrm{T2})}) = 2\mathrm{E}_g\{\mathrm{tr}(-\mathbf{\Lambda}_{\mathbf{I}}^{-1(\Delta)}\mathbf{\Gamma} - \mathbf{\Lambda}^{-1}\mathbf{\Gamma}_{\mathbf{I}}^{(\Delta)})\} = 0$ by construction and

$$
\mathrm{E}_g\{2(\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T2})})\} = n^{-1}d^{(\mathrm{T2})}, \tag{3.12}
$$

where the actual expression of $d^{(\mathrm{T2})}$ is given in Subsection A7 of the appendix.

The higher-order bias corrections of $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}\,(j=1,2)$ are given as follows:

Theorem 3. *Under possible model misspecification and some regularity conditions, define*

$$
n^{-1}\mathrm{CTIC}_{\mathrm{W}}^{(j)} \equiv n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)} - n^{-2}(\hat{b}_2 + \hat{d}^{(\mathrm{T}j)})\ (j=1,2), \tag{3.13}
$$

where $\hat{b}_2$ *and* $\hat{d}^{(\mathrm{T}j)}$ *are consistent estimators of* $b_2(=c_1+c_2+c_3)$ *and* $d^{(\mathrm{T}j)}$. *Then,* $\mathrm{E}_g(n^{-1}\mathrm{CTIC}_{\mathrm{W}}^{(j)} + 2\hat{\bar{l}}_{\mathrm{W}}^{*}) = O(n^{-3})$.

4. Asymptotic cumulants

In Section 2, the bias of $-2\hat{\bar{l}}_{\mathrm{W}}$ was defined as $-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}} - \hat{\bar{l}}_{\mathrm{W}}^*)$ (see (2.8)) with the definitions of $\hat{\bar{l}}_{\mathrm{W}}$ and $\hat{\bar{l}}_{\mathrm{W}}^*$ by (2.4) and (2.7), respectively. In this section, the asymptotic cumulants of $-2\hat{\bar{l}}_{\mathrm{W}}(=-2\bar{l}(\hat{\boldsymbol{\theta}}_{\mathrm{W}}, \mathbf{X}^*) = -2\bar{l}\{\boldsymbol{\theta}_{\mathrm{W}}(\mathbf{X}^*), \mathbf{X}^*\})$ using the density $g(\mathbf{X}^* \mid \boldsymbol{\zeta}_0)$ are given, where the bias is also defined as $-2\{\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}) - \bar{l}_0^*\}$ with $\bar{l}_0^*$ being the population counterpart of $\hat{\bar{l}}_{\mathrm{W}}$, which is the limiting value of $\hat{\bar{l}}_{\mathrm{W}}$ when *n* is infinitely large. The value and the notation of $\bar{l}_0^*$ are equal to those of (2.5) since

$$\begin{aligned}\bar{l}_0^* &= \mathrm{E}_g\{\bar{l}(\boldsymbol{\theta}_0 \mid \mathbf{X}^*)\} = \int_{R(\mathbf{X})} \bar{l}(\boldsymbol{\theta}_0 \mid \mathbf{X}) g(\mathbf{X} \mid \boldsymbol{\zeta}_0)\,\mathrm{d}\mathbf{X} \\ &= \int_{R(\mathbf{Z})} \bar{l}(\boldsymbol{\theta}_0 \mid \mathbf{Z}) g(\mathbf{Z} \mid \boldsymbol{\zeta}_0)\,\mathrm{d}\mathbf{Z} = \mathrm{E}_g\{\bar{l}(\boldsymbol{\theta}_0 \mid \mathbf{Z}^*)\}.\end{aligned} \tag{4.1}$$

The asymptotic cumulants of $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ and $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ are given before and after studentization up to the fourth order with the higher-order asymptotic variances. The studentization is for testing and interval estimation, where the population values of $-2\hat{\bar{l}}_{\mathrm{W}}$ are defined in two ways as $-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)$ and $-2\bar{l}_0^*$. While these two values are of order $O(1)$, the former depends on *n* in that the value is generally written as $O(1) + O(n^{-1}) + O(n^{-2}) + \cdots$. When *n* is infinitely large, $-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)$ becomes equal to $-2\bar{l}_0^*$. So, $-2\bar{l}_0^*$ is also of interest as well as $-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)$. Note that asymptotically unbiased point estimators of the latter up to order $O(n^{-1})$ are $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ under correct model specification and $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ under possible model misspecification.

Under possible model misspecification, assume that the following hold with the definitions of the asymptotic cumulants whose factors of $O(1)$ are $\alpha_{\mathrm{W}k}^{(\mathrm{A})}$ for $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ and $\alpha_{\mathrm{W}k}^{(\mathrm{T}j)}$ for $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ $(k=1,\ \Delta 1,\ 2,\ \Delta 2,\ 3,\ 4)$:

$$
\begin{aligned}
&\kappa_{g1}(n^{-1}\mathrm{AIC_W}) = -2(\bar{l}_0^*)_{O(1)} + n^{-1}\alpha_{\mathrm{W1}}^{(\mathrm{A})} + n^{-2}\alpha_{\mathrm{W\Delta 1}}^{(\mathrm{A})} + O(n^{-3}),\\
&\kappa_{g2}(n^{-1}\mathrm{AIC_W}) = n^{-1}\alpha_{\mathrm{W2}}^{(\mathrm{A})} + n^{-2}\alpha_{\mathrm{W\Delta 2}}^{(\mathrm{A})} + O(n^{-3}),\\
&\kappa_{g3}(n^{-1}\mathrm{AIC_W}) = n^{-2}\alpha_{\mathrm{W3}}^{(\mathrm{A})} + O(n^{-3}),\\
&\kappa_{g4}(n^{-1}\mathrm{AIC_W}) = n^{-3}\alpha_{\mathrm{W4}}^{(\mathrm{A})} + O(n^{-4}),\\
&\kappa_{g1}(n^{-1}\mathrm{TIC_W^{(\mathit{j})}}) = -2(\bar{l}_0^*)_{O(1)} + n^{-1}\alpha_{\mathrm{W1}}^{(\mathrm{T}j)} + n^{-2}\alpha_{\mathrm{W\Delta 1}}^{(\mathrm{T}j)} + O(n^{-3}),\\
&\kappa_{g2}(n^{-1}\mathrm{TIC_W^{(\mathit{j})}}) = n^{-1}\alpha_{\mathrm{W2}}^{(\mathrm{T}j)} + n^{-2}\alpha_{\mathrm{W\Delta 2}}^{(\mathrm{T}j)} + O(n^{-3}),\\
&\kappa_{g3}(n^{-1}\mathrm{TIC_W^{(\mathit{j})}}) = n^{-2}\alpha_{\mathrm{W3}}^{(\mathrm{T}j)} + O(n^{-3}),\\
&\kappa_{g4}(n^{-1}\mathrm{TIC_W^{(\mathit{j})}}) = n^{-3}\alpha_{\mathrm{W4}}^{(\mathrm{T}j)} + O(n^{-4}) \quad (j=1,\,2).
\end{aligned} \tag{4.2}
$$

From the asymptotic properties of $n^{-1}\mathrm{AIC_W}$ and $n^{-1}\mathrm{TIC_W^{(\mathit{j})}}(j=1,2)$ given earlier we have, $\kappa_{g1}\{n^{-1}\mathrm{AIC_W} + 2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)\} = O(n^{-1})$ under model misspecification and $\kappa_{f1}\{n^{-1}\mathrm{AIC_W} + 2\mathrm{E}_f(\hat{\bar{l}}_{\mathrm{W}}^*)\} = O(n^{-2})$ under correct model specification with $\alpha_{\mathrm{W1}}^{(\mathrm{A})*} = 0$ while $\kappa_{g1}\{n^{-1}\mathrm{TIC_W^{(\mathit{j})}} + 2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)\} = O(n^{-2})$ with $\alpha_{\mathrm{W1}}^{(\mathrm{T}j)*} = 0$ $(j=1,\,2)$ under possible model misspecification. Note that the asterisk in e.g., $\alpha_{\mathrm{W1}}^{(\mathrm{A})*}$ indicates that the corresponding cumulant is $\kappa_{f1}\{n^{-1}\mathrm{AIC_W} + 2\mathrm{E}_f(\hat{\bar{l}}_{\mathrm{W}}^*)\}$. Other asymptotic cumulants for $n^{-1}\mathrm{AIC_W} + 2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)$ and $n^{-1}\mathrm{TIC_W^{(\mathit{j})}} + 2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)$ using the notations $\alpha_{\mathrm{W}k}^{(\mathrm{A})*}$ and $\alpha_{\mathrm{W}k}^{(\mathrm{T}j)*}$ $(k=\Delta 1,\ 2,\ \Delta 2,\ 3,\ 4)$, respectively, are defined similarly to (4.2).

Recall that $n^{-1}\mathrm{AIC_W} = -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}2q$ (see (3.1)) with the corresponding symbolic expressions of the asymptotic expansions of $n^{-1}\mathrm{TIC_W^{(\mathit{j})}}(j=1,2)$ given by (3.4) and (3.5). Then, for the asymptotic cumulants of (4.2), we expand the main term $-2\hat{\bar{l}}_{\mathrm{W}}$ common to $n^{-1}\mathrm{AIC_W}$ and $n^{-1}\mathrm{TIC_W^{(\mathit{j})}}(j=1,2)$:

$$
\begin{aligned}
-2\hat{\bar{l}}_{\mathrm{W}} &= -2(\bar{l}_0)_{O_p(1)} - 2\sum_{j=1}^{4}\frac{1}{j!}\frac{\partial^j \bar{l}}{(\partial\boldsymbol{\theta}_0')^{<j>}}(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)^{<j>} + O_p(n^{-5/2}) \\
&= -2(\bar{l}_0)_{O_p(1)} - 2\sum_{j=1}^{4}\frac{1}{j!}\frac{\partial^j \bar{l}}{(\partial\boldsymbol{\theta}_0')^{<j>}}\left\{-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{k=1}^{3}\boldsymbol{\Lambda}^{(k)}\mathbf{l}_0^{(k)} + n^{-1}(\mathbf{l}_0^{(\mathrm{W})})_{O_p(n^{-1/2})}\right\}^{<j>} \\
&\quad + O_p(n^{-5/2})
\end{aligned} \tag{4.3}
$$

$$
\begin{aligned}
&= -2(\bar{l}_0)_{O_p(1)} - 2\left(\frac{\partial \bar{l}}{\partial\boldsymbol{\theta}_0'}\right)_{O_p(n^{-1/2})}\left\{-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{k=1}^{3}\boldsymbol{\Lambda}^{(k)}\mathbf{l}_0^{(k)} + n^{-1}(\mathbf{l}_0^{(\mathrm{W})})_{O_p(n^{-1/2})}\right\} \\
&\quad - \left\{\frac{\partial^2 \bar{l}}{(\partial\boldsymbol{\theta}_0')^{<2>}}\right\}_{O_p(1)}\left\{-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{k=1}^{3}\boldsymbol{\Lambda}^{(k)}\mathbf{l}_0^{(k)} + n^{-1}(\mathbf{l}_0^{(\mathrm{W})})_{O_p(n^{-1/2})}\right\}^{<2>} \\
&\quad - \frac{1}{3}\left\{\frac{\partial^3 \bar{l}}{(\partial\boldsymbol{\theta}_0')^{<3>}}\right\}_{O_p(1)}\left(-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{k=1}^{2}\boldsymbol{\Lambda}^{(k)}\mathbf{l}_0^{(k)}\right)^{<3>} \\
&\quad - \frac{1}{12}\mathrm{E}_g\left\{\frac{\partial^4 \bar{l}}{(\partial\boldsymbol{\theta}_0')^{<4>}}\right\}(\boldsymbol{\Lambda}^{(1)}\mathbf{l}_0^{(1)})^{<4>} + O_p(n^{-5/2}),
\end{aligned}
$$

which gives

$$
\begin{aligned}
&-2\hat{\bar{l}}_{\mathrm{W}} = -2(\bar{l}_0^*)_{O(1)} + \sum_{j=1}^{4}(\bar{l}_{\mathrm{ML}}^{(j)})_{O_p(n^{-j/2})} - (n^{-2}\mathbf{q}_0^{*}{}'\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^*)_{O(n^{-2})} + O_p(n^{-5/2}) \\
&(\bar{l}_{\mathrm{W}}^{(j)} = \bar{l}_{\mathrm{ML}}^{(j)},\ j = 1,...,4),
\end{aligned} \tag{4.4}
$$

where the derivation and actual expressions of $(\bar{l}_{\mathrm{ML}}^{(j)})_{O_p(n^{-j/2})}$ $(j = 1,...,4)$ are given in Subsection A8 of the appendix.

The last parenthetical results $\bar{l}_{\mathrm{W}}^{(j)} = \bar{l}_{\mathrm{ML}}^{(j)}$ $(j = 1,...,4)$ indicate that $-2\hat{\bar{l}}_{\mathrm{W}}$ is equal to $-2\hat{\bar{l}}_{\mathrm{ML}}$ up to order $O_p(n^{-3/2})$. The remaining two terms of order $O(n^{-2})$ and $O_p(n^{-2})$ are relevant only to $\alpha_{\mathrm{W}\Delta 1}^{(\mathrm{A})}$ and $\alpha_{\mathrm{W}\Delta 1}^{(\mathrm{T}j)}$ $(j = 1, 2)$ in (4.2).

Noting that $n^{-1}\mathrm{AIC}_{\mathrm{W}} = -2\hat{\bar{l}}_{\mathrm{W}} + n^{-1}2q$, (4.4) gives

Theorem 4. *Under possible model misspecification and regularity conditions for (4.2), the asymptotic cumulants of* $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ *up to the fourth order with the higher-order asymptotic bias and variance are given as follows*:

$$
\begin{aligned}
&\kappa_{g1}(n^{-1}\mathrm{AIC_W}) \\
&= -2(\overline{l}_0^*)_{O(1)} + n^{-1}\{n\mathrm{E}_g(\overline{l}_{\mathrm{ML}}^{(2)}) + 2q\}_{O(1)} + n^{-2}\{n^2\mathrm{E}_g(\overline{l}_{\mathrm{ML}}^{(3)} + \overline{l}_{\mathrm{ML}}^{(4)}) - \mathbf{q}_0^{*}{}'\mathbf{\Lambda}^{-1}\mathbf{q}_0^*\}_{O(1)} \\
&\quad + O(n^{-3}) \\
&= -2\overline{l}_0^* + n^{-1}\{\mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{\Gamma}) + 2q\} + n^{-2}\{n^2\mathrm{E}_g(\overline{l}_{\mathrm{ML}}^{(3)} + \overline{l}_{\mathrm{ML}}^{(4)}) - \mathbf{q}_0^{*}{}'\mathbf{\Lambda}^{-1}\mathbf{q}_0^*\} + O(n^{-3}) \qquad (4.5) \\
&\equiv -2\overline{l}_0^* + n^{-1}\alpha_{\mathrm{ML1}}^{(\mathrm{A})} + n^{-2}\alpha_{\mathrm{W\Delta1}}^{(\mathrm{A})} + O(n^{-3}) \\
&(\alpha_{\mathrm{W1}}^{(\mathrm{A})} = \alpha_{\mathrm{ML1}}^{(\mathrm{A})} = \mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{\Gamma}) + 2q),
\end{aligned}
$$

$$
\begin{aligned}
\kappa_{g2}(n^{-1}\mathrm{AIC_W}) &= n^{-1}[n\mathrm{E}_g\{(\overline{l}_{\mathrm{ML}}^{(1)})^2\}]_{O(1)} + n^{-2}[2n^2\mathrm{E}_g(\overline{l}_{\mathrm{ML}}^{(1)}\overline{l}_{\mathrm{ML}}^{(2)}) + 2n^2\mathrm{E}_g(\overline{l}_{\mathrm{ML}}^{(1)}\overline{l}_{\mathrm{ML}}^{(3)}) \\
&\quad + n^2\mathrm{E}_g\{(\overline{l}_{\mathrm{ML}}^{(2)})^2\} - (\alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q)^2] + O(n^{-3})
\end{aligned}
$$

$$
\equiv n^{-1}\alpha_{\mathrm{ML2}}^{(\mathrm{A})} + n^{-2}\alpha_{\mathrm{ML\Delta2}}^{(\mathrm{A})} + O(n^{-3}) \quad (n\mathrm{E}_g(\overline{l}_{\mathrm{ML}}^{(2)}) = \alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q = \mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{\Gamma}))
$$

$$
(\alpha_{\mathrm{W2}}^{(\mathrm{A})} = \alpha_{\mathrm{ML2}}^{(\mathrm{A})} = n\mathrm{E}_g\{(\overline{l}_{\mathrm{ML}}^{(1)})^2\} = 4\mathrm{E}_g\{(l_j - \overline{l}_0^*)^2\} = 4\,\mathrm{var}_g(l_j),\ \alpha_{\mathrm{W\Delta2}}^{(\mathrm{A})} = \alpha_{\mathrm{ML\Delta2}}^{(\mathrm{A})}),
$$

$$
\begin{aligned}
\kappa_{g3}(n^{-1}\mathrm{AIC_W}) &= n^{-2}[\, n^2\mathrm{E}_g\{(\overline{l}_{\mathrm{ML}}^{(1)})^3\} + 3n^2\mathrm{E}_g\{(\overline{l}_{\mathrm{ML}}^{(1)})^2\overline{l}_{\mathrm{ML}}^{(2)}\} - 3n\mathrm{E}_g(\overline{l}_{\mathrm{ML}}^{(2)})\alpha_{\mathrm{ML2}}^{(\mathrm{A})}\,] + O(n^{-3}) \\
&\equiv n^{-2}\alpha_{\mathrm{ML3}}^{(\mathrm{A})} + O(n^{-3}) \ (\alpha_{\mathrm{W3}}^{(\mathrm{A})} = \alpha_{\mathrm{ML3}}^{(\mathrm{A})}),
\end{aligned}
$$

$$
\begin{aligned}
&\kappa_{g4}(n^{-1}\mathrm{AIC_W}) = \mathrm{E}_g[\{n^{-1}\mathrm{AIC_W} - \mathrm{E}_g(n^{-1}\mathrm{AIC_W})\}^4] - 3\{n^{-1}\alpha_{\mathrm{ML2}}^{(\mathrm{A})} + n^{-2}\alpha_{\mathrm{ML\Delta2}}^{(\mathrm{A})}\}^2 + O(n^{-4}) \\
&= \mathrm{E}_g\{(n^{-1}\mathrm{AIC_W} + 2\overline{l}_0^*)^4\} + n^{-3}[-4(\alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q)\{\alpha_{\mathrm{ML3}}^{(\mathrm{A})} + 3(\alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q)\alpha_{\mathrm{ML2}}^{(\mathrm{A})}\} \\
&\quad + 6(\alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q)^2\alpha_{\mathrm{ML2}}^{(\mathrm{A})}] - 3n^{-2}(\alpha_{\mathrm{ML2}}^{(\mathrm{A})})^2 - 6n^{-3}\alpha_{\mathrm{ML2}}^{(\mathrm{A})}\alpha_{\mathrm{ML\Delta2}}^{(\mathrm{A})} + O(n^{-4}) \\
&= \mathrm{E}_g\{(n^{-1}\mathrm{AIC_W} + 2\overline{l}_0^*)^4\} - 3n^{-2}(\alpha_{\mathrm{ML2}}^{(\mathrm{A})})^2 \\
&\quad - n^{-3}\{4(\alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q)\alpha_{\mathrm{ML3}}^{(\mathrm{A})} + 6\alpha_{\mathrm{ML2}}^{(\mathrm{A})}\alpha_{\mathrm{ML\Delta2}}^{(\mathrm{A})} + 6\alpha_{\mathrm{ML2}}^{(\mathrm{A})}(\alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q)^2\} + O(n^{-4}) \\
&= n^{-3}[\, n^3\{\kappa_{g4}(\overline{l}_{\mathrm{ML}}^{(1)})\}_{O(n^{-3})} + 4n^3\mathrm{E}_g\{(\overline{l}_{\mathrm{ML}}^{(1)})^3\overline{l}_{\mathrm{ML}}^{(2)}\} + 6n^3\mathrm{E}_g\{(\overline{l}_{\mathrm{ML}}^{(1)})^2(\overline{l}_{\mathrm{ML}}^{(2)})^2\} \\
&\quad + 4n^3\mathrm{E}_g\{(\overline{l}_{\mathrm{ML}}^{(1)})^3\overline{l}_{\mathrm{ML}}^{(3)}\} - 4(\alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q)\alpha_{\mathrm{ML3}}^{(\mathrm{A})} - 6\alpha_{\mathrm{ML2}}^{(\mathrm{A})}\alpha_{\mathrm{ML\Delta2}}^{(\mathrm{A})} - 6\alpha_{\mathrm{ML2}}^{(\mathrm{A})}(\alpha_{\mathrm{ML1}}^{(\mathrm{A})} - 2q)^2\,] + O(n^{-4}) \\
&\equiv n^{-3}\alpha_{\mathrm{ML4}}^{(\mathrm{A})} + O(n^{-4}) \ (\alpha_{\mathrm{W4}}^{(\mathrm{A})} = \alpha_{\mathrm{ML4}}^{(\mathrm{A})}).
\end{aligned}
$$

In the case of the canonical parameters under correct model specification as in Corollary 1, using (4.4) (see also (2.12)), the asymptotic biases become as follows:

$$
\begin{aligned}
&\kappa_{f1}(n^{-1}\mathrm{AIC_W})\\
&=-2\overline{l}_0^{\,*}+n^{-1}\{\mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{\Gamma})+2q\}+n^{-2}\{n^2\mathrm{E}_f(\overline{l}_{\mathrm{ML}}^{(3)}+\overline{l}_{\mathrm{ML}}^{(4)})-\mathbf{q}_0^*{}'\mathbf{\Lambda}^{-1}\mathbf{q}_0^*\}+O(n^{-3})\\
&=-2\overline{l}_0^{\,*}+n^{-1}q\\
&\quad+n^{-2}\left[-\frac{1}{6}\boldsymbol{\kappa}_{f3}{}'(\tilde{\mathbf{x}}^*)\boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*)-\frac{1}{4}\boldsymbol{\kappa}_{f3}{}'(\tilde{\mathbf{x}}^*)[\mathbf{I}_{(q)}\otimes\{\mathrm{vec}(\mathbf{I}_{(q)})\mathrm{vec}'(\mathbf{I}_{(q)})\}]\boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*)\right. \qquad (4.7)\\
&\qquad\qquad\left.+\frac{1}{4}\boldsymbol{\kappa}_{f4}{}'(\tilde{\mathbf{x}}^*)\mathrm{vec}(\mathbf{I}_{(q^2)})+\mathbf{q}_0^*{}'\mathbf{I}_0^{-1}\mathbf{q}_0^*\right]+O(n^{-3})\\
&=-2\overline{l}_0^{\,*}+n^{-1}\alpha_{\mathrm{ML1}}^{(\mathrm{A})}+n^{-2}\alpha_{\mathrm{W\Delta1}}^{(\mathrm{A})}+O(n^{-3}),
\end{aligned}
$$

where $-1/6$ and +1/4 come from $(1/3)-(2/4)$ and $(1/12)\times 3$, respectively (see (A3.2) and (A8.3)).

The results for $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ corresponding to Theorem 4 are given from (3.4) and (3.5) of Definition 1.

Theorem 5. *Under possible model misspecification and regularity conditions for (4.2), the asymptotic cumulants of* $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ *up to the fourth order with the higher-order asymptotic bias and variance are given as follows*:

$$
\begin{aligned}
&\kappa_{g1}(n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)})\\
&=-2(\overline{l}_0^{\,*})_{O(1)}+n^{-1}\{\alpha_{\mathrm{ML1}}^{(\mathrm{A})}-2q+2\mathrm{tr}(-\mathbf{\Lambda}^{-1}\mathbf{\Gamma})\}+n^{-2}\{\alpha_{\mathrm{W\Delta1}}^{(\mathrm{A})}+2n^2\mathrm{E}_g(\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T}j)})\}+O(n^{-3})\\
&\equiv-2(\overline{l}_0^{\,*})_{O(1)}+n^{-1}\alpha_{\mathrm{ML1}}^{(\mathrm{T}\bullet)}+n^{-2}\alpha_{\mathrm{W\Delta1}}^{(\mathrm{T}j)}+O(n^{-3}) \qquad (4.7)\\
&(\alpha_{\mathrm{W1}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{ML1}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{ML1}}^{(\mathrm{A})}-2q+2\mathrm{tr}(-\mathbf{\Lambda}^{-1}\mathbf{\Gamma})=\mathrm{tr}(-\mathbf{\Lambda}^{-1}\mathbf{\Gamma})),\\
&\kappa_{g2}(n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)})=n^{-1}\alpha_{\mathrm{ML2}}^{(\mathrm{A})}+n^{-2}\{\alpha_{\mathrm{ML\Delta2}}^{(\mathrm{A})}+4n\mathrm{E}_g(\overline{l}_{\mathrm{ML}}^{(1)}\mathrm{tr}_{\Delta}^{(\mathrm{T}j)})\}+O(n^{-3})\\
&\qquad\equiv n^{-1}\alpha_{\mathrm{ML2}}^{(\mathrm{T}\bullet)}+n^{-2}\alpha_{\mathrm{W\Delta2}}^{(\mathrm{T}j)}+O(n^{-3})\quad(\alpha_{\mathrm{W2}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{ML2}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{W2}}^{(\mathrm{A})}=\alpha_{\mathrm{ML2}}^{(\mathrm{A})}),\\
&\kappa_{g3}(n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)})=\kappa_{g3}(n^{-1}\mathrm{AIC_{ML}})+O(n^{-3})\\
&\qquad=n^{-2}\alpha_{\mathrm{ML3}}^{(\mathrm{A})}+O(n^{-3})\quad(\alpha_{\mathrm{W3}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{ML3}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{W3}}^{(\mathrm{A})}=\alpha_{\mathrm{ML3}}^{(\mathrm{A})}),\\
&\kappa_{g4}(n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)})=\kappa_{g4}(n^{-1}\mathrm{AIC_{ML}})+O(n^{-3})\\
&\qquad=n^{-3}\alpha_{\mathrm{ML4}}^{(\mathrm{A})}+O(n^{-4})\quad(\alpha_{\mathrm{W4}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{ML4}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{W4}}^{(\mathrm{A})}=\alpha_{\mathrm{ML4}}^{(\mathrm{A})})\quad(j=1,2),
\end{aligned}
$$

where the superscript $(\mathrm{T}\cdot)$ *indicates a result common to* $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$.

In (4.7), $\alpha_{\mathrm{W3}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{ML3}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{W3}}^{(\mathrm{A})}=\alpha_{\mathrm{ML3}}^{(\mathrm{A})}$ stems from the property that the third asymptotic cumulants of $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ are given only by $\overline{l}_{\mathrm{W}}^{(1)}(=\overline{l}_{\mathrm{ML}}^{(1)})$ and $\overline{l}_{\mathrm{W}}^{(2)}(=\overline{l}_{\mathrm{ML}}^{(2)})$ of

$-2\hat{\bar{l}}_{\mathrm{W}}$ in (3.4) and (3.5) of Definition 1 (see the last parenthetical result of (4.4)) with the fixed term $\mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{\Gamma})$ in (3.4) and (3.5) being irrelevant to the cumulants except that of the first order. The additional stochastic terms $2n^{-1}(\mathrm{tr}_{\Delta}^{(\mathrm{T}\,j)})_{O_p(n^{-1/2})}$ for $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ in (3.4) and (3.5) with $\bar{l}_{\mathrm{W}}^{(j)}(=\bar{l}_{\mathrm{ML}}^{(j)})\ (j=1,2,3)$ in the expansion of $-2\hat{\bar{l}}_{\mathrm{W}}$ common to $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ and $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$ contribute to the higher-order added asymptotic variance $n^{-2}\alpha_{\mathrm{W}\Delta 2}^{(\mathrm{T}\,j)}$ $(j=1,\,2)$ in (4.7). However, the contributions by $2n^{-1}\mathrm{tr}_{\Delta}^{(\mathrm{T}\,j)}$ $(j=1,\,2)$ are canceled when we derive the (asymptotic) fourth cumulants, giving $\alpha_{\mathrm{W4}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{ML4}}^{(\mathrm{T}\bullet)}=\alpha_{\mathrm{W4}}^{(\mathrm{A})}=\alpha_{\mathrm{ML4}}^{(\mathrm{A})}$ in (4.7).

For interval estimation of the population quantity $-2\bar{l}_0^*$ as well as $-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)$ by $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ and $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$, the following studentized estimators are defined:

$$\begin{aligned} t_{\mathrm{W}}^{(\mathrm{A})} &\equiv \frac{n^{1/2}(n^{-1}\mathrm{AIC}_{\mathrm{W}}+2\bar{l}_0^*)}{(\hat{v}_{\mathrm{W}}^{(\mathrm{A})})^{1/2}},\ t_{\mathrm{W}}^{(\mathrm{T}\,j)} \equiv \frac{n^{1/2}(n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}+2\bar{l}_0^*)}{(\hat{v}_{\mathrm{W}}^{(\mathrm{A})})^{1/2}}\ (j=1,\,2), \\ t_{\mathrm{W}}^{(\mathrm{A})*} &\equiv \frac{n^{1/2}\{n^{-1}\mathrm{AIC}_{\mathrm{W}}+2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)\}}{(\hat{v}_{\mathrm{W}}^{(\mathrm{A})})^{1/2}},\ t_{\mathrm{W}}^{(\mathrm{T}\,j)*} \equiv \frac{n^{1/2}\{n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}+2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)\}}{(\hat{v}_{\mathrm{W}}^{(\mathrm{A})})^{1/2}}\ (j=1,\,2), \end{aligned} \tag{4.8}$$

where $t_{\mathrm{W}}^{(\mathrm{A})}$ and $t_{\mathrm{W}}^{(\mathrm{T}\,j)}\,(j=1,\,2)$ are for estimation of $-2\bar{l}_0^*$ while $t_{\mathrm{W}}^{(\mathrm{A})*}$ and $t_{\mathrm{W}}^{(\mathrm{T}\,j)*}\,(j=1,\,2)$ are for $-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}}^*)$ under possible model misspecification; $n^{-1}\hat{v}_{\mathrm{W}}^{(\mathrm{A})}$ is the robust estimator of the asymptotic variance $n^{-1}\alpha_{\mathrm{ML2}}^{(\mathrm{A})}$ common to $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ and $n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(j)}(j=1,2)$:

$$\hat{v}_{\mathrm{W}}^{(\mathrm{A})} \equiv 4(n-1)^{-1}\sum_{j=1}^{n}(\hat{l}_{\mathrm{W}j}-\hat{\bar{l}}_{\mathrm{W}})^2 = O_{p(1)} \tag{4.9}$$

with $\hat{l}_{\mathrm{W}j} \equiv l_j\,|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{W}}}$ $(j=1,\ldots,n)$ and $\hat{\bar{l}}_{\mathrm{W}} = n^{-1}\sum_{j=1}^{n}\hat{l}_{\mathrm{W}j}$ (for l_j see (2.1)).

Under correct model specification, in many cases $\alpha_{\mathrm{ML2}}^{(\mathrm{A})}$ may be explicitly obtained as a function of $\boldsymbol{\theta}_0$. However, since this result depends on a model employed, the four versions of robust studentization in (4.8) are considered in this section. Define the stochastic quantity using $\boldsymbol{\theta}_0$ in place of $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ in (4.9):

$$v_0^{(\mathrm{A})} \equiv 4(n-1)^{-1}\sum_{j=1}^{n}(l_{0j}-\overline{l}_0)^2 = O_{p(1)} \tag{4.11}$$

with $l_{0j} \equiv l_j \,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ $(j=1,...,n)$ and $\overline{l}_0 = n^{-1}\sum_{j=1}^{n} l_{0j}$. Then, $v_0^{(\mathrm{A})}$ is an exactly unbiased robust estimator of $\alpha_{\mathrm{ML2}}^{(\mathrm{A})}$ with $\mathrm{E}_g(v_0^{(\mathrm{A})}) = \alpha_{\mathrm{ML2}}^{(\mathrm{A})}$ though $v_0^{(\mathrm{A})}$ usually includes the unknown $\boldsymbol{\theta}_0$. Generally, the estimator $\hat{v}_{\mathrm{W}}^{(\mathrm{A})}$ is not an unbiased one but is a consistent estimator of $\alpha_{\mathrm{ML2}}^{(\mathrm{A})}$.

Under possible model misspecification, assume that the following hold with the asymptotic cumulants, whose factors of order $O(1)$ are $\alpha_{(t)\mathrm{W}k}^{(\mathrm{A})}$ $(k = 1,\ 2,\ \Delta 2,\ 3,\ 4)$ for $t_{\mathrm{W}}^{(\mathrm{A})}$:

$$\begin{aligned}
&\kappa_{g1}(t_{\mathrm{W}}^{(\mathrm{A})}) = n^{-1/2}\alpha_{(t)\mathrm{W}1}^{(\mathrm{A})} + O(n^{-3/2}),\\
&\kappa_{g2}(t_{\mathrm{W}}^{(\mathrm{A})}) = 1 + n^{-1}\alpha_{(t)\mathrm{W}\Delta 2}^{(\mathrm{A})} + O(n^{-2}) \ \ (\alpha_{(t)\mathrm{W}2}^{(\mathrm{A})} = 1),\\
&\kappa_{g3}(t_{\mathrm{W}}^{(\mathrm{A})}) = n^{-1/2}\alpha_{(t)\mathrm{W}3}^{(\mathrm{A})} + O(n^{-3/2}),\\
&\kappa_{g4}(t_{\mathrm{W}}^{(\mathrm{A})}) = n^{-1}\alpha_{(t)\mathrm{W}4}^{(\mathrm{A})} + O(n^{-2}).
\end{aligned} \tag{4.12}$$

Similarly, $\alpha_{(t)\mathrm{W}k}^{(\mathrm{T}j)}$ for $t_{\mathrm{W}}^{(\mathrm{T}j)}$, $\alpha_{(t)\mathrm{W}k}^{(\mathrm{A})*}$ for $t_{\mathrm{W}}^{(\mathrm{A})*}$ and $\alpha_{(t)\mathrm{W}k}^{(\mathrm{T}j)*}$ for $t_{\mathrm{W}}^{(\mathrm{T}j)*}(j=1,2)$, $(k = 1,\ 2,\ \Delta 2,\ 3,\ 4)$ are defined. These asymptotic cumulants are obtained. However, since their derivations and results are relatively involved, they are shown in the first supplement to this paper (Ogasawara, 2016a).

5. Examples for the asymptotic cumulants

Three examples are given in this section. Each of Examples 1 and 2 uses the MLE of a canonical parameter in the exponential family under model misspecification while Example 3 deals with the WSE of a canonical parameter in the exponential family under correct model specification. The asymptotic cumulants, obtained in Section 4, for the examples are shown in Tables 1 and 2, whose expository derivations are given in the supplements to this paper (Ogasawara, 2016a, 2016b).

Example 1: The MLE of the parameter in the exponential distribution is used when the gamma distribution with the shape parameter α being unequal to 1 holds. That is, the density

$$f(x^* = x \mid \lambda_0) = \lambda_0 \exp(-\lambda_0 x) \ (x > 0) \tag{5.1}$$

is used with $\theta_0 = \lambda_0$ when the true distribution is

$$g(x^* = x \mid \lambda_1, \alpha) = x^{\alpha-1}\lambda_1^{\alpha} \exp(-\lambda_1 x) / \Gamma(\alpha) \ (x > 0, \ \alpha \neq 1) \tag{5.2}$$

with $\boldsymbol{\zeta}_0 = (\lambda_1, \alpha)'$ and $\Gamma(\cdot)$ being the gamma function. By assumption $\alpha = 1$ is excluded. However, when $\alpha = 1$ in (5.2), this reduces to (5.1). The MLE of λ_0 is $1/\bar{x}$, where $\bar{x}$ is the sample mean of the observable variable. This gives the population λ_0 under model misspecification as

$$\lambda_0 = 1/\mathrm{E}_g(\bar{x}) = \lambda_1 / \alpha \, . \tag{5.3}$$

Example 2: The MLE of the mean in the univariate normal distribution with known variance σ^2 is used when the true distribution is non-normal with known variance σ^2. That is,

$$f(x^* = x \mid \mu_0, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu_0)^2}{2\sigma^2} \right\} \tag{5.4}$$

with $\hat{\theta}_{\mathrm{ML}} = \hat{\mu}_{\mathrm{ML}} = \bar{x}$. In this example,

$$\begin{aligned} \bar{l}_0^* &= \mathrm{E}_g(l_{0j}) = \mathrm{E}_f(l_{0j}) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}, \\ \mathrm{E}_g(\hat{\theta}_{\mathrm{ML}}) &= \mathrm{E}_f(\hat{\theta}_{\mathrm{ML}}) = \mu_0, \, n\,\mathrm{var}_g(\hat{\theta}_{\mathrm{ML}}) = n\,\mathrm{var}_f(\hat{\theta}_{\mathrm{ML}}) = \sigma^2. \end{aligned} \tag{5.5}$$

However, $\mathrm{var}_g(l_{0j}) = \frac{1}{4}\left\{ \kappa_{g4}\left(\frac{x^* - \mu_0}{\sigma} \right) + 2 \right\}$ under non-normality with $\kappa_{g4}(\cdot) \neq 0$ is not equal to $\mathrm{var}_f(l_{0j}) = 1/2$ under normality.

Example 3: The WSE of the logit in the Bernoulli distribution is used under correct model specification. That is,

$$\Pr(x^* = x \mid \theta_0) = \pi_0^x (1-\pi_0)^{1-x} \ (x = 0, 1), \, \pi_0 = \frac{1}{1+\exp(-\theta_0)} \, . \tag{5.6}$$

While $\hat{\theta}_{\mathrm{ML}} = \log\dfrac{\bar{x}}{1-\bar{x}}$ $(\bar{x} \neq 0, 1)$, where $\bar{x}$ is the usual sample proportion, $\hat{\theta}_{\mathrm{W}}$ in Example 3 is defined as the solution of θ which maximizes

$$\left\{\prod_{j=1}^{n} \pi^{x_j}(1-\pi)^{1-x_j}\right\}\{\pi(1-\pi)\}^{a/2} \text{ with } \pi = \frac{1}{1+\exp(-\theta)}, \tag{5.7}$$

where a is the sum of equal pseudocounts for two categories (do not confuse π with the circular constant used earlier). The solution is given when $\theta = \hat{\theta}_{\mathrm{W}} \equiv \log\dfrac{\bar{x}+n^{-1}0.5a}{1-\bar{x}+n^{-1}0.5a}$.

In the footnotes of the tables, general results associated with the tables are given (for derivation, see also Ogasawara, 2016a, b). In Examples 1 and 2, the results do not depend on scales since l (log-likelihood) except a fixed term is scale-free in these examples. Although $\alpha \neq 1$ is assumed in Example 1, $\alpha = 1$ gives the corresponding results under correct model specification. Note that in the latter case with $\alpha = 1$, all the results in Example 1 are given by fixed values. Under correct model specification, the bias-corrected $n^{-1}\mathrm{AIC}_{\mathrm{ML}}$ up to order $O(n^{-2})$, denoted by $n^{-1}\mathrm{CAIC}_{\mathrm{ML}}$, is given by as simple as

$$n^{-1}\mathrm{CAIC}_{\mathrm{ML}} = -2\hat{\bar{l}}_{\mathrm{ML}} + n^{-1}2 + n^{-2}2\,. \tag{5.8}$$

Similarly, under normality, the results for Example 2 in the tables are given only by fixed values, where κ_j's $(\equiv \kappa_{gj}\{(x^* - \mu_0)/\sigma\}$'s$)$ $(j \neq 2)$ vanish. Note also that $n^{-1}\mathrm{AIC}_{\mathrm{ML}} (= n^{-1}\mathrm{TIC}_{\mathrm{ML}}^{(j)},\ j = 1,2)$ in Example 2 is exactly unbiased even under non-normality (see (5.5) and Corollary 3). In Example 3, the results when $\hat{\theta}_{\mathrm{ML}}$ is used, are given by $a = 0$.

In Example 3, from Table 1 we have

Corollary 5. *Under the assumption that the Bernoulli distribution holds,* $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ *for estimation of* $-2\mathrm{E}_f(\hat{\bar{l}}_{\mathrm{W}}^*)$ *using* $\hat{\theta}_{\mathrm{W}}$ *as the weighted score estimator of the logit with the total number a of equal pseudocounts for two categories gives no asymptotic bias up to order* $O(n^{-2})$ *when a* = 1.

For the derivation of the higher-order asymptotic bias, see Ogasawara (2016b, Subsection S6.1). It is of interest to see that when a = 1, $\hat{\theta}_{\mathrm{W}}$ is also unbiased up to order $O(n^{-1})$ (see e.g., Ogasawara, 2015a, Section 6). On the other hand, for estimation of $-2\overline{l}_0^*$ the corresponding bias of $n^{-1}\mathrm{AIC}_{\mathrm{W}}$ up to order $O(n^{-2})$ is $n^{-1}+n^{-2}\{(1/6)(1-\overline{i}_0^{-1})+(a^2/4)(1-2\pi_0)^2\overline{i}_0^{-1}\}$, which is minimized when a = 0 and $\pi_0 \neq 1/2$ while a is irrelevant to the asymptotic bias when $\pi_0 = 1/2$.

Insert Tables 1 to 10 about here.

6. Simulation for model selection

Since in practice information criteria are used typically for model selection, simulations using the $n^{-1}\mathrm{AIC}\ (= n^{-1}\mathrm{AIC}_{\mathrm{ML}})$ and the bias corrected $n^{-1}\mathrm{AIC}$ i.e., $n^{-1}\mathrm{AIC} - n^{-2}\hat{c}_1$ denoted by $n^{-1}\mathrm{CAIC}\ (= n^{-1}\mathrm{CAIC}_{\mathrm{ML}})$ for selecting regressors are carried out in this section when a regression model holds under canonical parametrization. Four types of regression, logistic, Poisson, negative binomial and gamma regression are used, where a canonical parameter has a form of the linear combination of p regressors including an intercept when it is used.

Bias corrections of the AICs in logistic and Poisson regression are given by Yanagihara, Sekiguchi and Fujikoshi (2003) and Kamo, Yanagihara and Satoh (2013), respectively by different methods and expressions from those in this section. Since the unified result of bias correction for the $n^{-1}\mathrm{AIC}$ under canonical parametrization in the exponential family was derived earlier (see Corollary 2, (A3.1) and (A3.2)), logistic regression and Poisson regression are also deal with as special cases in this section. To the author’s knowledge, the results of bias corrections of $n^{-1}\mathrm{AICs}$ in negative binomial and gamma regression are new.

For computation, (A3.1) under correct model specification is written as

$$
\begin{aligned}
c_1 &= -\mathrm{vec}'(\mathbf{I}_0^{-1})\mathbf{J}_0^{(3)}{}'\mathbf{I}_0^{-1}\mathbf{J}_0^{(3)}\mathrm{vec}(\mathbf{J}_0^{(3)}) - \mathrm{vec}'(\mathbf{J}_0^{(3)})(\mathbf{I}_0^{-1})^{<3>}\mathrm{vec}(\mathbf{J}_0^{(3)}) \\
&\quad - \mathrm{vec}'(\mathbf{J}_0^{(4)})\mathrm{vec}\{(\mathbf{I}_0^{-1})^{<2>}\} \\
&= -\sum_{a,b,c,d,e,f=1}^{q} i_0^{ab}(\mathbf{J}_0^{(3)})_{(a,b,c)} i_0^{cd}(\mathbf{J}_0^{(3)})_{(d,e,f)} i_0^{ef} - \sum_{a,b,c,d,e,f=1}^{q} (\mathbf{J}_0^{(3)})_{(a,b,c)} i_0^{ad} i_0^{be} i_0^{cf} (\mathbf{J}_0^{(3)})_{(d,e,f)} \\
&\quad - \sum_{a,b,c,d=1}^{q} (\mathbf{J}_0^{(4)})_{(a,b,c,d)} i_0^{ad} i_0^{cd},
\end{aligned}
\tag{6.1}
$$

where $i_0^{ab} = (\mathbf{I}_0^{-1})_{ab}$ and q is the number of unknown parameters in a regression model. Let y_i^{**} be the dependent variable in a model under canonical parametrization. Define $\mathbf{x}_i$ be the $p \times 1$ vector for p covariates for the i-the observation (i = 1,…,n) and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)'$. Note that the expression y_i^* is retained for the usual response variable under possible non-canonical parametrization. The linear predictor using $p_0 \times 1$ vector of population regression coefficients $\boldsymbol{\beta}_0$ is assumed to be given by $\mathbf{x}_i{}'\boldsymbol{\beta}_0$ for the i-th observation. Then, $\boldsymbol{\Lambda}(= -\mathbf{I}_0)$, $\mathbf{J}_0^{(3)}$ and $\mathbf{J}_0^{(4)}$ are derived by the following unified expression when only $\boldsymbol{\beta}_0$ is unknown:

$$
\begin{aligned}
\boldsymbol{\Lambda} = -\mathbf{I}_0 &= \sum_{i=1}^{n} \frac{\partial^2 \overline{l}}{\partial\boldsymbol{\beta}_0 \partial\boldsymbol{\beta}_0{}'} = -\sum_{i=1}^{n} \kappa_{f2}(y_i^{**})\mathbf{x}_i\mathbf{x}_i{}', \\
\mathrm{vec}(\mathbf{J}_0^{(3)}) &= \sum_{i=1}^{n} \frac{\partial^3 \overline{l}}{(\partial\boldsymbol{\beta}_0)^{<3>}} = -\sum_{i=1}^{n} \kappa_{f3}(y_i^{**})\mathbf{x}_i^{<3>}, \\
\mathrm{vec}(\mathbf{J}_0^{(4)}) &= \sum_{i=1}^{n} \frac{\partial^4 \overline{l}}{(\partial\boldsymbol{\beta}_0)^{<4>}} = -\sum_{i=1}^{n} \kappa_{f4}(y_i^{**})\mathbf{x}_i^{<4>},
\end{aligned}
\tag{6.2}
$$

where $\kappa_{fk}(y_i^*)(k = 2, 3, 4)$ ($y_i^{**} = y_i^*$ except $y_i^{**} = -y_i^*$ in gamma regression) are shown in Table 3 under the headers of Variance, the numerator on the left-hand side of Skewness and that of Excess kurtosis, respectively without subscript i. Note that in the negative binomial and gamma distributions of Table 3, the shape parameters r and α, respectively are assumed to be given. When they are unknown, c_1 should be given from (6.1) with $q > 1$ even when only an intercept is used in a regression model. We find in the table that $-2/\alpha$ for c_1 in the gamma distribution gives (5.8) in Example 1 when $\alpha = 1$.

For clarity, the probability masses and density when the regression model with $\boldsymbol{\beta}_0$

holds under canonical parametrization are given as

logistic regression: $f(y_i^* = y_i \mid \pi_{0i}) = \pi_{0i}^{y_i}(1-\pi_{0i})^{1-y_i}$, $\pi_{0i} = 1/\{1+\exp(-\mathbf{x}_i{}'\boldsymbol{\beta}_0)\}$, (6.3)

$$y_i^{**} = y_i^*,\ \ y_i = 0, 1\ (i = 1,...,n);$$

Poisson regression: $f(y_i^* = y_i \mid \lambda_{0i}) = \lambda_{0i}^{y_i}\exp(-\lambda_{0i})/y_i!$, $\lambda_{0i} = \exp(\mathbf{x}_i{}'\boldsymbol{\beta}_0)$, (6.4)

$$y_i^{**} = y_i^*, y_i = 0,1,2,...\ (i = 1,...,n);$$

negative binomial regression:

$$f(y_i^* = y_i \mid \pi_{0i}, r_0) = \begin{pmatrix} y_i + r_0 - 1 \\ y_i \end{pmatrix} \pi_{0i}^{y_i}(1-\pi_{0i})^{r_0} = \frac{\Gamma(y_i + r_0)}{y_i!\Gamma(r_0)}\pi_{0i}^{y_i}(1-\pi_{0i})^{r_0},$$
$$0 < \pi_{0i} = \exp(\mathbf{x}_i{}'\boldsymbol{\beta}_0) < 1,\ \ r_0 > 0,\ \ y_i^{**} = y_i^*,\ \ y_i = 0,1,2,...\ (i = 1,...,n), \tag{6.5}$$

where r_0 is the population shape parameter or the given number of the occurrences of an event, when r_0 is a positive integer (the Pascal distribution), whose probability for an occurrence is $1-\pi_{0i}$ with π_{0i} being the probability for the complimentary event whose number of occurrences y_i, when $\mathbf{x}_i$ is given, is of primary interest;

and gamma regression: $f(y_i^* = y_i \mid \lambda_{0i}, \alpha_0) = y_i^{\alpha_0 - 1}\lambda_{0i}^{\alpha_0}\exp(-\lambda_{0i}y_i)/\Gamma(\alpha_0)$, (6.4)

$$\lambda_{0i} = \mathbf{x}_i{}'\boldsymbol{\beta}_0 > 0,\ \alpha_0 > 0,\ y_i^{**} = -y_i^*,\ y_i > 0\,(i = 1,...,n),$$

where α_0 is the population shape parameter, which gives the Arlang distribution when α_0 is a positive integer.

For computation of (6.1) in negative binomial and gamma regression when the shape parameter is unknown, the derivatives of the psi (digamma) function up to the third order (trigamma, tetragamma and pentagamma functions) are required, whose algorithm and software are available (Amos, 1983; psigamma() in R Core Team, 2015).

Note that canonical parametrization in logistic and Poison regression seems to be used most exclusively in practice although we have e.g., the probit and double exponential models in regression using the Bernoulli distribution. On the other hand, canonical parametrization in negative binomial regression is used by Hilbe (2011, Chapter 8) though other parametrizations may also be typical (see Lawless, 1987). In gamma regression, parametrizations using the mean and scale (the reciprocal of the rate parameter) are also typical especially when an event time is of primary interest rather than the rate of occurrence (see e.g., Ogasawara, 1995).

In the simulation, categorical regressors for grouping are used, where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}' \otimes \mathbf{1}_{(n/8)}, \tag{6.7}$$

$\mathbf{1}_{(n/8)}$ is the $(n/8)\times 1$ vector of 1's and n = 40, 80 and 160. Two population values $p_0 = 2$ and 3 are used, where candidate models are given by $p = 1,\ldots, p_0 + 1$. Note that the model of p = 1 and $\mathbf{X} = (1,\ldots,1)'$ is the model with an intercept only. Tables 4 to 9 show the proportions of model selection by the minimum $n^{-1}\mathrm{AIC}$ and $n^{-1}\mathrm{CAIC}(= n^{-1}\mathrm{AIC}_{\mathrm{ML}})$. In Tables 4 and 5 for logistic and Poisson regression, respectively, the mean (M) and standard deviation (SD) of the (biserial) correlations of the estimated linear predictor $\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ and y_i over i = 1,…, n are shown, where the correlation is defined as 0 when p =1. The Ms and SDs are given from 1,000 replications in a simulation. In the tables, the number of deleted cases due to non-convergence until 1,000 regular cases were obtained are also shown.

In Table 4 for logistic regression, the proportions of correct model selection (hereafter called as correct proportions) by the CAIC are greater than those by the AIC except the case of n = 40 and $p_0 = 3$ with the underscored AIC. In Table 5 for Poisson regression, the correct proportions by the CAIC are greater than or equal to those by the AIC. These results repeat similar known ones. Tables 6 and 7 give the results of negative binomial regression with r being known and unknown, respectively. In Table 7, the results of r = 4 are not shown since non-convergent cases occurred frequently. The correct proportions by the CAIC are mostly greater than those by the AIC. In Tables 8 and 9, the results for gamma regression corresponding to those in Tables 6 and 7, respectively are shown. Although many of the correct proportions by the CAIC are smaller than those by AIC when $p_0 = 3$, the difference becomes smaller or reversed when n becomes large.

It is known that in usual normal linear and Poisson regression, the AIC tends to choose overspecified models i.e., those including additional regressor(s) as well as true one(s) (see e.g., Hurvich & Tsai, 1989; Kamo et al., 2013). The results in the tables give similar tendencies. Note that the correction term $-n^{-2}\hat{c}_1$ in the $n^{-1}\mathrm{CAIC}$ penalizes models under canonical parametrization when squared (multivariate) skewness is large and the excess (multivariate) kurtosis is negative with its absolute value being large. It is found in the tables that the $n^{-1}\mathrm{CAIC}$ corrects the tendency of choosing relatively complicated models to some extent (note that c_1's in Table 10 are all negative) .

7. Simulation for interval estimation of $-2\mathrm{E}_f(\hat{\bar{l}}^*_{\mathrm{ML}})$

A simulation for interval estimation of $-2\mathrm{E}_f(\hat{\bar{l}}^*_{\mathrm{ML}})$, whose unbiased point estimator up to order $O(n^{-1})$ under correct model specification is $n^{-1}\mathrm{AIC}$, is carried out in this section as an application of the asymptotic cumulants of the studentized and non-studentized $n^{-1}\mathrm{AIC}$ in the case of the exponential distribution. Note that the asymptotic cumulants up to the forth order are given from Tables 1 and 2 when $\alpha = 1$ in Example 1, which do not depend on the rate parameter λ_0.

When the asymptotic cumulants are fixed values as in this case, it is known that under some regularity conditions the following lower endpoint of a one-sided confidence interval has the third-order accuracy as defined below (see Ogasawara, 2012, Equation (2.5)):

$$\begin{aligned} L(\alpha;n^{-3/2}) &= -2\hat{\bar{l}}_{\mathrm{ML}} - n^{-1/2}(\hat{v}^{(\mathrm{A})}_{\mathrm{ML}})^{1/2} z_\alpha - n^{-1}(\hat{v}^{(\mathrm{A})}_{\mathrm{ML}})^{1/2}\{\alpha^{(\mathrm{A})*}_{(t)\mathrm{ML1}} + (\alpha^{(\mathrm{A})*}_{(t)\mathrm{ML3}}/6)(z_\alpha^2 - 1)\} \\ &\quad - n^{-3/2}(\hat{v}^{(\mathrm{A})}_{\mathrm{ML}})^{1/2}\left\{\frac{1}{2}\alpha^{(\mathrm{A})*}_{(t)\mathrm{ML}\Delta 2} z_\alpha + (\alpha^{(\mathrm{A})*}_{(t)\mathrm{ML3}})^2\left(-\frac{z_\alpha^3}{18} + \frac{5}{36}z_\alpha\right) + \alpha^{(\mathrm{A})*}_{(t)\mathrm{ML4}}\left(\frac{z_\alpha^3}{24} - \frac{z_\alpha}{8}\right)\right\}, \end{aligned} \tag{7.1}$$

where

$$\int_{-\infty}^{z_\alpha} (1/\sqrt{2})\exp(-z^2/2)\,\mathrm{d}z = \alpha \quad \text{and} \quad \Pr\{-2\mathrm{E}_f(\hat{\bar{l}}^*_{\mathrm{ML}}) > L(\alpha;n^{-3/2})\} = \alpha + O(n^{-3/2}) \,. \tag{7.2}$$

The above results are based on Cornish-Fisher expansions. In (7.1), the value up to order $O_p(n^{-1/2})$ is an endpoint of the usual Wald confidence interval. The values up to orders $O_p(n^{-1})$ and $O_p(n^{-3/2})$ are the second- and third-order accurate confidence intervals, respectively. From Table 2 we have

$$\alpha_{(t)\mathrm{ML1}}^{(\mathrm{A})*} = 0,\ \alpha_{(t)\mathrm{ML3}}^{(\mathrm{A})*} (= \alpha_{(t)\mathrm{ML3}}^{(\mathrm{A})}) = -1,\ \alpha_{(t)\mathrm{ML}\Delta 2}^{(\mathrm{A})*} = 5.5 \ \text{ and } \ \alpha_{(t)\mathrm{ML4}}^{(\mathrm{A})*} (= \alpha_{(t)\mathrm{ML4}}^{(\mathrm{A})}) = 14 \qquad (7.3)$$

which can be used in (7.1) with $\hat{v}_{\mathrm{ML}}^{(\mathrm{A})}$ (see (4.9)). Although $\alpha_{\mathrm{ML2}}^{(\mathrm{A})*} (= \alpha_{\mathrm{ML2}}^{(\mathrm{A})}) = 4$ for the non-studentized $n^{-1}\mathrm{AIC}$ is a fixed value, robust $\hat{v}_{\mathrm{ML}}^{(\mathrm{A})}$ against possible model misspecification is used for illustration.

When $\alpha_{\mathrm{ML2}}^{(\mathrm{A})*}$ is used, the following standardized statistic is defined in the case of $n^{-1}\mathrm{AIC}$ similarly to (4.8)

$$z_{\mathrm{ML}}^{(\mathrm{A})*} \equiv \frac{n^{1/2}\{n^{-1}\mathrm{AIC} + 2\mathrm{E}_f(\hat{\bar{l}}_{\mathrm{ML}}^{*})\}}{(\alpha_{\mathrm{ML2}}^{(\mathrm{A})*})^{1/2}}. \qquad (7.4)$$

In the exponential distribution, from Table 1 when $\alpha = 1$, the factors of order $O(1)$ for the asymptotic cumulants of $z_{\mathrm{ML}}^{(\mathrm{A})*}$ are

$$\alpha_{(z)\mathrm{ML1}}^{(\mathrm{A})*} = 0,\ \alpha_{(z)\mathrm{ML3}}^{(\mathrm{A})*} (= \alpha_{(z)\mathrm{ML3}}^{(\mathrm{A})}) = -1,\ \alpha_{(z)\mathrm{ML}\Delta 2}^{(\mathrm{A})*} (= \alpha_{(z)\mathrm{ML}\Delta 2}^{(\mathrm{A})}) = 0.5$$

$$\text{and } \ \alpha_{(z)\mathrm{ML4}}^{(\mathrm{A})*} (= \alpha_{(z)\mathrm{ML4}}^{(\mathrm{A})}) = 2, \qquad (7.5)$$

where $\alpha_{(z)\mathrm{ML}j}^{(\mathrm{A})*} = \alpha_{\mathrm{ML}j}^{(\mathrm{A})} / (\alpha_{\mathrm{ML2}}^{(\mathrm{A})})^{j/2} = \alpha_{\mathrm{ML}j}^{(\mathrm{A})} / 2^j \, (j = 1,\ \Delta 2,\ 3,\ 4)$. The expression corresponding to (7.1) is given by replacing $\alpha_{(t)\mathrm{ML}j}^{(\mathrm{A})*}$ and $(\hat{v}_{\mathrm{ML}}^{(\mathrm{A})})^{1/2}$ by $\alpha_{(z)\mathrm{ML}j}^{(\mathrm{A})*} (j = 1, \Delta 2, 3, 4)$ and 2, respectively.

A simulation is performed in the following way. An arbitrary population value λ_0, three sample sizes n = 25, 50, 200 and seven nominal confidence levels (coverages) α = 0.005, 0.025, 0.05, 0.5, 0.95, 0.975, 0.995 for z_α are used. Note that different λ_0's give the same results as far as coverages are concerned when the same seeds for random numbers are used and simulated realized values of the observable variable are given proportionately to the population scale $1/\lambda_0$. This was also confirmed by the simulation. While $-2\bar{l}_0^{*} = -2\mathrm{E}_f(\bar{l}_0) = -2\{-\lambda_0 \mathrm{E}_f(\bar{x}) + \log\lambda_0\} = -2(\log\lambda_0 - 1)$ is available when λ_0 is known in the simulation, it is difficult to have $-2\mathrm{E}_f(\hat{\bar{l}}_{\mathrm{ML}}^{*}) = -2\mathrm{E}_f(-\hat{\lambda}_{\mathrm{ML}}\lambda_0 + \log\hat{\lambda}_{\mathrm{ML}})$ in closed form without using an infinite series. In this section, this value is numerically given by a simulation with 10^6 replications.

Table 10 shows the simulated coverages corresponding to the seven nominal values. The first and second blocks of the table are given by studentization when λ_0=1 and 4, respectively with different seeds for random numbers while the third block by standardization when λ_0=1 using the same seeds as those for the first block. Three confidence intervals by Wald and Cornish-Fisher with second- and third order accuracies (denoted by CF2 and CF3, respectively) are used. The simulated coverages show that CF2 improves the coverages by Wald and CF3 those by CF2 when the nominal values are less than 0.5 at the small expense of over correction when the nominal values are greater than or equal to 0.5 in the table. The same coverages by CF2 and CF3, when the nominal value is 0.5, is due to $z_{0.5} = 0$. The results by standardization are somewhat different from those by studentization. However, they are mostly similar. Overall, advantages of the confidence intervals by CF2 and CF3 over those by Wald are shown.

Appendix

A1. Proof of Theorem 1

We obtain an expression of b_2 which is different from that of Konishi and Kitagawa (2003) with b_1 being well known. For the expression, we use the formula of the expansion of $\hat{\boldsymbol{\theta}}_{\mathrm{W}} = \boldsymbol{\theta}_{\mathrm{W}}(\mathbf{X}^*)$ given by Ogasawara (2015a, Equation (2.1) (see also 2015b for correction); 2014, Equation (2.4)):

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0 &= -n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{j=1}^{3}\boldsymbol{\Lambda}^{(j)}\mathbf{l}_0^{(j)} - n^{-1}(\hat{\mathbf{L}}_{\mathrm{W}}^{-1}\hat{\mathbf{q}}_{\mathrm{W}}^* - \boldsymbol{\Lambda}^{-1}\mathbf{q}_0^*)_{O_p(n^{-1/2})} + O_p(n^{-2}) \\
&= -n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{j=1}^{3}\boldsymbol{\Lambda}^{(j)}\mathbf{l}_0^{(j)} + n^{-1}\Bigg[\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* - \boldsymbol{\Lambda}^{-1}\frac{\partial \mathbf{q}^*}{\partial \boldsymbol{\theta}_0{}'}\boldsymbol{\Lambda}^{(1)}\mathbf{l}_0^{(1)} \\
&\qquad - \boldsymbol{\Lambda}^{-1}\,\mathrm{E}_g(\mathbf{J}_0^{(3)})\{(\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^*)\otimes\boldsymbol{\Lambda}^{-1}\}\mathbf{l}_0^{(1)}\Bigg] + O_p(n^{-2}) \qquad \text{(A1.1)} \\
&\equiv -n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{j=1}^{3}\boldsymbol{\Lambda}^{(j)}\mathbf{l}_0^{(j)} + n^{-1}(\mathbf{l}_0^{(\mathrm{W})})_{O_p(n^{-1/2})} + O_p(n^{-2}),
\end{aligned}$$

where $\boldsymbol{\Lambda} = \mathrm{E}_g(\partial^2 \overline{l} / \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}) \equiv \mathrm{E}_g(\partial^2 \overline{l} / \partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0') = O(1)$, $\mathbf{q}_0^* = \mathbf{q}^*(\boldsymbol{\theta}_0)$,

$\boldsymbol{\Lambda}^{(j)} = O(1)$, $\mathbf{l}_0^{(j)} = O_p(n^{-j/2})$ $(j = 1,\ 2,\ 3)$, $\hat{\mathbf{L}}_{\mathrm{W}} = \dfrac{\partial^2 \overline{l}}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{W}}} \equiv \dfrac{\partial^2 \overline{l}}{\partial\hat{\boldsymbol{\theta}}_{\mathrm{W}}\,\partial\hat{\boldsymbol{\theta}}_{\mathrm{W}}'}$,

$\mathbf{q}_{\mathrm{W}}^* = \mathbf{q}^*(\hat{\boldsymbol{\theta}}_{\mathrm{W}})$, $\mathbf{M} = \dfrac{\partial^2 \overline{l}}{\partial\boldsymbol{\theta}_0\,\partial\boldsymbol{\theta}_0'} - \boldsymbol{\Lambda} = O_p(n^{-1/2})$, $\dfrac{\partial\mathbf{q}^*}{\partial\boldsymbol{\theta}_0'} = \dfrac{\partial\mathbf{q}^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$,

$\mathbf{J}_0^{(3)} = \dfrac{\partial^3 \overline{l}}{\partial\boldsymbol{\theta}_0(\partial\boldsymbol{\theta}_0')^{<2>}}$, $\mathbf{x}^{<k>} = \mathbf{x}\otimes\cdots\otimes\mathbf{x}$ (k times of $\mathbf{x}$), $\otimes$ denotes the Kronecker product, and $(\cdot)_{O_p(n^{-1/2})}$ indicates that $(\cdot)$ is of order $O_p(n^{-1/2})$ with other similar expressions.

The term $\sum_{j=1}^{3}\boldsymbol{\Lambda}^{(j)}\mathbf{l}_0^{(j)}$ in (A1.1) (Ogasawara, 2010, Equation (2.4)) is given from the following expansion:

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}_0 = \sum_{j=1}^{3}\boldsymbol{\Lambda}^{(j)}\mathbf{l}_0^{(j)} + O_p(n^{-2}), \tag{A1.2}$$

$$\boldsymbol{\Lambda}^{(1)}\mathbf{l}_0^{(1)} = -\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0},$$

$$\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)} = \boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0} - \frac{1}{2}\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0}\right)^{<2>}$$

$$\begin{aligned}\boldsymbol{\Lambda}^{(3)}\mathbf{l}_0^{(3)} = &-\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0} + \frac{1}{2}\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0}\right)^{<2>} \\ &+\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\left(\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0}\right)\otimes\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0}\right)\right\} - \frac{1}{2}\boldsymbol{\Lambda}^{-1}\{\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})\}\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0}\right)^{<2>} \\ &-\frac{1}{2}\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left[\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0}\right)\otimes\left\{\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0}\right)^{<2>}\right\}\right] \\ &+\frac{1}{6}\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(4)})\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\overline{l}}{\partial\boldsymbol{\theta}_0}\right)^{<3>},\end{aligned}$$

$$\mathbf{J}_0^{(4)} \equiv \frac{\partial^4 \bar{l}}{\partial \boldsymbol{\theta}_0 (\partial \boldsymbol{\theta}_0 ')^{<3>}}, \ \mathbf{l}_0^{(1)} = \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0},$$

$$\mathbf{l}_0^{(2)} = \left\{ \mathrm{v}'(\mathbf{M}) \otimes \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '}, \left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} \right)^{<2>} \right\}' \equiv (\mathbf{l}_0^{(2-1)}{}', \mathbf{l}_0^{(2-2)}{}')' = O_p(n^{-1}),$$

$$\mathbf{l}_0^{(3)} = \left[\mathrm{v}'(\mathbf{M})^{<2>} \otimes \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '}, \mathrm{v}'(\mathbf{M}) \otimes \left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} \right)^{<2>}, \mathrm{vec}'\{\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})\} \otimes \left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} \right)^{<2>} \right.$$
$$\left. \left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} \right)^{<3>} \right]'$$
$$\equiv (\mathbf{l}_0^{(3-1)}{}', \mathbf{l}_0^{(3-2)}{}', \mathbf{l}_0^{(3-3)}{}', \mathbf{l}_0^{(3-4)}{}')' = O_p(n^{-3/2}),$$

where $\boldsymbol{\Lambda}^{(2-j)} = O(1) \ (j = 1,2)$ and $\boldsymbol{\Lambda}^{(3-j)} = O(1) \ (j = 1,...,4)$ are defined implicitly by $\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)} = \sum_{j=1}^{2} \boldsymbol{\Lambda}^{(2-j)}\mathbf{l}_0^{(2-j)}$ and $\boldsymbol{\Lambda}^{(3)}\mathbf{l}_0^{(3)} = \sum_{j=1}^{4} \boldsymbol{\Lambda}^{(3-j)}\mathbf{l}_0^{(3-j)}$; $\mathrm{v}'(\mathbf{M})^{<2>} = [\{\mathrm{v}(\mathbf{M})\}']^{<2>}$; $\mathrm{v}(\cdot)$ is the vectorizing operator taking the non-duplicated elements of a symmetric matrix in parentheses; and $\mathrm{vec}(\cdot)$ is the vectorizing operator stacking the columns of a matrix sequentially.

Expand $-2\hat{\bar{l}}_{\mathrm{W}}$ and $-2\hat{\bar{l}}_{\mathrm{W}}^*$ as

$$-2\hat{\bar{l}}_{\mathrm{W}} = -2(\bar{l}_0)_{O_p(1)} - 2\sum_{j=1}^{4} \frac{1}{j!} \left\{ \frac{\partial^j \bar{l}}{(\partial \boldsymbol{\theta}_0 ')^{<j>}} \right\}_{O_p(1)} \{(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)^{<j>}\}_{O_p(n^{-j/2})} + O_p(n^{-5/2}) \quad \text{(A1.3)}$$

and $-2\hat{\bar{l}}_{\mathrm{W}}^* = -2(\bar{l}_0^*)_{O(1)} - 2\sum_{j=1}^{4} \frac{1}{j!} \left[\mathrm{E}_g \left\{ \frac{\partial^j \bar{l}}{(\partial \boldsymbol{\theta}_0 ')^{<j>}} \right\} \right]_{O(1)} \{(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)^{<j>}\}_{O_p(n^{-j/2})} + O_p(n^{-5/2})$,

respectively. Then, recalling $\mathrm{E}_g(\bar{l}_0) = \bar{l}_0^*$, we have

$$-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{W}} - \hat{\bar{l}}_{\mathrm{W}}^*)$$
$$= -2\mathrm{E}_g \left[\sum_{j=1}^{3} \frac{1}{j!} \left[\frac{\partial^j \bar{l}}{(\partial \boldsymbol{\theta}_0 ')^{<j>}} - \mathrm{E}_g \left\{ \frac{\partial^j \bar{l}}{(\partial \boldsymbol{\theta}_0 ')^{<j>}} \right\} \right]_{O_p(n^{-1/2})} (\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)^{<j>} \right] + O(n^{-3}) \quad \text{(A1.4)}$$

$$= -2\mathrm{E}_g\left\{\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0'}(\hat{\boldsymbol{\theta}}_{\mathrm{W}}-\boldsymbol{\theta}_0)\right\}_{\to O(n^{-2})} - \mathrm{E}_g\{\mathrm{vec}'(\mathbf{M})(\hat{\boldsymbol{\theta}}_{\mathrm{W}}-\boldsymbol{\theta}_0)^{<2>}\}_{\to O(n^{-2})}$$

$$-\frac{1}{3}\mathrm{E}_g\{\mathrm{vec}'\{\mathbf{J}_0^{(3)}-\mathrm{E}_g(\mathbf{J}_0^{(3)})\}(\hat{\boldsymbol{\theta}}_{\mathrm{W}}-\boldsymbol{\theta}_0)^{<3>}\}_{\to O(n^{-2})} + O(n^{-3}),$$

where the term of $j = 4$ in $\sum_{j=1}^{4}(\cdot)$ of (A1.3), when the expectation is taken, is absorbed in the remainder term of order $O(n^{-3})$; and $\mathrm{E}_g(\cdot)_{\to O(n^{-2})}$ indicates that the expectation is taken up to order $O(n^{-2})$.

Let $\boldsymbol{\Gamma}=n\,\mathrm{E}_g\left(\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0'}\right)$. When the model is true, $\boldsymbol{\Gamma}=-\boldsymbol{\Lambda}=\mathbf{I}_0$, where $\mathbf{I}_0$ is the population Fisher information matrix per observation. Under possible model misspecification, the last three expectations in (A1.4) are given as

$$-2\mathrm{E}_g\left\{\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0'}(\hat{\boldsymbol{\theta}}_{\mathrm{W}}-\boldsymbol{\theta}_0)\right\}$$

$$= -2\mathrm{E}_g\left\{\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0'}\left(-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^{*}+\sum_{j=1}^{3}\boldsymbol{\Lambda}^{(j)}\mathbf{l}_0^{(j)}+n^{-1}\mathbf{l}_0^{(\mathrm{W})}\right)\right\}$$

$$= \left\{2\mathrm{E}_g\left(\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0'}\boldsymbol{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\right\}_{O(n^{-1})} - \left\{2\mathrm{E}_g\left(\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0'}\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)}\right)\right\}_{O(n^{-2})}$$

$$- \left\{2\mathrm{E}_g\left(\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0'}\boldsymbol{\Lambda}^{(3)}\mathbf{l}_0^{(3)}\right)\right\}_{O(n^{-2})} - \left\{2\mathrm{E}_g\left(n^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0'}\mathbf{l}_0^{(\mathrm{W})}\right)\right\}_{O(n^{-2})} + O(n^{-3})$$

$$= 2n^{-1}\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}) - 2n^{-2}\underset{(\mathrm{A})}{\Bigg[}\sum_{a\ge b}\sum_{c,d=1}^{q}(\boldsymbol{\Lambda}^{(2-1)})_{(d:ab,\,c)}n^2\mathrm{E}_g\left(m_{ab}\frac{\partial\bar{l}}{\partial\theta_{0c}}\frac{\partial\bar{l}}{\partial\theta_{0d}}\right) \qquad \text{(A1.5)}$$

$$+\sum_{a,b,c=1}^{q}(\boldsymbol{\Lambda}^{(2-2)})_{(c:a,b)}n^2\mathrm{E}_g\left(\frac{\partial\bar{l}}{\partial\theta_{0a}}\frac{\partial\bar{l}}{\partial\theta_{0b}}\frac{\partial\bar{l}}{\partial\theta_{0c}}\right)+\sum_{a\ge b}\sum_{c\ge d}\sum_{e,f=1}^{q}(\boldsymbol{\Lambda}^{(3-1)})_{(f:ab,\,cd,e)}$$

$$\times\left\{n\,\mathrm{cov}_g(m_{ab},m_{cd})\gamma_{ef}+\sum_{(e,f)}^{2}n\,\mathrm{cov}_g\left(m_{ab},\frac{\partial\bar{l}}{\partial\theta_{0e}}\right)n\,\mathrm{cov}_g\left(m_{cd},\frac{\partial\bar{l}}{\partial\theta_{0f}}\right)\right\}$$

$$+\sum_{a\ge b}\sum_{c,d,e=1}^{q}(\mathbf{\Lambda}^{(3-2)})_{(e:ab,\,c,d)}\sum_{(c,d,e)}^{3} n\,\mathrm{cov}_g\left(m_{ab},\frac{\partial\bar{l}}{\partial\theta_{0c}}\right)\gamma_{de}+\sum_{a,b,c,d,e,f=1}^{q}(\mathbf{\Lambda}^{(3-3)})_{(f:abc,\,d,e)}$$

$$\times\sum_{(d,e,f)}^{3} n\,\mathrm{cov}_g\left\{(\mathbf{J}_0^{(3)})_{(a,b,c)},\frac{\partial\bar{l}}{\partial\theta_{0d}}\right\}\gamma_{ef}+\sum_{a,b,c,d=1}^{q}(\mathbf{\Lambda}^{(3-4)})_{(d:a,b,c)}(\gamma_{ab}\gamma_{cd}+\gamma_{ac}\gamma_{bd}+\gamma_{ad}\gamma_{bc})$$

$$+\sum_{a,b,c=1}^{q}\lambda^{ab}(\mathbf{\Lambda}^{-1}\mathbf{q}_0^{*})_c\, n\,\mathrm{cov}_g\left(\frac{\partial\bar{l}}{\partial\theta_{0a}},m_{bc}\right)+\mathrm{tr}\left(\frac{\partial\mathbf{q}^{*}}{\partial\boldsymbol{\theta}_0{}'}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}\mathbf{\Lambda}^{-1}\right)$$

$$\left.-\mathrm{tr}[\mathrm{E}_g(\mathbf{J}_0^{(3)})\{(\mathbf{\Lambda}^{-1}\mathbf{q}_0^{*})\otimes(\mathbf{\Lambda}^{-1}\mathbf{\Gamma}\mathbf{\Lambda}^{-1})\}]\quad\right]_{(\mathrm{A})}+O(n^{-3})$$

$$\equiv n^{-1}b_1+n^{-2}c_1+O(n^{-3})\ \ (b_1=2\mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{\Gamma}),\ c_1=-2\underset{(\mathbf{A})}{[}\cdot\underset{(\mathbf{A})}{]}),$$

where $(\mathbf{\Lambda}^{(2-1)})_{(d:ab,\,c)}$ indicates the element of the d-th row and the column corresponding to $(\mathbf{M})_{ab}\equiv m_{ab}$ (the (a, b)th element of $\mathbf{M}$) and $\partial\bar{l}/\partial(\boldsymbol{\theta}_0)_c\equiv\partial\bar{l}/\partial\theta_{0c}$ of $\mathbf{\Lambda}^{(2-1)}$ with $(\cdot)_c$ being the c-th element of a vector with other expressions defined similarly; $\sum_{a\ge b}(\cdot)\equiv\sum_{b=1}^{a}\sum_{a=1}^{q}(\cdot)$, $\sum_{e,f=1}^{q}(\cdot)=\sum_{e=1}^{q}\sum_{f=1}^{q}(\cdot)$; $\mathrm{cov}_g(\cdot)$ is the covariance using the distribution $g(\mathbf{X}^{*}\,|\,\boldsymbol{\zeta}_0)$; $\sum_{(e,f)}^{2}(\cdot)$ is the sum of two symmetric terms with respect to e and f with $\sum_{(c,d,e)}^{3}(\cdot)$ defined similarly; and $\underset{(\mathbf{A})}{[}\cdot\underset{(\mathbf{A})}{]}$ is for ease of finding correspondence;

$$-\mathrm{E}_g\{\mathrm{vec}'(\mathbf{M})(\hat{\boldsymbol{\theta}}_{\mathrm{W}}-\boldsymbol{\theta}_0)^{<2>}\}$$

$$=-\mathrm{E}_g\left[\mathrm{vec}'(\mathbf{M})\left\{2(-n^{-1}\mathbf{\Lambda}^{-1}\mathbf{q}_0^{*})\otimes\left(-\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)+\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)^{<2>}\right.\right.$$

$$\left.\left.+2\left(-\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\otimes(\mathbf{\Lambda}^{(2)}\mathbf{l}_0^{(2)})\right\}\right]$$

$$
\begin{aligned}
&= -n^{-2}\Bigg[\underset{\text{(A)}}{}\; 2\sum_{a,b,c=1}^{q} (\mathbf{\Lambda}^{-1}\mathbf{q}_0^*)_a \lambda^{bc} n\,\mathrm{cov}_g\left(m_{ab}, \frac{\partial \bar{l}}{\partial \theta_{0c}} \right) + \sum_{a,b,c,d=1}^{q} \lambda^{ac}\lambda^{bd} n^2 \mathrm{E}_g\left(m_{ab}\frac{\partial \bar{l}}{\partial \theta_{0c}}\frac{\partial \bar{l}}{\partial \theta_{0d}} \right) \\
&\quad -2\sum_{a,b,c=1}^{q}\sum_{d \ge e}^{q}\sum_{f=1} (\mathbf{\Lambda}^{(2-1)})_{(b:de,\,f)} \lambda^{ac} \left\{ \sum_{(c,f)}^{2} n\,\mathrm{cov}_g\left(m_{ab}, \frac{\partial \bar{l}}{\partial \theta_{0c}} \right) n\,\mathrm{cov}_g\left(m_{de}, \frac{\partial \bar{l}}{\partial \theta_{0f}} \right) \right. \\
&\quad \left. + n\,\mathrm{cov}_g(m_{ab}, m_{de})\gamma_{cf} \right\} \qquad\qquad \text{(A1.6)} \\
&\quad -2\sum_{a,b,c,d,e=1}^{q} (\mathbf{\Lambda}^{(2-2)})_{(b:d,e)} \lambda^{ac} \sum_{(c,d,e)}^{3} n\,\mathrm{cov}_g\left(m_{ab}, \frac{\partial \bar{l}}{\partial \theta_{0c}} \right)\gamma_{de} \;\underset{\text{(A)}}{\Bigg]} + O(n^{-3}) \\
&\equiv n^{-2}c_2 + O(n^{-3}),
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{1}{3}\mathrm{E}_g[\mathrm{vec}'\{\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})\}(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)^{<3>}] \\
&= -\frac{1}{3}\mathrm{E}_g\left[\mathrm{vec}'\{\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})\}\left(-\mathbf{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \theta_{0d}} \right)^{<3>} \right] + O(n^{-3}) \qquad\qquad \text{(A1.7)} \\
&= n^{-2}\sum_{a,b,c,d,e,f=1}^{q} \lambda^{ad}\lambda^{be}\lambda^{cf} n\,\mathrm{cov}_g\left\{ (\mathbf{J}_0^{(3)})_{(a,b,c)}, \frac{\partial \bar{l}}{\partial \theta_{0d}} \right\}\gamma_{ef} + O(n^{-3}) \\
&\equiv n^{-2}c_3 + O(n^{-3}),
\end{aligned}
$$

where $\lambda^{bc} = (\mathbf{\Lambda}^{-1})_{bc}$. Then, from (A1.5) to (A1.7) we have (2.9).

A2. Proof of Corollary 1

Under canonical parametrization in the exponential family, it is known that $\dfrac{\partial^j \bar{l}}{(\partial \boldsymbol{\theta}_0)^{<j>}} = \mathrm{E}_g\left\{ \dfrac{\partial^j \bar{l}}{(\partial \boldsymbol{\theta}_0)^{<j>}} \right\} (j = 2, 3, \ldots)$, which gives c_1 of (2.11) from (A1.5) with $\mathbf{M} = \mathbf{O}$ (a zero matrix of an appropriate size) and $\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)}) = \mathbf{O}$. The results of $c_2 = c_3 = 0$ are derived similarly from (A1.6) and (A1.7) with $\mathbf{M} = \mathbf{O}$ and $\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)}) = \mathbf{O}$, respectively.

A3. Proof of Corollary 2

In the case of the MLE, the two terms associated with $\mathbf{q}_0^*$ in (2.11) vanish and recalling (A1.2) for $\mathbf{\Lambda}^{(2-2)}$ and $\mathbf{\Lambda}^{(3-4)}$ in c_1 of (2.11), we have

$$
\begin{aligned}
c_1 &= -2\left\{ \sum_{a,b,c=1}^{q} (\mathbf{\Lambda}^{(2-2)})_{(c:a,b)} n^2 \mathrm{E}_g \left(\frac{\partial \bar{l}}{\partial \theta_{0a}} \frac{\partial \bar{l}}{\partial \theta_{0b}} \frac{\partial \bar{l}}{\partial \theta_{0c}} \right) \right. \\
&\qquad \left. + \sum_{a,b,c,d=1}^{q} (\mathbf{\Lambda}^{(3-4)})_{(d:a,b,c)} (\gamma_{ab}\gamma_{cd} + \gamma_{ac}\gamma_{bd} + \gamma_{ad}\gamma_{bc}) \right\} \\
&= -2\left[\sum_{a,b,c=1}^{q} \left\{ -\frac{1}{2} \mathbf{\Lambda}^{-1} \mathbf{J}_0^{(3)} (\mathbf{\Lambda}^{-1})^{<2>} \right\}_{(c:a,b)} n^2 \mathrm{E}_g \left(\frac{\partial \bar{l}}{\partial \theta_{0a}} \frac{\partial \bar{l}}{\partial \theta_{0b}} \frac{\partial \bar{l}}{\partial \theta_{0c}} \right) \right. \\
&\qquad - \sum_{a,b,c,d=1}^{q} \frac{1}{2} (\mathbf{\Lambda}^{-1})_{d\bullet} \mathbf{J}_0^{(3)} \left[(\mathbf{\Lambda}^{-1})_{\bullet a} \otimes [\mathbf{\Lambda}^{-1} \mathbf{J}_0^{(3)} \{ (\mathbf{\Lambda}^{-1})_{\bullet b} \otimes (\mathbf{\Lambda}^{-1})_{\bullet c} \}] \right] \\
&\qquad\qquad \times (\gamma_{ab}\gamma_{cd} + \gamma_{ac}\gamma_{bd} + \gamma_{ad}\gamma_{bc}) \\
&\qquad \left. + \sum_{a,b,c,d=1}^{q} \frac{1}{6} \{ \mathbf{\Lambda}^{-1} \mathbf{J}_0^{(4)} (\mathbf{\Lambda}^{-1})^{<3>} \}_{(d:a,b,c)} 3\gamma_{ab}\gamma_{cd} \right] \\
&= -\mathrm{vec}'(\mathbf{J}_0^{(3)}) n^2 \mathrm{E}_g \left\{ \left(-\mathbf{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)^{<3>} \right\} + \mathrm{vec}'(\mathbf{\Lambda}^{-1}\mathbf{\Gamma}\mathbf{\Lambda}^{-1}) \mathbf{J}_0^{(3)}{}' \mathbf{\Lambda}^{-1} \mathbf{J}_0^{(3)} \mathrm{vec}(\mathbf{\Lambda}^{-1}\mathbf{\Gamma}\mathbf{\Lambda}^{-1}) \\
&\qquad + 2\mathrm{vec}'(\mathbf{J}_0^{(3)}) \{ \mathbf{\Lambda}^{-1} \otimes (\mathbf{\Lambda}^{-1}\mathbf{\Gamma}\mathbf{\Lambda}^{-1})^{<2>} \} \mathrm{vec}(\mathbf{J}_0^{(3)}) - \mathrm{vec}'(\mathbf{J}_0^{(4)}) \mathrm{vec}\{ (\mathbf{\Lambda}^{-1}\mathbf{\Gamma}\mathbf{\Lambda}^{-1})^{<2>} \}, \qquad \text{(A3.1)}
\end{aligned}
$$

where $(\cdot)_{d\bullet}$ is the *d*-th row of a matrix and $(\cdot)_{\bullet a}$ is the *a*-th column of a matrix.

Under correct model specification and canonical parametrization, since $\partial l_j / \partial \boldsymbol{\theta}_0 = \mathbf{x}^* - \mathrm{E}_f(\mathbf{x}^*)$ and $-\mathbf{\Lambda} = \mathbf{\Gamma} = \mathbf{I}_0$, (A3.1) becomes

$$
\begin{aligned}
c_1 &= \boldsymbol{\kappa}_{f3}{}'(\mathbf{x}^*) \boldsymbol{\kappa}_{f3} \left(\mathbf{I}_0^{-1} \frac{\partial l_j}{\partial \boldsymbol{\theta}_0} \right) - \mathrm{vec}'(\mathbf{I}_0^{-1}) \mathbf{J}_0^{(3)}{}' \mathbf{I}_0^{-1} \mathbf{J}_0^{(3)} \mathrm{vec}(\mathbf{I}_0^{-1}) \\
&\qquad - 2\mathrm{vec}'(\mathbf{J}_0^{(3)}) (\mathbf{I}_0^{-1})^{<3>} \mathrm{vec}(\mathbf{J}_0^{(3)}) + \boldsymbol{\kappa}_{f4}{}'(\mathbf{x}^*) \mathrm{vec}\{ (\mathbf{I}_0^{-1})^{<2>} \} \\
&= \boldsymbol{\kappa}_{f3}{}'(\mathbf{I}_0^{-1/2}\mathbf{x}^*) \boldsymbol{\kappa}_{f3} \left(\mathbf{I}_0^{-1/2} \frac{\partial l_j}{\partial \boldsymbol{\theta}_0} \right) - \boldsymbol{\kappa}_{f3}{}'(\mathbf{I}_0^{-1/2}\mathbf{x}^*) [\mathbf{I}_{(q)} \otimes \{ \mathrm{vec}(\mathbf{I}_{(q)}) \mathrm{vec}'(\mathbf{I}_{(q)}) \}] \boldsymbol{\kappa}_{f3}(\mathbf{I}_0^{-1/2}\mathbf{x}^*) \qquad \text{(A3.2)} \\
&\qquad - 2\boldsymbol{\kappa}_{f3}{}'(\mathbf{I}_0^{-1/2}\mathbf{x}^*) \boldsymbol{\kappa}_{f3}(\mathbf{I}_0^{-1/2}\mathbf{x}^*) + \boldsymbol{\kappa}_{f4}{}'(\mathbf{I}_0^{-1/2}\mathbf{x}^*) \mathrm{vec}\{ (\mathbf{I}_{(q)})^{<2>} \} \\
&= \boldsymbol{\kappa}_{f3}{}'(\tilde{\mathbf{x}}^*) \boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*) - \boldsymbol{\kappa}_{f3}{}'(\tilde{\mathbf{x}}^*) [\mathbf{I}_{(q)} \otimes \{ \mathrm{vec}(\mathbf{I}_{(q)}) \mathrm{vec}'(\mathbf{I}_{(q)}) \}] \boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*) \\
&\qquad - 2\boldsymbol{\kappa}_{f3}{}'(\tilde{\mathbf{x}}^*) \boldsymbol{\kappa}_{f3}(\tilde{\mathbf{x}}^*) + \boldsymbol{\kappa}_{f4}{}'(\tilde{\mathbf{x}}^*) \mathrm{vec}(\mathbf{I}_{(q^2)})
\end{aligned}
$$

which gives (2.12).

A4. Proof of Corollary 4

Since $\mathbf{J}_0^{(j)} \equiv \dfrac{\partial^j \bar{l}}{\partial \boldsymbol{\theta}_0 (\partial \boldsymbol{\theta}_0 ')^{<j-1>}} = \mathrm{E}_g \left\{ \dfrac{\partial^j \bar{l}}{\partial \boldsymbol{\theta}_0 (\partial \boldsymbol{\theta}_0 ')^{<j-1>}} \right\} (j = 2,3,...)$ under canonical parametrization, the asymptotic expansion using the MLE corresponding to (A1.4) higher than (A1.4) is given only by the first term $-2\mathrm{E}_g \left\{ \dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} (\hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}_0) \right\}$, which is also given only by $-2\mathrm{E}_g \left\{ \dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} \left(-\boldsymbol{\Lambda}^{-1} \dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \right\}$ and $-2\mathrm{E}_g \{ h(\mathbf{J}_0^{(3)}, \mathbf{J}_0^{(4)}, ...)\}$, where $h(\cdot)$ is the sum of multiplicative functions of the powers of the arguments.

In the only non-vanishing term $-2\mathrm{E}_g \left\{ \dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} (\hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}_0) \right\}$ for the expansion of the left-hand side of (2.15), $-2\mathrm{E}_g \left\{ \dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} \left(-\boldsymbol{\Lambda}^{-1} \dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \right\} = -2\mathrm{tr} \left\{ \boldsymbol{\Sigma} \mathrm{E}_g \left(\dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0 '} \right) \right\}$ $= -n^{-1} 2\mathrm{tr}(\boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1}) = -n^{-1} 2q$ under arbitrary distributions as long as $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ exist. The remaining terms $-2\mathrm{E}_g \{ h(\mathbf{J}_0^{(3)}, \mathbf{J}_0^{(4)}, ...)\}$ vanish when we use the normal distribution even under non-normality since $\mathbf{J}_0^{(j)} = \mathbf{O} (j = 3,4,...)$ in this case.

An alternative direct proof of is given as follows. Let $\mathbf{z}_j (j = 1,...,n)$ be independent copies of $\mathbf{x}^*$ and $\mathrm{E}_{\mathbf{Z}^*}(\cdot)$ denote an expectation over the distribution of $\mathbf{Z}^*$ or $\mathbf{z}_j (j = 1,...,n)$. Then, by definition,

$$\begin{aligned} -2\hat{\bar{l}}_{\mathrm{ML}}^* &= -2\mathrm{E}_{\mathbf{Z}^*} \left[-\frac{n^{-1}}{2} \sum_{j=1}^n (\mathbf{z}_j - \bar{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\mathbf{z}_j - \bar{\mathbf{x}}) - \frac{1}{2} \log\{(2\pi)^q | \boldsymbol{\Sigma} |\} \right] \\ &= \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}) + (\boldsymbol{\mu}_0 - \bar{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_0 - \bar{\mathbf{x}})' + \log\{(2\pi)^q | \boldsymbol{\Sigma} |\} \\ &= q + (\boldsymbol{\mu}_0 - \bar{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_0 - \bar{\mathbf{x}})' + \log\{(2\pi)^q | \boldsymbol{\Sigma} |\}, \end{aligned} \qquad \text{(A4.1)}$$

which gives $-2\mathrm{E}_g (\hat{\bar{l}}_{\mathrm{ML}}^*) = (1 + n^{-1})q + \log\{(2\pi)^q | \boldsymbol{\Sigma} |\}$. On the other hand,

$$
\begin{aligned}
-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{ML}}) &= -2\mathrm{E}_g\left[-\frac{n^{-1}}{2}\sum_{j=1}^{n}(\mathbf{x}_j-\overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j-\overline{\mathbf{x}})-\frac{1}{2}\log\{(2\pi)^q\,|\,\boldsymbol{\Sigma}\,|\}\right] \\
&= (1-n^{-1})\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma})+\log\{(2\pi)^q\,|\,\boldsymbol{\Sigma}\,|\} \\
&= (1-n^{-1})q+\log\{(2\pi)^q\,|\,\boldsymbol{\Sigma}\,|\}.
\end{aligned} \tag{A4.2}
$$

Consequently, (A4.1) and (A4.2) yield $-2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{ML}}-\hat{\bar{l}}^{*}_{\mathrm{ML}})=-n^{-1}2q$.

A5. Expressions of $-\boldsymbol{\Lambda}_{\mathbf{M}}^{-1(\Delta)}$, $-\boldsymbol{\Lambda}_{\mathbf{M}}^{-1(\Delta\Delta)}$, $\boldsymbol{\Gamma}_{\mathbf{M}}^{(\Delta)}$ and $\boldsymbol{\Gamma}_{\mathbf{M}}^{(\Delta\Delta)}$

Let $\mathbf{L}_0=\left(\dfrac{\partial^2\bar{l}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0\,'}\right)_{O_p(1)}$, then

$$
\begin{aligned}
-\hat{\mathbf{L}}_{\mathrm{W}}^{-1} &= -\mathbf{L}_0^{-1}+\sum_{j=1}^{q}\mathbf{L}_0^{-1}\frac{\partial\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_j}\mathbf{L}_0^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{W}}-\boldsymbol{\theta}_0)_j \\
&\quad+\sum_{j,k=1}^{q}\left\{-\mathbf{L}_0^{-1}\frac{\partial\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_j}\mathbf{L}_0^{-1}\frac{\partial\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_k}\mathbf{L}_0^{-1}+\frac{1}{2}\mathbf{L}_0^{-1}\frac{\partial^2\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_j\partial(\boldsymbol{\theta}_0)_k}\mathbf{L}_0^{-1}\right\} \\
&\qquad\times(\hat{\boldsymbol{\theta}}_{\mathrm{W}}-\boldsymbol{\theta}_0)_j(\hat{\boldsymbol{\theta}}_{\mathrm{W}}-\boldsymbol{\theta}_0)_k+O_p(n^{-3/2})
\end{aligned}
$$

$$
\begin{aligned}
&= -\boldsymbol{\Lambda}^{-1}+\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}-\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1} \\
&+(\boldsymbol{\Lambda}^{-1}-\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1})\sum_{j=1}^{q}\left[\mathrm{E}_g\left(\frac{\partial\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_j}\right)+\left\{\frac{\partial\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_j}-\mathrm{E}_g\left(\frac{\partial\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_j}\right)\right\}\right](\boldsymbol{\Lambda}^{-1}-\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1}) \\
&\quad\times\left(-\boldsymbol{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^{*}+\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)}\right)_j \\
&+\sum_{j,k=1}^{q}\left\{-\boldsymbol{\Lambda}^{-1}\mathrm{E}_g\left(\frac{\partial\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_j}\right)\boldsymbol{\Lambda}^{-1}\mathrm{E}_g\left(\frac{\partial\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_k}\right)\boldsymbol{\Lambda}^{-1}+\frac{1}{2}\boldsymbol{\Lambda}^{-1}\mathrm{E}_g\left(\frac{\partial^2\mathbf{L}_0}{\partial(\boldsymbol{\theta}_0)_j\partial(\boldsymbol{\theta}_0)_k}\right)\boldsymbol{\Lambda}^{-1}\right\} \\
&\quad\times\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)_j\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)_k+O_p(n^{-3/2})
\end{aligned}
$$

$$
\begin{aligned}
&= -\mathbf{\Lambda}^{-1} + \left[\mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1} - \mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\mathbf{\Lambda}^{-1} \otimes \left(\mathbf{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\right\}\right]_{O_p(n^{-1/2})} \\
&+ \Bigg[_{(\mathrm{A})} -\mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1} + \mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\mathbf{\Lambda}^{-1} \otimes \left(\mathbf{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\right\} \\
&+ \mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{(\mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1}) \otimes \left(\mathbf{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\right\} - \mathbf{\Lambda}^{-1}\{\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})\}\left\{\mathbf{\Lambda}^{-1} \otimes \left(\mathbf{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\right\} \\
&+ \mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\mathbf{\Lambda}^{-1} \otimes (-n^{-1}\mathbf{\Lambda}^{-1}\mathbf{q}_0^* + \mathbf{\Lambda}^{(2)}\mathbf{l}_0^{(2)})\right\} - \mathbf{\Lambda}^{-1}\left[\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\mathbf{\Lambda}^{-1} \otimes \left(\mathbf{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\right\}\right]^2 \\
&+ \frac{1}{2}\mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(4)})\left\{\mathbf{\Lambda}^{-1} \otimes \left(\mathbf{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)^{<2>}\right\} \Bigg]_{(\mathrm{A})O_p(n^{-1})} + O_p(n^{-3/2}) \\
&\equiv -\mathbf{\Lambda}^{-1} + (-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)})_{O_p(n^{-1/2})} + (-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2}).
\end{aligned}
\tag{A5.1}
$$

Let

$$
\begin{aligned}
&\mathbf{G} \equiv \mathbf{G}(\boldsymbol{\theta}) \equiv \mathbf{G}(\boldsymbol{\theta}, \mathbf{X}^*) \equiv \left(n^{-1}\sum_{j=1}^{n}\frac{\partial l_j}{\partial \boldsymbol{\theta}}\frac{\partial l_j}{\partial \boldsymbol{\theta}'}\right)_{O_p(1)},\ \mathbf{G}_0 \equiv \mathbf{G}(\boldsymbol{\theta}_0) \equiv \mathbf{G}(\boldsymbol{\theta}_0, \mathbf{X}^*), \\
&\mathbf{G}_0 = \mathbf{\Gamma} + (\mathbf{M}_{\mathbf{G}})_{O_p(n^{-1/2})},\ \mathrm{E}_g(\mathbf{G}_0) = \mathbf{\Gamma},\ \mathbf{G}_{0(j)}^{(3)} = \partial \mathbf{G}_0 / \partial(\boldsymbol{\theta}_0)_j, \\
&\mathbf{G}_{0(j,k)}^{(4)} = \partial^2 \mathbf{G}_0 / \partial(\boldsymbol{\theta}_0)_j \partial(\boldsymbol{\theta}_0)_j\ (j,k = 1,...,q),
\end{aligned}
\tag{A5.2}
$$

then

$$
\begin{aligned}
\hat{\mathbf{\Gamma}}_{\mathrm{W}} &= \mathbf{G}(\hat{\boldsymbol{\theta}}_{\mathrm{W}}, \mathbf{X}^*) = (\mathbf{G}_0)_{O_p(1)} + (\hat{\mathbf{\Gamma}}_{\mathrm{W}} - \mathbf{G}_0)_{O_p(n^{-1/2})} \\
&= \mathbf{\Gamma} + \mathbf{M}_{\mathbf{G}} + \sum_{j=1}^{q}\mathbf{G}_{0(j)}^{(3)}(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)_j + \frac{1}{2}\sum_{j,k=1}^{q}\mathbf{G}_{0(j,k)}^{(4)}(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)_j(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0)_k + O_p(n^{-3/2}) \\
&= \mathbf{\Gamma} + \mathbf{M}_{\mathbf{G}} + \sum_{j=1}^{q}\mathbf{G}_{0(j)}^{(3)}\left(-n^{-1}\mathbf{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{k=1}^{2}\mathbf{\Lambda}^{(k)}\mathbf{l}_0^{(k)}\right)_j + \frac{1}{2}\sum_{j,k=1}^{q}\mathbf{G}_{0(j,k)}^{(4)}(\mathbf{\Lambda}^{(1)}\mathbf{l}_0^{(1)})_j(\mathbf{\Lambda}^{(1)}\mathbf{l}_0^{(1)})_k \\
&\quad + O_p(n^{-3/2})
\end{aligned}
\tag{A5.3}
$$

$$= \boldsymbol{\Gamma} + \left\{ \mathbf{M}_{\mathbf{G}} - \sum_{j=1}^{q} \mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)}) \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)_j \right\}_{O_p(n^{-1/2})}$$

$$+ \Bigg[- \sum_{j=1}^{q} \{ \mathbf{G}_{0(j)}^{(3)} - \mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)}) \} \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)_j + \sum_{j=1}^{q} \mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})(-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)})_j$$

$$+ \frac{1}{2} \sum_{j,k=1}^{q} \mathrm{E}_g(\mathbf{G}_{0(j,k)}^{(4)}) \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)_j \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)_k \Bigg]_{O_p(n^{-1})}^{(\mathrm{A})} + O_p(n^{-3/2})$$

$$\equiv \boldsymbol{\Gamma} + (\boldsymbol{\Gamma}_{\mathbf{M}}^{(\Delta)})_{O_p(n^{-1/2})} + (\boldsymbol{\Gamma}_{\mathbf{M}}^{(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2}).$$

A6. Actual expressions of $-\boldsymbol{\Lambda}_{\mathbf{I}}^{-1(\Delta)}$, $-\boldsymbol{\Lambda}_{\mathbf{I}}^{-1(\Delta\Delta)}$, $\boldsymbol{\Gamma}_{\mathbf{I}}^{(\Delta)}$ and $\boldsymbol{\Gamma}_{\mathbf{I}}^{(\Delta\Delta)}$

Omitting terms with $\mathbf{M}$, $\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})$, $\mathbf{M}_{\mathbf{G}}$ and $\mathbf{G}_{0(j)}^{(3)} - \mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})$ in (3.6) and (3.7), we obtain

$$\hat{\mathbf{I}}_{\mathrm{W}}^{(-\boldsymbol{\Lambda})-1} = -\boldsymbol{\Lambda}^{-1} - \left[\boldsymbol{\Lambda}^{-1} \mathrm{E}_g(\mathbf{J}_0^{(3)}) \left\{ \boldsymbol{\Lambda}^{-1} \otimes \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \right\} \right]_{O_p(n^{-1/2})}$$

$$+ \Bigg[\boldsymbol{\Lambda}^{-1} \mathrm{E}_g(\mathbf{J}_0^{(3)}) \{ \boldsymbol{\Lambda}^{-1} \otimes (-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)}) \} - \boldsymbol{\Lambda}^{-1} \left[\mathrm{E}_g(\mathbf{J}_0^{(3)}) \left\{ \boldsymbol{\Lambda}^{-1} \otimes \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \right\} \right]^2$$

$$+ \frac{1}{2} \boldsymbol{\Lambda}^{-1} \mathrm{E}_g(\mathbf{J}_0^{(4)}) \left\{ \boldsymbol{\Lambda}^{-1} \otimes \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)^{<2>} \right\} \Bigg]_{O_p(n^{-1})}^{(\mathrm{A})} + O_p(n^{-3/2})$$

$$\equiv -\boldsymbol{\Lambda}^{-1} + (-\boldsymbol{\Lambda}_{\mathbf{I}}^{-1(\Delta)})_{O_p(n^{-1/2})} + (-\boldsymbol{\Lambda}_{\mathbf{I}}^{-1(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2}),$$

(A6.1)

$$\hat{\mathbf{I}}_{\mathrm{W}}^{(\boldsymbol{\Gamma})} = \boldsymbol{\Gamma} - \left\{ \sum_{j=1}^{q} \mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)}) \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)_j \right\}_{O_p(n^{-1/2})} + \Bigg[\sum_{j=1}^{n} \mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})(-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)})_j$$

$$+ \frac{1}{2} \sum_{j,k=1}^{q} \mathrm{E}_g(\mathbf{G}_{0(j,k)}^{(4)}) \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)_j \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)_k \Bigg]_{O_p(n^{-1})}^{(\mathrm{A})} + O_p(n^{-3/2})$$

$$\equiv \boldsymbol{\Gamma} + (\boldsymbol{\Gamma}_{\mathbf{I}}^{(\Delta)})_{O_p(n^{-1/2})} + (\boldsymbol{\Gamma}_{\mathbf{I}}^{(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2}).$$

$$\tag{A6.2}$$

A7. Actual expressions of $d^{(\mathrm{T1})}$ in $\mathrm{E}_g\{2(\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T1})})\}$ and $d^{(\mathrm{T2})}$ in $\mathrm{E}_g\{2(\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T2})})\}$

$$\begin{aligned}
&\mathrm{E}_g\{2(\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T1})})\} \\
&= 2\mathrm{E}_g\{\mathrm{tr}(-\mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta)}\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta)} - \mathbf{\Lambda}_{\mathbf{M}}^{-1(\Delta\Delta)}\mathbf{\Gamma} - \mathbf{\Lambda}^{-1}\mathbf{\Gamma}_{\mathbf{M}}^{(\Delta\Delta)})\} \\
&= n^{-1}2\mathrm{E}_g\mathrm{tr}\underset{(\mathrm{A})}{\Bigg[}\; n\left[\mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1} - \mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\mathbf{\Lambda}^{-1}\otimes\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\right\}\right] \\
&\qquad\qquad \times\left\{\mathbf{M}_{\mathbf{G}} - \sum_{j=1}^{q}\mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)_j\right\} \\
&+n\underset{(\mathrm{B})}{\Bigg[}\; -\mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1} + \mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\mathbf{\Lambda}^{-1}\otimes\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\right\} \\
&\quad + \mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{(\mathbf{\Lambda}^{-1}\mathbf{M}\mathbf{\Lambda}^{-1})\otimes\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\right\} - \mathbf{\Lambda}^{-1}\{\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})\}\left\{\mathbf{\Lambda}^{-1}\otimes\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\right\} \\
&\quad + \mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\mathbf{\Lambda}^{-1}\otimes(-n^{-1}\mathbf{\Lambda}^{-1}\mathbf{q}_0^{*} + \mathbf{\Lambda}^{(2)}\mathbf{l}_0^{(2)})\right\} - \mathbf{\Lambda}^{-1}\left[\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\mathbf{\Lambda}^{-1}\otimes\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\right\}\right]^2 \\
&\quad + \frac{1}{2}\mathbf{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(4)})\left\{\mathbf{\Lambda}^{-1}\otimes\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)^{<2>}\right\}\underset{(\mathrm{B})}{\Bigg]}\mathbf{\Gamma} \\
&-n\mathbf{\Lambda}^{-1}\underset{(\mathrm{C})}{\Bigg[}\; -\sum_{j=1}^{q}\{\mathbf{G}_{0(j)}^{(3)} - \mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})\}\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)_j + \sum_{j=1}^{q}\mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})(-n^{-1}\mathbf{\Lambda}^{-1}\mathbf{q}_0^{*} + \mathbf{\Lambda}^{(2)}\mathbf{l}_0^{(2)})_j \\
&\quad + \frac{1}{2}\sum_{j,k=1}^{q}\mathrm{E}_g(\mathbf{G}_{0(j,k)}^{(4)})\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)_j\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)_k\;\underset{(\mathrm{C})}{\Bigg]}\;\underset{(\mathrm{A})}{\Bigg]}
\end{aligned} \tag{A7.1}$$

$$
\begin{aligned}
&= n^{-1} 2 \underset{(\mathrm{A})}{\Bigg[} \ \mathrm{vec}'(\boldsymbol{\Lambda}^{-1}) n \mathrm{E}_g (\mathbf{M}_{\mathbf{G}} \otimes \mathbf{M}) \mathrm{vec}(\boldsymbol{\Lambda}^{-1}) \\
&\quad - \sum_{a,b,c=1}^{q} (\boldsymbol{\Lambda}^{-1})_{a\bullet} \mathrm{E}_g (\mathbf{J}_0^{(3)}) \{ (\boldsymbol{\Lambda}^{-1})_{\bullet b} \otimes (\boldsymbol{\Lambda}^{-1})_{\bullet c} \} n \mathrm{E}_g \left\{ \frac{\partial \bar{l}}{\partial \theta_{0c}} (\mathbf{M}_{\mathbf{G}})_{ba} \right\} \\
&\quad - \sum_{a,b,c,d,e=1}^{q} \lambda^{ac} \lambda^{db} \sum_{j=1}^{q} \mathrm{E}_g (\mathbf{G}_{0(j)}^{(3)})_{ab} \lambda^{je} n \mathrm{E}_g \left(m_{cd} \frac{\partial \bar{l}}{\partial \theta_{0e}} \right) \\
&\quad + \sum_{a,b,c,d=1}^{q} (\boldsymbol{\Lambda}^{-1})_{a\bullet} \mathrm{E}_g (\mathbf{J}_0^{(3)}) \{ (\boldsymbol{\Lambda}^{-1})_{\bullet b} \otimes (\boldsymbol{\Lambda}^{-1})_{\bullet c} \} \mathrm{E}_g (\mathbf{G}_{0(j)}^{(3)})_{ab} \lambda^{jd} \gamma_{cd} \\
&+ \underset{(\mathrm{B})}{\Bigg[} - \mathrm{vec}'(\boldsymbol{\Lambda}^{-1}) n \mathrm{E}_g (\mathbf{M}^{<2>}) \mathrm{vec}(\boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1}) \\
&\quad + 2 \sum_{a,b,c=1}^{q} (\boldsymbol{\Lambda}^{-1})_{a\bullet} \mathrm{E}_g (\mathbf{J}_0^{(3)}) \{ (\boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1})_{\bullet b} \otimes (\boldsymbol{\Lambda}^{-1})_{\bullet c} \} n \mathrm{E}_g \left(m_{ab} \frac{\partial \bar{l}}{\partial \theta_{0c}} \right) \\
&\quad - \sum_{a=1}^{q} \mathrm{tr} \left[n \mathrm{E}_g \left\{ \{ \mathbf{J}_0^{(3)} - \mathrm{E}_g (\mathbf{J}_0^{(3)}) \} \frac{\partial \bar{l}}{\partial \theta_{0a}} \right\} \{ (\boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1}) \otimes (\boldsymbol{\Lambda}^{-1})_{\bullet a} \} \right] \\
&\quad + \mathrm{tr} [\mathrm{E}_g (\mathbf{J}_0^{(3)}) \{ (\boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1}) \otimes \boldsymbol{\alpha}_{\mathrm{W}1} \}] \\
&\quad - \sum_{a,b=1}^{q} \mathrm{tr} [\boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1} \mathrm{E}_g (\mathbf{J}_0^{(3)}) \{ \boldsymbol{\Lambda}^{-1} \otimes (\boldsymbol{\Lambda}^{-1})_{\bullet a} \} \mathrm{E}_g (\mathbf{J}_0^{(3)}) \{ \boldsymbol{\Lambda}^{-1} \otimes (\boldsymbol{\Lambda}^{-1})_{\bullet b} \}] \gamma_{ab} \\
&\quad + \frac{1}{2} \mathrm{vec}' \mathrm{E}_g (\mathbf{J}_0^{(4)}) \{ \mathrm{vec}(\boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1}) \}^{<2>} \underset{(\mathrm{B})}{\Bigg]} \\
&\quad - \underset{(\mathrm{C})}{\Bigg[} - \sum_{a,b=1}^{q} \mathrm{tr} \, n \mathrm{E}_g \left[\boldsymbol{\Lambda}^{-1} \{ \mathbf{G}_{0(a)}^{(3)} - \mathrm{E}_g (\mathbf{G}_{0(a)}^{(3)}) \} \frac{\partial \bar{l}}{\partial \theta_{0b}} \right] \lambda^{ab} \\
&\quad + \sum_{j=1}^{q} \mathrm{tr} \{ \boldsymbol{\Lambda}^{-1} \mathrm{E}_g (\mathbf{G}_{0(j)}^{(3)}) (\boldsymbol{\alpha}_{\mathrm{W}1})_j \} + \frac{1}{2} \sum_{j,k=1}^{q} \mathrm{tr} \{ \boldsymbol{\Lambda}^{-1} \mathrm{E}_g (\mathbf{G}_{0(j,k)}^{(4)}) (\boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Lambda}^{-1})_{jk} \} \underset{(\mathrm{C})}{\Bigg]} \underset{(\mathrm{A})}{\Bigg]} \\
&\quad \equiv n^{-1} d^{(\mathrm{T}1)},
\end{aligned}
$$

where $n^{-1}\boldsymbol{\alpha}_{\mathrm{W}1} \equiv -n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \mathrm{E}_g(\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)})$ is the vector of the asymptotic biases of $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ up to order $O(n^{-1})$ under possible model misspecification.

On the other hand,

$$\mathrm{E}_g\{2(\mathrm{tr}_{\Delta\Delta}^{(\mathrm{T2})})\}$$

$$= 2\mathrm{E}_g\{\mathrm{tr}(-\boldsymbol{\Lambda}_{\mathbf{I}}^{-1(\Delta)}\boldsymbol{\Gamma}_{\mathbf{I}}^{(\Delta)} - \boldsymbol{\Lambda}_{\mathbf{I}}^{-1(\Delta\Delta)}\boldsymbol{\Gamma} - \boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}_{\mathbf{I}}^{(\Delta\Delta)})\}$$

$$= n^{-1}2\mathrm{E}_g\mathrm{tr}\underset{(\mathrm{A})}{\Bigg[}\; n\Bigg[-\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\boldsymbol{\Lambda}^{-1}\otimes\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\right\}\left\{-\sum_{j=1}^{q}\mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)_j\right\}\Bigg]$$

$$+ n\underset{(\mathrm{B})}{\Bigg[}\;\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\boldsymbol{\Lambda}^{-1}\otimes(-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^{*} + \boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)})\right\} \tag{A7.2}$$

$$-\boldsymbol{\Lambda}^{-1}\left[\mathrm{E}_g(\mathbf{J}_0^{(3)})\left\{\boldsymbol{\Lambda}^{-1}\otimes\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)\right\}\right]^2$$

$$+\frac{1}{2}\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(4)})\left\{\boldsymbol{\Lambda}^{-1}\otimes\left(\boldsymbol{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\boldsymbol{\theta}_0}\right)^{<2>}\right\}\underset{(\mathrm{B})}{\Bigg]}\boldsymbol{\Gamma}\underset{(\mathrm{A})}{\Bigg]} - n^{-1}2\mathrm{E}_g\{\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}_{\mathbf{I}}^{(\Delta\Delta)})\}$$

$$= n^{-1}2\underset{(\mathrm{A})}{\Bigg[}\sum_{a,b,c,d=1}^{q}(\boldsymbol{\Lambda}^{-1})_{a\cdot}\mathrm{E}_g(\mathbf{J}_0^{(3)})\{(\boldsymbol{\Lambda}^{-1})_{\cdot b}\otimes(\boldsymbol{\Lambda}^{-1})_{\cdot c}\}\sum_{j=1}^{q}\mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})_{ab}\lambda^{jd}\gamma_{cd}$$

$$+\underset{(\mathrm{B})}{\Bigg[}\;\mathrm{tr}[\mathrm{E}_g(\mathbf{J}_0^{(3)})\{(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1})\otimes\boldsymbol{\alpha}_{\mathrm{W1}}\}]$$

$$-\sum_{a,b=1}^{q}\mathrm{tr}[\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{J}_0^{(3)})\{\boldsymbol{\Lambda}^{-1}\otimes(\boldsymbol{\Lambda}^{-1})_{\cdot a}\}\mathrm{E}_g(\mathbf{J}_0^{(3)})\{\boldsymbol{\Lambda}^{-1}\otimes(\boldsymbol{\Lambda}^{-1})_{\cdot b}\}\gamma_{ab}$$

$$+\frac{1}{2}\mathrm{vec}'\mathrm{E}_g(\mathbf{J}_0^{(4)})\{\mathrm{vec}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1})\}^{<2>}\underset{(\mathrm{B})}{\Bigg]}$$

$$-\underset{(\mathrm{C})}{\Bigg[}\sum_{j=1}^{q}\mathrm{tr}\{\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{G}_{0(j)}^{(3)})(\boldsymbol{\alpha}_{\mathrm{W1}})_j\} + \frac{1}{2}\sum_{j,k=1}^{q}\mathrm{tr}\{\boldsymbol{\Lambda}^{-1}\mathrm{E}_g(\mathbf{G}_{0(j,k)}^{(4)})\}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1})_{jk}\underset{(\mathrm{C})}{\Bigg]}\underset{(\mathrm{A})}{\Bigg]}$$

$$\equiv n^{-1}d^{(\mathrm{T2})}.$$

A8. The derivation and actual expressions of $(\bar{l}_{\mathrm{ML}}^{(j)})_{O_p(n^{-j/2})}$ $(j = 1,\ldots,4)$

The five terms up to order $O_p(n^{-2})$ in the last expression of (4.3) are further expanded one by one as follows:

(i)

$$-2(\bar{l}_0)_{O_p(1)} = -2\mathrm{E}_g(\bar{l}_0) - 2\{\bar{l}_0 - \mathrm{E}_g(\bar{l}_0)\}$$
$$\equiv -2(\bar{l}_0^*)_{O(1)} - 2(\bar{l}_0 - \bar{l}_0^*)_{O_p(n^{-1/2})},$$

(ii)

$$-2\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'}\left(-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{k=1}^{3}\boldsymbol{\Lambda}^{(k)}\mathbf{l}_0^{(k)} + n^{-1}\mathbf{l}_0^{(\mathrm{W})}\right)$$
$$= 2\left(n^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^*\right)_{O_p(n^{-3/2})} + 2\left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'}\boldsymbol{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)_{O_p(n^{-1})} - 2\left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'}\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)}\right)_{O_p(n^{-3/2})} \qquad \text{(A8.1)}$$
$$-2\left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'}\boldsymbol{\Lambda}^{(3)}\mathbf{l}_0^{(3)}\right)_{O_p(n^{-2})} - 2\left(n^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'}\mathbf{l}_0^{(\mathrm{W})}\right)_{O_p(n^{-2})},$$

(iii)

$$-\left\{\frac{\partial^2 \bar{l}}{(\partial \boldsymbol{\theta}_0{}')^{<2>}}\right\}_{O_p(1)}\left(-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{k=1}^{3}\boldsymbol{\Lambda}^{(k)}\mathbf{l}_0^{(k)} + n^{-1}\mathbf{l}_0^{(\mathrm{W})}\right)^{<2>}$$
$$= -\mathrm{vec}'\{\boldsymbol{\Lambda} + (\mathbf{M})_{O_p(n^{-1/2})}\}\left(-n^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^* + \sum_{k=1}^{3}\boldsymbol{\Lambda}^{(k)}\mathbf{l}_0^{(k)} + n^{-1}(\mathbf{l}_0^{(\mathrm{W})})\right)^{<2>}$$
$$= -\{n^{-2}\mathrm{vec}'(\boldsymbol{\Lambda})(\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^*)^{<2>}\}_{O(n^{-2})} - 2\left[n^{-1}\mathrm{vec}'(\boldsymbol{\Lambda})\left\{(\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^*)\otimes\left(\boldsymbol{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\right\}\right]_{O_p(n^{-3/2})}$$
$$+2[n^{-1}\mathrm{vec}'(\boldsymbol{\Lambda})\{(\boldsymbol{\Lambda}^{-1}\mathbf{q}_0^*)\otimes(\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)})\}]_{O_p(n^{-2})} - \left\{\mathrm{vec}'(\boldsymbol{\Lambda})\left(\boldsymbol{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)^{<2>}\right\}_{O_p(n^{-1})}$$
$$+2\left[\mathrm{vec}'(\boldsymbol{\Lambda})\left\{\left(\boldsymbol{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\otimes(\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)})\right\}\right]_{O_p(n^{-3/2})}$$
$$+2\left[\mathrm{vec}'(\boldsymbol{\Lambda})\left\{\left(\boldsymbol{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\otimes(\boldsymbol{\Lambda}^{(3)}\mathbf{l}_0^{(3)})\right\}\right]_{O_p(n^{-2})}$$
$$-\{\mathrm{vec}'(\boldsymbol{\Lambda})(\boldsymbol{\Lambda}^{(2)}\mathbf{l}_0^{(2)})^{<2>}\}_{O_p(n^{-2})} + 2\left[n^{-1}\mathrm{vec}'(\boldsymbol{\Lambda})\left\{\left(\boldsymbol{\Lambda}^{-1}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0}\right)\otimes\mathbf{l}_0^{(\mathrm{W})}\right\}\right]_{O_p(n^{-2})}$$

$$-2\left[n^{-1}\mathrm{vec}'(\mathbf{M})\left\{(\mathbf{\Lambda}^{-1}\mathbf{q}_0^*)\otimes\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)\right\}\right]_{O_p(n^{-2})}$$

$$-\left\{\mathrm{vec}'(\mathbf{M})\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)^{<2>}\right\}_{O_p(n^{-3/2})}+2\left[\mathrm{vec}'(\mathbf{M})\left\{\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)\otimes(\mathbf{\Lambda}^{(2)}\mathbf{l}_0^{(2)})\right\}\right]_{O_p(n^{-2})},$$

(iv)

$$-\frac{1}{3}\frac{\partial^3\bar{l}}{(\partial\mathbf{\theta}_0{}')^{<3>}}\left(-n^{-1}\mathbf{\Lambda}^{-1}\mathbf{q}_0^*+\sum_{k=1}^{2}\mathbf{\Lambda}^{(k)}\mathbf{l}_0^{(k)}\right)^{<3>}$$

$$=-\frac{1}{3}\mathrm{vec}'[\mathrm{E}_g(\mathbf{J}_0^{(3)})+\{\mathbf{J}_0^{(3)}-\mathrm{E}_g(\mathbf{J}_0^{(3)})\}_{O_p(n^{-1/2})}]\left(-n^{-1}\mathbf{\Lambda}^{-1}\mathbf{q}_0^*+\sum_{k=1}^{2}\mathbf{\Lambda}^{(k)}\mathbf{l}_0^{(k)}\right)^{<3>}$$

$$=\left[n^{-1}\mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\}\left\{(\mathbf{\Lambda}^{-1}\mathbf{q}_0^*)\otimes\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)^{<2>}\right\}\right]_{O_p(n^{-2})}$$

$$+\frac{1}{3}\left[\mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\}\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)^{<3>}\right]_{O_p(n^{-3/2})}$$

$$-\left[\mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\}\left\{\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)^{<2>}\otimes(\mathbf{\Lambda}^{(2)}\mathbf{l}_0^{(2)})\right\}\right]_{O_p(n^{-2})}$$

$$+\frac{1}{3}\left[\mathrm{vec}'\{\mathbf{J}_0^{(3)}-\mathrm{E}_g(\mathbf{J}_0^{(3)})\}\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)^{<3>}\right]_{O_p(n^{-2})},$$

(v)

$$-\frac{1}{12}\mathrm{E}_g\left\{\frac{\partial^4\bar{l}}{(\partial\mathbf{\theta}_0{}')^{<4>}}\right\}(\mathbf{\Lambda}^{(1)}\mathbf{l}_0^{(1)})^{<4>}=-\frac{1}{12}\left[\mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(4)})\}\left(\mathbf{\Lambda}^{-1}\frac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)^{<4>}\right]_{O_p(n^{-2})}.$$

Using $\dfrac{\partial\bar{l}}{\partial\mathbf{\theta}_0{}'}\mathbf{\Lambda}^{-1}\dfrac{\partial\bar{l}}{\partial\mathbf{\theta}_0}=\mathrm{vec}'(\mathbf{\Lambda})\left(\mathbf{\Lambda}^{-1}\dfrac{\partial\bar{l}}{\partial\mathbf{\theta}_0}\right)^{<2>}$ and similar results in (A8.1), (4.3) becomes

$$
\begin{aligned}
-2\hat{\bar{l}}_{\mathrm{W}} &= -2(\bar{l}_0^{*})_{O(1)} - 2(\bar{l}_0 - \bar{l}_0^{*})_{O_p(n^{-1/2})} + \left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'} \boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)_{O_p(n^{-1})} \\
&+ \underset{(\mathrm{A})}{\Bigg[} \; 2n^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'} \boldsymbol{\Lambda}^{-1} \mathbf{q}_0^{*} - 2 \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'} \boldsymbol{\Lambda}^{(2)} \mathbf{l}_0^{(2)} - 2n^{-1} \mathrm{vec}'(\boldsymbol{\Lambda}) \left\{ (\boldsymbol{\Lambda}^{-1} \mathbf{q}_0^{*}) \otimes \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \right\} \\
&\quad + 2\mathrm{vec}'(\boldsymbol{\Lambda}) \left\{ \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \otimes (\boldsymbol{\Lambda}^{(2)} \mathbf{l}_0^{(2)}) \right\} - \mathrm{vec}'(\mathbf{M}) \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)^{<2>} \\
&\quad + \frac{1}{3} \mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\} \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)^{<3>} \; \Bigg]_{(\mathrm{A})\, O_p(n^{-3/2})} \\
&- \{ n^{-2} \mathrm{vec}'(\boldsymbol{\Lambda}) (\boldsymbol{\Lambda}^{-1} \mathbf{q}_0^{*})^{<2>} \}_{O(n^{-2})} \\
&+ \underset{(\mathrm{B})}{\Bigg[} \; \underbrace{-2 \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'} \boldsymbol{\Lambda}^{(3)} \mathbf{l}_0^{(3)}}_{(1)} \underbrace{- 2n^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'} \mathbf{l}_0^{(\mathrm{W})}}_{(2)} + \underbrace{2n^{-1} \mathrm{vec}'(\boldsymbol{\Lambda}) \{ (\boldsymbol{\Lambda}^{-1} \mathbf{q}_0^{*}) \otimes (\boldsymbol{\Lambda}^{(2)} \mathbf{l}_0^{(2)}) \}}_{(3)} \\
&\quad + \underbrace{2\mathrm{vec}'(\boldsymbol{\Lambda}) \left\{ \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \otimes (\boldsymbol{\Lambda}^{(3)} \mathbf{l}_0^{(3)}) \right\}}_{-(1)} \underbrace{- \mathrm{vec}'(\boldsymbol{\Lambda}) (\boldsymbol{\Lambda}^{(2)} \mathbf{l}_0^{(2)})^{<2>}}_{(4)} \\
&\quad + \underbrace{2n^{-1} \mathrm{vec}'(\boldsymbol{\Lambda}) \left\{ \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \otimes \mathbf{l}_0^{(\mathrm{W})} \right\}}_{-(2)} \underbrace{- 2n^{-1} \mathrm{vec}'(\mathbf{M}) \left\{ (\boldsymbol{\Lambda}^{-1} \mathbf{q}_0^{*}) \otimes \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \right\}}_{-(3)} \\
&\quad \underbrace{+ 2\mathrm{vec}'(\mathbf{M}) \left\{ \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right) \otimes (\boldsymbol{\Lambda}^{(2)} \mathbf{l}_0^{(2)}) \right\}}_{-2\times(4)} \underbrace{+ n^{-1} \mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\} \left\{ (\boldsymbol{\Lambda}^{-1} \mathbf{q}_0^{*}) \otimes \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)^{<2>} \right\}}_{-(3)} \\
&\quad \underbrace{- \mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\} \left\{ \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)^{<2>} \otimes (\boldsymbol{\Lambda}^{(2)} \mathbf{l}_0^{(2)}) \right\}}_{-2\times(4)} + \frac{1}{3} \mathrm{vec}'\{\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})\} \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)^{<3>} \\
&\quad - \frac{1}{12} \mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(4)})\} \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0} \right)^{<4>} \; \Bigg]_{(\mathrm{B})\, O_p(n^{-2})} + O_p(n^{-5/2})
\end{aligned}
\tag{A8.2}
$$

$$= -2(\overline{l}_0^*)_{O(1)} - 2(\overline{l}_0 - \overline{l}_0^*)_{O_p(n^{-1/2})} + \left(\frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0{}'} \boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right)_{O_p(n^{-1})}$$

$$+ \left[-\mathrm{vec}'(\mathbf{M}) \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right)^{<2>} + \frac{1}{3} \mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\} \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right)^{<3>} \right]_{O_p(n^{-3/2})}$$

$$-(n^{-2} \mathbf{q}_0^*{}' \boldsymbol{\Lambda}^{-1} \mathbf{q}_0^*)_{O(n^{-2})}$$

$$+ \left[\mathrm{vec}'(\boldsymbol{\Lambda})(\boldsymbol{\Lambda}^{(2)} \mathbf{l}_0^{(2)})^{<2>} + \frac{1}{3} \mathrm{vec}'\{\mathbf{J}_0^{(3)} - \mathrm{E}_g(\mathbf{J}_0^{(3)})\} \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right)^{<3>} \right.$$

$$\left. - \frac{1}{12} \mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(4)})\} \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right)^{<4>} \right]_{O_p(n^{-2})} + O_p(n^{-5/2})$$

$$\equiv -2(\overline{l}_0^*)_{O(1)} + \sum_{j=1}^{4} (\overline{l}_{\mathrm{ML}}^{(j)})_{O_p(n^{-j/2})} - (n^{-2} \mathbf{q}_0^*{}' \boldsymbol{\Lambda}^{-1} \mathbf{q}_0^*)_{O(n^{-2})} + O_p(n^{-5/2})$$

$$(\overline{l}_{\mathrm{W}}^{(j)} = \overline{l}_{\mathrm{ML}}^{(j)}, j = 1, ..., 4),$$

where the underline with a number in parentheses indicates a quantity and the negative number e.g., "$-a \times (4)...$" indicates $-a$ times the quantity which has the sign "(4)…" when the quantities with "$-a \times (4)...$" are summed.

In the last result of (A8.2), the first term for $\overline{l}_{\mathrm{W}}^{(4)}$ can also be written as

$$\begin{aligned}
&\mathrm{vec}'(\boldsymbol{\Lambda})(\boldsymbol{\Lambda}^{(2)} \mathbf{l}_0^{(2)})^{<2>} \\
&= \mathrm{vec}'(\mathbf{M}) \left\{ \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right) \otimes \left(\boldsymbol{\Lambda}^{-1} \mathbf{M} \boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right) \right\} \\
&\quad - \mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\} \left\{ \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right)^{<2>} \otimes \left(\boldsymbol{\Lambda}^{-1} \mathbf{M} \boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right) \right\} \qquad \text{(A8.3)} \\
&\quad + \frac{1}{4} \mathrm{vec}'\{\mathrm{E}_g(\mathbf{J}_0^{(3)})\} \left[\left(\boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right)^{<2>} \otimes \left\{ \boldsymbol{\Lambda}^{-1} \mathrm{E}_g(\mathbf{J}_0^{(3)}) \left(\boldsymbol{\Lambda}^{-1} \frac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0} \right)^{<2>} \right\} \right]
\end{aligned}$$

(recall (A1.2)).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *Proceedings of the 2nd international symposium on information theory* (pp. 267-281). Budapest: Académiai Kiado.

Amos, D. E. (1983). ALGORITHM 610 A portable FORTRAN subroutine for derivatives of the psi function. *ACM Transactions on Mathematical Software, 9*, 494–502.

Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference – A practical information-theoretic approach* (2nd ed.). New York: Springer.

Hilbe, J. M. (2011). *Negative binomial regression* (2nd ed.). Cambridge: Cambridge University Press.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*, 297-307.

Ishiguro, M., Sakamoto, Y., & Kitagawa, G. (1997). Bootstrapping log-likelihood and EIC, An extension of AIC. *Annals of the Institute of Statistical Mathematics, 49*, 411-434.

Kamo, K., Yanagihara, H., & Satoh, K. (2013). Bias-corrected AIC for selecting variables in Poisson regression models. *Communications in Statistics – Theory and Methods, 42*, 1911-1921.

Kishino, H., & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution, 29*, 170-179.

Konishi, S., & Kitagawa, G. (1996). Generalized information criteria in model selection. *Biometrika, 83*, 875-890.

Konishi, S., & Kitagawa, G. (2003). Asymptotic theory for information criteria in model selection – functional approach. *Journal of Statistical Planning and Inference, 114*, 45-61.

Konishi, S., & Kitagawa, G. (2008). *Information criteria and statistical modeling*. New-York: Springer.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Statistics, 22*, 79-86.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics, 15*, 209-225.

Linhart, H. (1988). A test whether two AIC's differ significantly. *South African Statistical*

Journal, 22, 153-161.

Linhart, H., & Zucchini, W. (1986). *Model selection*. New York: Wiley.

Ogasawara, H. (1995). The gamma-gamma regression for the distribution model of event times. *Japanese Psychological Research, 37*, 70-79.

Ogasawara, H. (2010). Asymptotic expansions for the pivots using log-likelihood derivatives with an application in item response theory. *Journal of Multivariate Analysis, 101*, 2149-2167.

Ogasawara, H. (2012). Cornish-Fisher expansions using sample cumulants and monotonic transformations. *Journal of Multivariate Analysis, 103*, 1-18.

Ogasawara, H. (2014). Optimization of the Gaussian and Jeffreys power priors with emphasis on the canonical parameters in the exponential family. *Behaviormetrika, 41,* 195-223.

Ogasawara, H. (2015a). Bias adjustment minimizing the asymptotic mean square error. *Communications in Statistics – Theory and Methods, 44,* 3503-3522.

Ogasawara, H. (2015b). An expository supplement to the paper "Bias adjustment minimizing the asymptotic mean square error" with errata. *Economic Review* (*Otaru University of Commerce), 65* (4), 121-130. http://www.res.otaru-uc.ac.jp/~hogasa/, http://barrel.ih.otaru-uc.ac.jp/.

Ogasawara, H. (2016a). Supplement I to the paper "Asymptotic cumulants of some information criteria" – Asymptotic cumulants of the studentized information criteria and Example 1. To appear in *Economic Review* (*Otaru University of Commerce*).

Ogasawara, H. (2016b). Supplement II to the paper "Asymptotic cumulants of some information criteria" – Examples 2 and 3. To appear in *Economic Review (Otaru University of Commerce*).

R Core Team R (2015). *R: A language and environment for statistical computing*. Vienna : R Foundation for Statistical Computing. URL https://www.R-project.org/.

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Dordrecht: Reidel.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464.

Shibata, R. (1989). Statistical aspects of model selection. In J. C. Willems (Ed,), *From data to model* (pp.215-240). Berlin: Springer.

Shimodaira, H. (1997). Assessing the error probability of the model selection test. *Annals of the Institute of Statistical Mathematics, 49*, 395-410.

Shimodaira, H., & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution, 16*, 1114-1116.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50*, 253-264.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society, B, 39*, 44-47.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics – Theory and Methods, 7*, 13-26.

Takeuchi, K. (1976). Distributions of information statistics and criteria of the goodness of models. *Mathematical Science, 153*, 12-18. (in Japanese)

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *The Annals of Mathematical Statistics, 18*, 309-348.

Withers, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *The Annals of Statistics, 11*, 577-587.

Yanagihara, H., Sekiguchi, R., & Fujikoshi, Y. (2003). Bias correction of AIC in logistic regression models. *Journal of Statistical Planning and Inference, 115*, 349-360.

Yanagihara, H., & Ohmoto, C. (2005). On distribution of AIC in linear regression models. *Journal of Statistical Planning and Inference, 133*, 417-433.

Table 1. Asymptotic cumulants of $n^{-1}\mathrm{AIC}_{\mathrm{ML(W)}}$ and $n^{-1}\mathrm{TIC}_{\mathrm{ML(W)}}^{(j)}$ $(j=1,2)$ before studentization

	Example 1	Example 2	Example 3
Model distribution	Exponential	normal with known σ^2	Bernoulli
True distribution	gamma, $\alpha \neq 1$	non-normal	Bernoulli
Parameter	canonical (the reciprocal of the scale)	canonical (mean)	canonical (logit)
AIC	$n^{-1}\mathrm{AIC}_{\mathrm{ML}}$	$n^{-1}\mathrm{AIC}_{\mathrm{ML}}(=n^{-1}\mathrm{TIC}_{\mathrm{ML}}^{(\bullet)})$	$n^{-1}\mathrm{AIC}_{\mathrm{W}}(=n^{-1}\mathrm{TIC}_{\mathrm{W}}^{(\bullet)})$
$\alpha_{\mathrm{ML(W)1}}^{(\mathrm{A})}$	$2-\alpha^{-1}$	1	1
$\alpha_{\mathrm{ML(W)\Delta 1}}^{(\mathrm{A})}$	$-(1/6)\alpha^{-2}$	0	$(1/6)(1-\overline{i}_0^{-1})$ $+(a^2/4)(1-2\pi_0)^2\overline{i}_0^{-1}$
$\alpha_{\mathrm{ML(W)1}}^{(\mathrm{A})*}$	$2-2\alpha^{-1}$	0	0
$\alpha_{\mathrm{ML(W)\Delta 1}}^{(\mathrm{A})*}$	$-2\alpha^{-2}$	0	$(a-1)\{(1-2\pi_0)^2\overline{i}_0^{-1}+2\}$
$\alpha_{\mathrm{ML(W)2}}^{(\mathrm{A})}$	$4\alpha^{-1}$	κ_4+2	$4\theta_0^2\overline{i}_0$
$\alpha_{\mathrm{ML(W)\Delta 2}}^{(\mathrm{A})}$	$2\alpha^{-2}$	$-2(\kappa_4+1)$	2
$\alpha_{\mathrm{ML(W)3}}^{(\mathrm{A})}$	$-8\alpha^{-2}$	$\kappa_6+12\kappa_4+4\kappa_3^2+8$	$-\{8\theta_0^3(1-2\pi_0)$ $+24\theta_0^2\}\overline{i}_0$
$\alpha_{\mathrm{ML(W)4}}^{(\mathrm{A})}$	$32\alpha^{-3}$	$\kappa_8+24\kappa_6+32\kappa_5\kappa_3$ $+32\kappa_4^2+144\kappa_4$ $+96\kappa_3^2+48$	$\{16\theta_0^4(1-6\pi_0+6\pi_0^2)$ $+128\theta_0^3(1-2\pi_0)$ $+192\theta_0^2\}\overline{i}_0$
Higher-order bias correction	(see the case of $n^{-1}\mathrm{TIC}_{\mathrm{ML}}^{(\bullet)}$ below)	$n^{-1}\mathrm{AIC}_{\mathrm{ML}}(=n^{-1}\mathrm{TIC}_{\mathrm{ML}}^{(\bullet)})$ is unbiased	$n^{-1}\mathrm{AIC}_{\mathrm{W}\to O(n^{-2})}$ $=n^{-1}\mathrm{TIC}_{\mathrm{W}\to O(n^{-2})}^{(\bullet)}$ $=-2\hat{\overline{l}}_{\mathrm{W}}+n^{-1}2$ $+n^{-2}(1-a)[(1-2\overline{x})^2$ $\times\{\overline{x}(1-\overline{x})\}^{-1}+2]$

(to be continued)

Table 1. (continued)

	Example 1 ($\alpha_{\mathrm{ML}\Delta 2}^{(\mathrm{T}\bullet)} = \alpha_{\mathrm{ML}\Delta 2}^{(\mathrm{A})}$ in this example)
TIC	$n^{-1}\mathrm{TIC}_{\mathrm{ML}}^{(\bullet)}$
$\alpha_{\mathrm{ML}1}^{(\mathrm{T}\bullet)}$	α^{-1}
$\alpha_{\mathrm{ML}\Delta 1}^{(\mathrm{T}\bullet)}$	$-\dfrac{1}{6\alpha^2} + \dfrac{\alpha\psi''(\alpha) + \psi'(\alpha)}{\alpha\{\alpha\psi'(\alpha) - 1\}^2}$
$\alpha_{\mathrm{ML}1}^{(\mathrm{T}\bullet)*}$	0
$\alpha_{\mathrm{ML}\Delta 1}^{(\mathrm{T}\bullet)*}$	$-\dfrac{2}{\alpha^2} + \dfrac{\alpha\psi''(\alpha) + \psi'(\alpha)}{\alpha\{\alpha\psi'(\alpha) - 1\}^2}$
Higher-order bias correction	$n^{-1}\mathrm{TIC}_{\mathrm{ML}\to O(n^{-2})}^{(\bullet)} = -2\hat{\bar{l}}_{\mathrm{ML}} + n^{-1}\dfrac{2}{\hat{\alpha}} + n^{-2}\left[\dfrac{2}{\hat{\alpha}^2} - \dfrac{\hat{\alpha}\psi''(\hat{\alpha}) + \psi'(\hat{\alpha})}{\hat{\alpha}\{\hat{\alpha}\psi'(\hat{\alpha}) - 1\}^2}\right]$

Note. $\alpha_{\mathrm{ML}\,j}^{(\mathrm{A})}$ are for $\kappa_{g\,j}(n^{-1}\mathrm{AIC}_{\mathrm{ML}} + 2\hat{\bar{l}}_0^*)$ while $\alpha_{\mathrm{ML}\,j}^{(\mathrm{A})*}$ are for $\kappa_{g\,j}\{n^{-1}\mathrm{AIC}_{\mathrm{ML}} + 2\mathrm{E}_g(\hat{\bar{l}}_{\mathrm{ML}}^*)\}$ $(j = 1,\ \Delta 1,\ 2,\ \Delta 2,\ 3,\ 4)$ in Examples 1 and 2. Similarly, $\alpha_{\mathrm{W}\,j}^{(\mathrm{A})}$ and $\alpha_{\mathrm{W}\,j}^{(\mathrm{A})*}$ are defined in Example 3. $\psi'(\cdot)$ and $\psi''(\cdot)$ are the first and second derivatives of the digamma function $\psi(\cdot)$, respectively. Generally, $\alpha_{\mathrm{ML}\,j}^{(\mathrm{A})} = \alpha_{\mathrm{ML}\,j}^{(\mathrm{A})*}, \alpha_{\mathrm{W}\,j}^{(\mathrm{A})} = \alpha_{\mathrm{W}\,j}^{(\mathrm{A})*}$ $(j = 2,\ \Delta 2,\ 3,\ 4)$, $\alpha_{\mathrm{W}\,j}^{(\mathrm{A})} = \alpha_{\mathrm{ML}\,j}^{(\mathrm{A})}$ $(j = 1,\ 2,\ \Delta 2,\ 3,\ 4)$, $\alpha_{\mathrm{W}\Delta 1}^{(\mathrm{A})} \neq \alpha_{\mathrm{ML}\Delta 1}^{(\mathrm{A})}$ and $\alpha_{\mathrm{W}\,j}^{(\mathrm{T}\bullet)} = \alpha_{\mathrm{ML}\,j}^{(\mathrm{T}\bullet)} = \alpha_{\mathrm{W}\,j}^{(\mathrm{A})} = \alpha_{\mathrm{ML}\,j}^{(\mathrm{A})}$ $(j = 2,\ 3,\ 4)$.

In Example 1, $\kappa_j \equiv \kappa_{g\,j}\{(x^* - \mu_0)/\sigma\}$ and in Example 3 $\bar{i}_0 = \pi_0(1 - \pi_0)$ is the population Fisher information per observation.

Table 2. Asymptotic cumulants of $n^{-1}\mathrm{AIC}_{\mathrm{ML(W)}}$ and $n^{-1}\mathrm{TIC}^{(j)}_{\mathrm{ML(W)}}$ $(j=1, 2)$ after studentization

	Example 1	Example 2	Example 3
AIC	$n^{-1}\mathrm{AIC}_{\mathrm{ML}}$	$n^{-1}\mathrm{AIC}_{\mathrm{ML}}(=n^{-1}\mathrm{TIC}^{(\bullet)}_{\mathrm{ML}})$	$n^{-1}\mathrm{AIC}_{\mathrm{W}}(=n^{-1}\mathrm{TIC}^{(\bullet)}_{\mathrm{W}})$
$\alpha^{(\mathrm{A})}_{(t)\mathrm{ML(W)}1}$	$\alpha^{1/2}-(1/2)\alpha^{-1/2}$	$(\kappa_4+2)^{-1/2}-(1/2)(\kappa_4+2)^{-3/2}$ $\times(\kappa_6+12\kappa_4+6\kappa_3^2+8)$	$\{(3/2)\theta_0^{-1}$ $+(1/2)(1-2\pi_0)\}\bar{i}_0^{-1/2}$
$\alpha^{(\mathrm{A})*}_{(t)\mathrm{ML(W)}1}$	$\alpha^{1/2}-\alpha^{-1/2}$	$-(1/2)(\kappa_4+2)^{-3/2}$ $\times(\kappa_6+12\kappa_4+6\kappa_3^2+8)$	$\{\theta_0^{-1}+(1/2)(1-2\pi_0)\}\bar{i}_0^{-1/2}$
$\alpha^{(\mathrm{A})}_{(t)\mathrm{ML(W)}2}$	1	1	1
$\alpha^{(\mathrm{A})}_{(t)\mathrm{ML(W)}\Delta 2}$	$(7/2)\alpha^{-1}+2$	$2-2(\kappa_4+2)^{-1}+(\kappa_4+2)^{-2}(-\kappa_6$ $+8\kappa_5\kappa_3+2\kappa_4^2-4\kappa_4+50\kappa_3^2)$ $+(\kappa_4+2)^{-3}(7/4)$ $\times(\kappa_6+12\kappa_4+6\kappa_3^2+8)^2$	$\left\{\frac{7}{4}(1-2\pi_0)^2+\left(-\frac{a}{4}+\frac{9}{2}\right)\right.$ $\left.\times\theta_0^{-1}(1-2\pi_0)+\frac{11}{2}\theta_0^{-2}\right\}\bar{i}_0^{-1}$ $+2$
$\alpha^{(\mathrm{A})*}_{(t)\mathrm{ML(W)}\Delta 2}$	$(7/2)\alpha^{-1}+2$ $(=\alpha^{(\mathrm{A})}_{(t)\mathrm{ML}\Delta 2}$ in Example 1)	$\alpha^{(\mathrm{A})}_{(t)\mathrm{ML}\Delta 2}+(\kappa_4+2)^{-2}$ $\times(\kappa_6+12\kappa_4+6\kappa_3^2+8)$	$\left\{\frac{7}{4}(1-2\pi_0)^2+\left(-\frac{a}{4}+4\right)\right.$ $\left.\times\theta_0^{-1}(1-2\pi_0)+\frac{9}{2}\theta_0^{-2}\right\}\bar{i}_0^{-1}$ $+2$
$\alpha^{(\mathrm{A})}_{(t)\mathrm{ML(W)}3}$	$-\alpha^{-1/2}$	$-2(\kappa_4+2)^{-3/2}$ $\times(\kappa_6+12\kappa_4+7\kappa_3^2+8)$	$\{2(1-2\pi_0)+3\theta_0^{-1}\}\bar{i}_0^{-1/2}$
$\alpha^{(\mathrm{A})}_{(t)\mathrm{ML(W)}4}$	$8\alpha^{-1}+6$	$12-18(\kappa_4+2)^{-1}+(\kappa_4+2)^{-2}$ $\times(-2\kappa_8-48\kappa_6-64\kappa_5\kappa_3-70\kappa_4^2$ $-294\kappa_4-144\kappa_3^2-84)$ $+(\kappa_4+2)^{-3}$ $\times\{12(\kappa_6+12\kappa_4+6\kappa_3^2+8)^2$ $+12(\kappa_6+12\kappa_4+6\kappa_3^2+8)\kappa_3^2\}$	$\{10(1-2\pi_0)^2$ $+26\theta_0^{-1}(1-2\pi_0)$ $+26\theta_0^{-2}\}\bar{i}_0^{-1}+10$
TIC	$n^{-1}\mathrm{TIC}^{(\bullet)}_{\mathrm{ML}}$		
$\alpha^{(\mathrm{T}\bullet)}_{(t)\mathrm{ML}1}$	$(1/2)\alpha^{-1/2}$		
$\alpha^{(\mathrm{T}\bullet)*}_{(t)\mathrm{ML}1}$	0 (not a general result)		
$\alpha^{(\mathrm{T}\bullet)}_{(t)\mathrm{ML}\Delta 2}$	$(7/2)\alpha^{-1}+2$ $(=\alpha^{(\mathrm{A})}_{(t)\mathrm{ML}\Delta 2}$ in Example 1)		
$\alpha^{(\mathrm{T}\bullet)*}_{(t)\mathrm{ML}\Delta 2}$	$(7/2)\alpha^{-1}+2$ $(=\alpha^{(\mathrm{A})}_{(t)\mathrm{ML}\Delta 2}$ in Example 1)		

Note. Generally, $\alpha^{(\mathrm{A})}_{(t)\mathrm{W}\,j}=\alpha^{(\mathrm{A})}_{(t)\mathrm{ML}\,j}$ $(j=1,\ 2,\ 3,\ 4)$, $\alpha^{(\mathrm{A})*}_{(t)\mathrm{W}2}=\alpha^{(\mathrm{A})*}_{(t)\mathrm{ML}2}=\alpha^{(\mathrm{A})}_{(t)\mathrm{W}2}=\alpha^{(\mathrm{A})}_{(t)\mathrm{ML}2}=1$ and $\alpha^{(\mathrm{A})*}_{(t)\mathrm{W}\,j}=\alpha^{(\mathrm{A})*}_{(t)\mathrm{ML}\,j}=\alpha^{(\mathrm{A})}_{(t)\mathrm{W}\,j}=\alpha^{(\mathrm{A})}_{(t)\mathrm{ML}\,j}$ $(j=3,\ 4)$. Generally, $\alpha^{(\mathrm{T}\bullet)*}_{(t)\mathrm{W}2}=\alpha^{(\mathrm{T}\bullet)*}_{(t)\mathrm{ML}2}=\alpha^{(\mathrm{T}\bullet)}_{(t)\mathrm{W}2}=\alpha^{(\mathrm{T}\bullet)}_{(t)\mathrm{ML}2}=1$ and $\alpha^{(\mathrm{T}\bullet)*}_{(t)\mathrm{W}\,j}=\alpha^{(\mathrm{T}\bullet)*}_{(t)\mathrm{ML}\,j}=\alpha^{(\mathrm{T}\bullet)}_{(t)\mathrm{W}\,j}=\alpha^{(\mathrm{T}\bullet)}_{(t)\mathrm{ML}\,j}=\alpha^{(\mathrm{A})}_{(t)\mathrm{W}\,j}=\alpha^{(\mathrm{A})}_{(t)\mathrm{ML}\,j}$ $(j=3,\ 4)$.

Table 3. c_1 in the higher-order correction term $-n^{-2}\hat{c}_1$ for the n^{-1}AIC under canonical parametrization for one-parameter cases

Distribution	Variance	Skewness (= sk)	Excess kurtosis (= kt)	c_1 $(=-2\mathrm{sk}^2+\mathrm{kt})$
Bernoulli	$\pi_0(1-\pi_0)$	$\frac{(1-2\pi_0)\pi_0(1-\pi_0)}{\{\pi_0(1-\pi_0)\}^{3/2}} = \frac{1-2\pi_0}{\{\pi_0(1-\pi_0)\}^{1/2}}$	$\frac{(1-6\pi_0+6\pi_0^2)\pi_0(1-\pi_0)}{\{\pi_0(1-\pi_0)\}^2} = \frac{1-6\pi_0+6\pi_0^2}{\pi_0(1-\pi_0)}$	$-\frac{1-2\pi_0+2\pi_0^2}{\pi_0(1-\pi_0)}$
Poisson	λ_0	$\lambda_0/\lambda_0^{3/2}=\lambda_0^{-1/2}$	$\lambda_0/\lambda_0^2=\lambda_0^{-1}$	$-\lambda_0^{-1}$
Negative binomial (fixed r)	$\frac{r\pi_0}{(1-\pi_0)^2}$	$\frac{r\pi_0\left\{\frac{1}{(1-\pi_0)^2}+\frac{2\pi_0}{(1-\pi_0)^3}\right\}}{\{r\pi_0/(1-\pi_0)^2\}^{3/2}} = \frac{1+\pi_0}{(r\pi_0)^{1/2}}$	$\frac{r\pi_0\left\{\frac{1}{(1-\pi_0)^2}+\frac{6\pi_0}{(1-\pi_0)^3}+\frac{6\pi_0^2}{(1-\pi_0)^4}\right\}}{\{r\pi_0/(1-\pi_0)^2\}^2} = \frac{(1-\pi_0)^2+6\pi_0(1-\pi_0)+6\pi_0^2}{r\pi_0} = \frac{(1-\pi_0)^2+6\pi_0}{r\pi_0}$	$-\frac{1+\pi_0^2}{r\pi_0}$
Gamma (fixed α)	$\frac{\alpha}{\lambda_0^2}$	$\frac{2\alpha}{\lambda_0^3}\Big/\left(\frac{\alpha}{\lambda_0^2}\right)^{3/2} = 2/\alpha^{1/2}$	$\frac{6\alpha}{\lambda_0^4}\Big/\left(\frac{\alpha}{\lambda_0^2}\right)^2 = 6/\alpha$	$-2/\alpha$

Table 4. 1,000 times the proportions of model selection and associated statistics in logistic regression by the AIC and CAIC

		$p_0 = 2,\ \boldsymbol{\beta}_0 = (-1, 1)'$				$p_0 = 3,\ \boldsymbol{\beta}_0 = (-1, 1, 1)'$				
	p:	1	2	3	Deleted	1	2	3	4	Deleted
$n = 40$	AIC	421	447	132	2	242	237	<u>389</u>	132	3
	CAIC	443	456	101		265	267	387	81	
Cor	M	0	0.250	0.299		0	0.246	0.373	0.404	
	(SD)	(0	0.143	0.133)		(0	0.136	0.128	0.122)	
$n = 80$	AIC	239	597	164	3	69	147	629	155	0
	CAIC	255	606	139		78	166	634	122	
Cor	M	0	0.233	0.261		0	0.234	0.353	0.368	
	(SD)	(0	0.105	0.101)		(0	0.107	0.098	0.094)	
$n = 160$	AIC	44	795	161	0	3	37	810	150	0
	CAIC	44	797	159		3	37	816	144	
Cor	M	0	0.237	0.251		0	0.233	0.339	0.348	
	(SD)	(0	0.075	0.073)		(0	0.073	0.069	0.068)	

Note. p_0 = the true number of regressors including an intercept, p = the number of regressors including an intercept in a model, Deleted = the number of deleted cases in the simulation, Cor = the correlation between $\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ and y_i over i = 1,…,n, M and SD = the mean and standard deviation of Cor’s over 1,000 replications. An underscore indicates that the proportion of correct model selection by the AIC is larger than that by the CAIC.

Table 5. 1,000 times the proportions of model selection and associated statistics in Poisson regression by the AIC and CAIC

		$p_0 = 2,\ \boldsymbol{\beta}_0 = (0.7,\ 0.7)'$				$p_0 = 3,\ \boldsymbol{\beta}_0 = (0.7,\ 0.7,\ 0.7)'$				
	p:	1	2	3	Deleted	1	2	3	4	Deleted
$n = 40$	AIC	6	846	148	0	0	0	832	168	0
	CAIC	6	847	147		0	0	834	166	
Cor	M	0	0.516	0.534		0	0.507	0.720	0.729	
	(SD)	(0	0.107	0.104)		(0	0.086	0.062	0.060)	
$n = 80$	AIC	0	849	151	0	0	0	858	142	0
	CAIC	0	850	150		0	0	861	139	
Cor	M	0	0.503	0.513		0	0.496	0.710	0.714	
	(SD)	(0	0.080	0.078)		(0	0.063	0.045	0.045)	
$n = 160$	AIC	0	830	170	0	0	0	850	150	0
	CAIC	0	830	170		0	0	850	150	
Cor	M	0	0.506	0.511		0	0.498	0.706	0.709	
	(SD)	(0	0.054	0.054)		(0	0.043	0.033	0.032)	

Note. p_0 = the true number of regressors including an intercept, p = the number of regressors including an intercept in a model, Deleted = the number of deleted cases in the simulation, Cor = the correlation between $\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ and y_i over i = 1,…,n, M and SD = the mean and standard deviation of Cor's over 1,000 replications.

Table 6. 1,000 times the proportions of model selection in negative binomial regression by the AIC and CAIC when the shape parameter is given

		$p_0 = 2,\ \boldsymbol{\beta}_0 = (-0.02,\ -0.02)'$				$p_0 = 3,\ \boldsymbol{\beta}_0 = (-0.02,\ -0.02,\ -0.02)'$				
	p:	1	2	3	Deleted	1	2	3	4	Deleted
$r = 1$										
$n = 40$	AIC	199	654	147	9	150	200	<u>498</u>	152	46
	CAIC	208	657	135		172	210	494	124	
$n = 80$	AIC	35	798	167	0	27	94	<u>729</u>	150	5
	CAIC	35	812	153		32	98	726	144	
$n = 160$	AIC	5	834	161	0	1	16	827	156	0
	CAIC	5	837	158		1	16	829	154	
$r = 2$										
$n = 40$	AIC	38	805	157	0	26	112	718	144	6
	CAIC	41	807	152		27	113	723	137	
$n = 80$	AIC	2	843	155	0	0	13	818	169	0
	CAIC	2	849	149		0	13	820	167	
$n = 160$	AIC	0	821	170	0	0	0	838	162	0
	CAIC	0	821	179		0	0	839	161	
$r = 4$										
$n = 40$	AIC	2	832	166	0	0	16	855	129	0
	CAIC	2	836	162		0	16	858	126	
$n = 80$	AIC	0	852	148	0	0	0	835	165	0
	CAIC	0	855	145		0	0	837	163	
$n = 160$	AIC	0	843	157	0	0	0	849	151	0
	CAIC	0	843	157		0	0	849	151	

Note. p_0 = the true number of regressors including an intercept, p = the number of regressors including an intercept in a model, Deleted = the number of deleted cases in the simulation, r = the given shape parameter. An underscore indicates that the proportion of correct model selection by the AIC is larger than that by the CAIC.

Table 7. 1,000 times the proportions of model selection in negative binomial regression by the AIC and CAIC when the shape parameter is unknown

		$p_0 = 2,\ \boldsymbol{\beta}_0 = (-0.02,\ -0.02)'$				$p_0 = 3,\ \boldsymbol{\beta}_0 = (-0.02,\ -0.02,\ -0.02)'$				
	p:	1	2	3	Deleted	1	2	3	4	Deleted
$r = 1$										
$n = 40$	AIC	191	639	170	41	144	176	<u>508</u>	172	186
	CAIC	237	642	121		218	203	457	122	
$n = 80$	AIC	35	793	172	2	27	92	718	163	39
	CAIC	43	811	146		34	111	726	129	
$n = 160$	AIC	5	830	165	0	1	13	826	160	6
	CAIC	5	843	152		1	15	837	147	
$r = 2$										
$n = 40$	AIC	38	797	165	92	36	86	697	181	798
	CAIC	58	815	127		52	117	700	131	
$n = 80$	AIC	2	829	169	18	0	19	807	174	756
	CAIC	2	856	142		1	19	832	148	
$n = 160$	AIC	0	813	187	0	0	0	831	169	694
	CAIC	0	828	172		0	0	846	154	

Note. p_0 = the true number of regressors including an intercept, p = the number of regressors including an intercept in a model, Deleted = the number of deleted cases in the simulation, r = the unknown shape parameter. An underscore indicates that the proportion of correct model selection by the AIC is larger than that by the CAIC.

Table 8. 1,000 times the proportions of model selection in gamma regression by the AIC and CAIC when the shape parameter is given

		$p_0 = 2,\ \boldsymbol{\beta}_0 = (1,\ 1)'$				$p_0 = 3,\ \boldsymbol{\beta}_0 = (2,\ 1,\ 1)'$				
	p:	1	2	3	Deleted	1	2	3	4	Deleted
$\alpha = 1$										
$n = 40$	AIC	189	<u>656</u>	155	6	433	210	<u>261</u>	96	50
	CAIC	210	654	136		460	209	247	84	
$n = 80$	AIC	41	797	162	0	235	213	<u>437</u>	115	2
	CAIC	41	805	154		246	213	429	112	
$n = 160$	AIC	2	845	153	0	74	139	<u>647</u>	140	0
	CAIC	2	849	149		75	142	643	140	
$\alpha = 2$										
$n = 40$	AIC	41	822	137	0	243	218	<u>429</u>	110	2
	CAIC	42	827	131		253	220	421	106	
$n = 80$	AIC	2	843	155	0	80	137	<u>652</u>	131	0
	CAIC	2	849	149		85	140	645	130	
$n = 160$	AIC	0	848	152	0	6	37	819	138	0
	CAIC	0	850	150		6	37	820	137	
$\alpha = 4$										
$n = 40$	AIC	1	845	154	0	70	154	629	147	0
	CAIC	1	846	153		71	154	629	146	
$n = 80$	AIC	0	821	179	0	2	48	816	134	0
	CAIC	0	823	177		2	48	816	134	
$n = 160$	AIC	0	835	165	0	0	1	848	151	0
	CAIC	0	836	164		0	1	848	151	

Note. p_0 = the true number of regressors including an intercept, p = the number of regressors including an intercept in a model, Deleted = the number of deleted cases in the simulation, α = the given shape parameter. An underscore indicates that the proportion of correct model selection by the AIC is larger than that by the CAIC.

Table 9. 1,000 times the proportions of model selection in gamma regression by the AIC and CAIC when the shape parameter is unknown

		$p_0 = 2$, $\boldsymbol{\beta}_0 = (1,\ 1)'$				$p_0 = 3$, $\boldsymbol{\beta}_0 = (2,\ 1,\ 1)'$				
	p:	1	2	3	Deleted	1	2	3	4	Deleted
$\alpha = 1$										
$n = 40$	AIC	185	647	168	27	394	215	270	121	72
	CAIC	228	645	127		475	197	239	89	
$n = 80$	AIC	37	790	173	2	230	203	444	123	8
	CAIC	47	810	143		258	215	423	104	
$n = 160$	AIC	2	844	154	0	73	135	647	145	0
	CAIC	2	857	141		79	139	646	136	
$\alpha = 2$										
$n = 40$	AIC	39	812	149	31	236	203	433	128	30
	CAIC	47	827	126		294	218	403	85	
$n = 80$	AIC	1	832	167	5	81	135	647	137	2
	CAIC	3	850	147		93	154	633	120	
$n = 160$	AIC	0	836	164	0	6	38	816	140	0
	CAIC	0	850	150		6	41	823	130	
$\alpha = 4$										
$n = 40$	AIC	0	833	167	615	70	141	630	159	408
	CAIC	2	872	126		79	165	616	120	
$n = 80$	AIC	0	807	193	457	2	42	813	143	228
	CAIC	0	830	170		4	50	828	118	
$n = 160$	AIC	0	841	159	212	0	1	835	164	82
	CAIC	0	850	150		0	1	845	154	

Note. p_0 = the true number of regressors including an intercept, p = the number of regressors including an intercept in a model, Deleted = the number of deleted cases in the simulation, α = the unknown shape parameter. An underscore indicates that the proportion of correct model selection by the AIC is larger than that by the CAIC.

Table 10. 10,000 times the simulated proportions of endpoints of one-sided confidence intervals below the population value $-2\mathrm{E}_f(\hat{\bar{l}}^*_{\mathrm{ML}})$ for the exponential family

Nominal coverage		50	250	500	5000	9500	9750	9950
λ_0=1		Based on studentization						
n = 25	Wald	208	617	957	5762	9633	9806	9963
-2E$_f$()	CF2	151	532	889	5886	9583	9762	9939
= 2.043	CF3	77	350	707	5886	9709	9860	9983
n = 50	Wald	145	495	801	5511	9612	9814	9961
-2E$_f$()	CF2	117	438	754	5606	9578	9781	9943
= 2.020	CF3	77	344	670	5606	9652	9839	9978
n = 200	Wald	81	323	613	5166	9559	9776	9971
-2E$_f$()	CF2	70	300	585	5222	9543	9755	9959
= 2.005	CF3	62	283	559	5222	9559	9774	9971
λ_0=4		Based on studentization						
n = 25	Wald	215	596	958	5734	9652	9823	9956
-2E$_f$()	CF2	164	504	868	5856	9602	9783	9934
= -0.730	CF3	67	339	687	5856	9718	9869	9980
n = 50	Wald	146	437	731	5423	9561	9776	9962
-2E$_f$()	CF2	114	398	678	5513	9529	9731	9935
= -0.752	CF3	63	316	590	5513	9608	9824	9968
n = 200	Wald	82	353	623	5250	9568	9786	9951
-2E$_f$()	CF2	69	332	609	5291	9554	9765	9946
= -0.768	CF3	58	306	587	5291	9570	9785	9951
λ_0=1		Based on standardization using the ASE of $n^{-1}\mathrm{AIC} = 2/n^{1/2}$						
n = 25	Wald	124	471	838	5762	9720	9879	9986
-2E$_f$()	CF2	82	385	761	5893	9683	9844	9969
= 2.043	CF3	77	377	741	5893	9693	9860	9976
n = 50	Wald	117	402	747	5511	9674	9848	9982
-2E$_f$()	CF2	86	353	702	5604	9638	9822	9965
= 2.020	CF3	80	348	696	5604	9643	9827	9967
n = 200	Wald	67	307	590	5166	9565	9794	9969
-2E$_f$()	CF2	61	291	573	5222	9545	9778	9959
= 2.005	CF3	61	287	572	5222	9548	9779	9960

Note. $-2\mathrm{E}_f() = -2\mathrm{E}_f(\hat{\bar{l}}^*_{\mathrm{ML}})$, CF2 (CF3) = Cornish-Fisher confidence interval with second (third)-order accuracy, ASE = asymptotic standard error.

This **Discussion Paper Series** is published by the **Center for Business Creation**(changed from the Institute of Economic Research on April 1999) and integrates two old ones published separately by the Department of Economics and the Department of Commerce.

Discussion Paper Series
Institute of Economic Research
Otaru University of Commerce

No.	Title	Author/s	Date
1.	ﾎｰｷﾝｽﾞ=ｻｲﾓﾝの条件に関する諸説の統合について	ﾀﾞｽｸﾞﾌﾟﾀ,ﾃﾞｨﾊﾟﾝｶｰ	Jul.1992
2.	Motivation and Causal Inferences in the Budgetary Control	Yoshihiro Naka	Aug.1992
3.	Проблемы управления рабочей силой на предприятиях Далънего Востока(социологические аспекты)	Анатолий Михайловичн Шкурки	Nov.1992
4.	Dynamic Tax Incidence in a Finite Horizon Model	Jun-ichi Itaya	Jan.1993
5.	Business Cycles with Asset Price Bubbles and the Role of Monetary Policy	Hiroshi Shibuya	Jun.1993
6.	Continuous Double-Sided Auctions in Foreign Exchange Markets	Ryosuke Wada	Aug.1993
7.	The Existence of Ramsey Equilibrium with Consumption Externality	Sadao Kanaya & Tomoichi Shinotsuka	Sep.1993
8.	Money, Neutrality of Consumption Taxes, and Growth in Intertemporal Optimizing Models	Jun-ichi Itaya	Nov.1993
9.	Product Returns in the Japanese Distribution System:A Case Study of a Japanese Wholesaler's Return Reduction Efforts	Jeffery Alan Brunson	Mar.1994
10.	Dynamics, Consistent Conjectures and Heterogeneous Agents in the Private Provision of Public Goods	Jun-ichi Itaya & Dipankar Dasgupta	Jun.1994
11.	Intra-industry Investment and Imperfect Markets A Geometric approach in Simple General Equilibrium	Laixun Zhao	Oct.1994
12.	Sit-Down to Split:Flint GM Workers in 1937-1939	Satoshi Takata	Dec.1994
13.	The Complementarity between Endogenous Protection and Direct foreign Investment	Laixun Zhao	Feb.1995
14.	Consumption Taxation and Tax Prepayment approach in Dynamic General equilibrium Models with Consumer Durables	Jun-ichi Itaya	Mar.1995
15.	Regulatory System and Supervision of the Financial Institutions in Japan	Osamu Ito	May 1995
16.	Financial Restructuring and the U. S. Regulatory Framework	Jane W. D'Arista	May 1995
17.	The Legacy of the Bubble Economy in Japan:Declining cross Shareholding and Capital Formation	Hiroo Hojo	May 1995
18.	Stockownership in the U. S.:Capital Formation and Regulation	Marshall E. Blume	May 1995
19.	International Joint Ventures and Endogenous Protection a Political-Economy Approach	Laixun Zhao	Nov.1995
20.	GM社をめぐるｱﾒﾘｶ労働史研究：ﾌｧｲﾝとｴｯﾂﾞﾌｫｰｽの現場像の吟味	高田聡	Feb.1996
21.	卸売業の経営と戦略――卸売流通研究会ﾋｱﾘﾝｸﾞ調査録(1):日用雑貨卸売企業	卸売流通研究会（代表　高宮城朝則）	Apr.1996
22.	卸売業の経営と戦略――卸売流通研究会ﾋｱﾘﾝｸﾞ調査録(2):食品・酒類卸売企業	卸売流通研究会（代表　高宮城朝則）	Apr.1996
23.	A Note on the Impacts of Price Shocks on Wage in Unionized Economies	Laixun Zhao	May 1996
24.	Transfer Pricing and the Nature of the subsidiary firm	Laixun Zhao	Jun.1996
25.	The Incidence of a Tax on Pure in an Altruistic Overlapping Generations Economy	Jun-ichi Itaya	Sep.1996
26.	'Small Government' in the 21st Century	Hiroshi Shibuya	Sep.1996
27.	Characteristics and Reforms of Public Health Insurance System in Japan	Takashi Nakahama	Sep.1996
28.	The Role of Local Governments in Urban Development Policy	Yoshinori Akiyama	Sep.1996
29.	Optimal Taxation and the Private Provision of Public Goods	Jun-ichi Itaya & David de Meza & Gareth D. Myles	Oct.1996
30.	Comparison of Agricultural Policy in the U. S. and the Japan	Toshikazu Tateiwa	Oct.1996
31.	US Health Insurance:Types, Patterns of Coverage and Constraints to Reform	Dwayne A. Banks	Oct.1996
32.	International Capital Flows and National Macroeconomic Policies	Jane W. D'Arista	Oct.1996
33.	Financial Liberalization and Securitization in Housing Finance and the Changing Roles of the Government	Syn-ya Imura	Oct.1996
34.	Social Efficiency and the 'Market Revolution' in US Housing Finance	Gary Dymski & Dorene Isenberg	Oct.1996
35.	Government Expenditure and the Balance of Payments:Budget Deficit, Financial Integration, and Economic Diplomacy	Hiroshi Shibuya	Nov.1996
36.	A History of PBGC and Its Roles	C. David Gustafson	Nov.1996
37.	Dynamic Provision of Public Goods as Environmental Externalities	Toshihiro Ihori & Jun-ichi Itaya	Mar.1997

38.	A Comparative Static Analysis of the Balanced Budget Incidence in the Presence of Sector-Specific Unemployment	Koh Sumino	Mar.1997
39.	An Econometric Study of Trade Creation and Trade Diversion in the EEC,LAFTA and CMEA:A Simple Application of the Gravity Model	Masahiro Endoh	Apr.1997
40.	A Dynamic Model of Fiscal Reconstruction	Toshihiro Ihori & Jun-ichi Itaya	Apr.1997
41.	The Japanese Way of Solving Financial Institution Failures	Osamu Ito	Jul.1997
42.	The Federal Role in Community Development in the U.S. :Evolution vs. Devolution	Jane Knodell	Oct.1997
43.	Rent-Seeking Behavior in the War of Attrition	Jun-ichi Itaya & Hiroyuki Sano	Oct.1997
44.	ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第1号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	May 1998
45.	購買部門の戦略性と企業間連携について	伊藤 一	Jun.1998
46.	The Formation of Customs Unions and the Effect on Government Policy Objectives	Masahiro Endoh	Jul.1998
47.	The Transition of Postwar Asia-Pacific Trade Relations	Masahiro Endoh	Jul.1998
48.	地域型ﾍﾞﾝﾁｬｰ支援ｼｽﾃﾑの研究 Ⅰー道内製造業系ﾍﾞﾝﾁｬｰ企業のｹｰｽｽﾀﾃﾞｨｰ	地域経済社会ｼｽﾃﾑ研究会 日本開発銀行札幌支店	Jul.1998
49.	Fiscal Reconstruction Policy and Free Riding Behavior of Interest Groups	Toshihiro Ihori & Jun-ichi Itaya	Aug.1998
50.	Quellen zum Markwesen des Osnabrücker Landes im Niedersächsischen Staatsarchiv Osnabrück(mit Schwerpunkt :Verfassung,Hölting,Siedlung und Konflikten im 17.und 18.Jahrhundert)	Susumu Hirai	Sep.1998
51.	Equity and Continuity with a Continuum of Generations	Tomoichi Shinotsuka	Dec.1998
52.	Public Resources Allocation and Election System	Akihiko Kawaura	Mar.1999

Discussion Paper Series
Center for Business Creation
Otaru University of Commerce

53.	消費者の価格ﾌﾟﾛﾓｰｼｮﾝ反応への影響を考慮した広告効果測定結果ﾓﾃﾞﾙの構築	奥瀬喜之	Jun.1999
54.	地域型ﾍﾞﾝﾁｬｰ支援ｼｽﾃﾑの研究Ⅱ－地域型ﾍﾞﾝﾁｬｰ･ｲﾝｷｭﾍﾞｰｼｮﾝの設計－	小樽商科大学ﾋﾞｼﾞﾈｽ創造ｾﾝﾀｰ & 日本開発銀行札幌支店	Jul.1999
55.	ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第2号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	May 1999
56.	石鹸洗剤ﾒｰｶｰにおけるﾏｰｹﾃｨﾝｸﾞ･ﾁｬﾈﾙ行動の変遷	高宮城朝則	Dec.1999
57.	長期的取引関係における資源蓄積と展開	近藤公彦&坂川裕司	Dec.1999
58.	Exernalities:A Pigovian Tax vs. A Labor Tax	Ko Sumino	Dec.1999
59.	A New Dimension of Service Quality:An Empirical Study in Japan.	Makoto Matsuo & Carolus Praet & Yoshiyuki Okuse	Dec.1999
60.	Aftermath of the Flint Sit-Down Strike:Grass-Roots Unionism and African-American Workers, 1937-1939	Satoshi Takata	Mar.2000
61.	Tariff induced dumping in the intermediate-good market	Chisato Shibayama	Apr.2000
62.	Deregulation, Monitoring and Ownership structure:A Case Study of Japanese Banks	Akihiko Kawaura	Apr.2000
63.	ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第3号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	Apr.2000
64.	A Cooperative and Competitive Organizational Culture, Innovation, and Performance: An Empirical Study of Japanese Sales Departments	Makoto Matsuo	May 2000
65.	Foreign Exchange Market Maker's Optimal Spread with Heterogeneous Expectations	Ryosuke Wada	Jun.2000
66.	ダンピングとダンピング防止法の起源 歴史的文脈における「不公正貿易」概念の成立	柴山千里	Oct.2000
67.	The Organizational Learning Process: A Review	Makoto Matsuo	Dec.2000
68.	The Weak Core of Simple Games with Ordinal Preferences: Implementation in Nash Equilibrium	Tomoichi Shinotsuka & Koji Takamiya	Jan.2001
69.	業態開発におけるイノベーションと競争ービブレのケースー	近藤公彦	Jan.2001
70.	Budget Distribution Problem	Tomoichi Shinotsuka	Feb.2001
71.	小売バイヤー組織の機能と顧客対応	伊藤 一	May 2001
72.	The Effect of Intra-Organizational Competition on Knowledge Creation:Case Study of a Japanese Financial Company	Makoto Matsuo	May 2001
73.	ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第4号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	Mar.2001
74.	The Weak Core of Simple Games with Ordinal Preferences: Implementation in Nash Equilibrium	Tomoichi Shinotsuka & Koji Takamiya	Oct.2001
75.	環境保全型河川計画と景観構築に係る計画技術の研究	地域環境問題研究会（代表 八木宏樹）	Oct.2001
76.	Additivity, Bounds, and Continuity in Budget Distribution Problem	Tomoichi Shinotsuka	Dec.2001

77. Monetary Policy in Bhutan: Implications of Indian Rupee Circulation	Akihiko Kawaura	Dec.2001
78. Optimal Multiobject Auctions with Correlated Types	Tomoichi Shinotsuka & Simon Wilkie	Feb.2002
79. ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第5号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	Mar.2002
80. The Case Study of Retail Buying Organization in Japanese Context	Hajime Itoh	Mar.2002
81. 宿泊業のサービスのサービス構成要素に関する重要度調査法に関しての一考察 北海道への台湾人観光客の事例を中心に	稲葉由之&沈潔如&伊藤 一	Feb.2003
82. ブティック経営における販売要素の分析 -AHPによる経営者・販売員間における重要度認識比較に関する一考察-	伊藤 一&橋詰敦樹	Mar.2003
83. 温泉地に対するイメージギャップに関する調査	伊藤 一	Mar.2003
84. Literature Review on Retail Buyer from Research on Industrial Purchasing	Hajime Itoh	
85. The Comparison Study on Retail Buyer Behaviour between UK, Australia and Japan	Hajime Itoh	
86. 社会科学研究の基礎－大学院生のための研究法－	ダン・レメニイ他著 抄訳 稲葉由之&奥瀬善之&近藤公彦&玉井健一&高宮城朝則&松尾 睦	Mar.2002
87. マーケティング行為からみた小売業による需要創造 －明治期呉服店の経営行為を考察対象として－	坂川裕司	May 2002
88. Interdependent Utility Functions in an Intergenerational Context	Tomoichi Shinotsuka	May 2002
89. Internal and External Views of the Corporate Reputation in the Japanese Hotel Industry	Hajime Itoh	Feb.2003
90. ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第6号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	Mar.2003
91.小売購買行動研究に関する展望 －「買い手視角」での小売購買行動研究に向けて－	坂川裕司	May 2003
92.商品購買における「情報システムの逆機能」 －リスク回避的バイヤーにみる合理性とその弊害－	坂川裕司	Sep.2003
93.An Experiment of Round-Robin Tournament by Excel's Macro -Using 160 Students' Data from Cournot Duopoly Game-	Masaru Uzawa	Apr.2004
94.Earnings Management through Deferred Tax Aseets -In Case of Banking Company-	Hiroshi Onuma	Jun.2004
97.Competition between Matching Markets	Koji Yokota	May 2005
98.On the role of asymmetric information in the aggregate matching function	Koji Yokota	Apr.2006
99.A note on Optimal Taxation in the Presence of Externalities	Tomoichi Shinotsuka & Ko Sumino	Feb.2005
100. A Note on Jones' Model of Growth	Mutsuhiro Kato	Mar.2005
101. 整数ナップサック問題が多項式時間で解ける特殊な場合を定める条件について	飯田浩志	Jul.2005
102. ＩＴ技術者の熟達化と経験学習	松尾 睦	Sep.2005
103. Product De-listing by Retail Buyers: Relational Antecedents and Consequences	Gary Davies & Hajime Itoh	Dec.2005
104. 米国地域経営史における多文化主義的発展－１９３０年代ミシガン州フリントにおけるアフリカ系コミュニティの起業基盤を中心に－	高田 聡	May 2006
105. 環境便益を反映させた環境指標の開発 Developing an environmental indicator including environmental benefits	山本 充	Apr.2006
106. A Critical Investigation of Long-run Properties of Endogenous Growth Models	Mutsuhiro Kato	May 2006
107. What is National Income in Jones' Model of Growth? :An Expository Annotation	Mutsuhiro Kato	Jun.2006
108. A Further Analysis of the Consumer Behavior in Jones' R&D-Based Model of Economic Growth	Mutsuhiro Kato	Aug.2006
109. 看護師の経験学習プロセス	松尾 睦 & 正岡経子 & 吉田真奈美 & 丸山知子 & 荒木奈緒	Feb.2007
110. Comments on knapsack problems with a penalty	Iida Hiroshi	Mar.2007
111. 看護師の経験学習に関する記述的分析	松尾 睦 & 正岡経子 & 吉田真奈美 & 丸山知子 & 荒木奈緒	Jul.2007
112. 頂点被覆へのリスト減少法の解析に関する一考察	飯田浩志	Dec.2007
113. 小中学校における校長の経営観－探索的分析－	松尾 睦	Dec.2007
114. インタビュー調査：戦後復興期大阪における自転車部品製造業者・問屋の経営活動	田中幹大	Apr.2008
115. Partitionのある風景	飯田浩志	Jun.2008
116. Multiproduct Firms and Dumping	Chisato Shibayama & Yasunori Ishii	Jul.2008
117. モスクワの低層住宅団地開発―2つのケース－	小田福男	Mar.2009
118. 整数ナップサックの周期性について	飯田浩志	Mar.2009

119.	Discussion paper series no.118への補遺	飯田浩志	Jul.2009
120.	環境フィードバック効果を考慮したSandmoモデルによる二重配当仮説の再考察	角野　浩	Jul.2009
121.	部分線形モデルの差分推定量の漸近理論	劉　慶豊	Oct.2009
122.	モデル平均理論の新展開	劉　慶豊	Oct.2009
123.	Production Theory with Convex Labor Friction: Foundation of an Optimal Non-market-clearing Economy	Koji Yokota	Dec.2009
124.	19世紀ドイツの農村ゲマインデ制と政治参加資格 ---北西ドイツ・ハノーファーを中心に-----	平井　進	Feb.2010
125.	環境経営と企業業績に関する実証研究（再検討:2003-2008)	加賀田和弘	Apr.2010
126.	「北海道ブランド」の仕入れに関する研究 ―台湾小売バイヤーの視点から―	沈　潔如	Apr.2010
127.	Generalized Cp Model Averaging for Heteroskedastic models	Qingfeng Liu	Oct.2010
128.	How to solve the collapsing subset-sum problem revisited	Hiroshi Iida	Jan.2011
129.	顧客関係のマネジメントの系譜	近藤公彦	Feb.2011
130.	An Application of Forecast Combination Methods to Default Risk Prediction	Qingfeng Liu	Feb.2011
131.	An effect of consumer's earlier decision to purchase a discount ticket	Ryosuke Ishii & Kuninori Nakagawa	Feb.2011
132.	On the Behavior of money flows on the real side and the financial side in Hokkaido prefecture	Toshiaki Kanzaki	Mar.2011
133.	星野リゾート― 顧客志向の組織マネジメント ―	乙政佐吉 & 近藤隆史	Mar.2011
134.	(ケース）札幌ビズカフェ―地域企業家ネットワークにおける中間主導型組織の役割―	加藤敬太	Mar.2011
135.	二重配当効果の最適課税ルールにしたがった再評価	角野　浩	Mar.2011
136.	18・19世紀前半北海沿岸農村社会の地域役職者：Landschaft Eiderstedt	平井　進	Mar.2011
137.	Tax Collecting Efforts and Local Allocation Tax Grants in Japan: The Effect of Administrative Reform Incentive Assessment on Local Tax Collection Rates	Mitsunari Ishida	Mar.2011
138.	The bargaining family with strategic interaction	Atsue Mizushima & Koichi Futagami	Mar.2011
139.	Generalized Cp Model Averaging for Heteroskedastic Models (Revised Version)	Qingfeng Liu	Apr.2011
140.	Exclusion of agents, virtual surplus and a transversality condition in adverse selection	Naoki Kojima	May 2011
141.	Implementability by a canonical indirect mechanism of an optimal two-dimensional direct mechanism	Naoki Kojima	Jun.2011
142.	18・19世紀前半北西ドイツ北海沿岸地方の領邦官吏と自治組織役職者：Landschaft S?derdithmarschen	平井　進	Jun.2011
143.	CRMにおける顧客関係のマネジメント	近藤公彦	Aug.2011
144.	企業家ネットワーキングによる地域企業のビジネスシステム・イノベーション―サムライ日本プロジェクトの事例分析―	加藤敬太	Oct.2011
145.	Observable Actions	Ryosuke Ishii	Nov.2012
146.	Dumping in Transition Economies and the Effects of Anti-Dumping Policy	Chisato Shibayama & Yasunori Ishii	Mar.2012
147.	Time Discount and Convex Hiring Cost	Koji Yokota	May 2012
148.	Two-dimensional Mechanism Design and Implementability by an Indirect Mechanism	Naoki Kojima	Jun.2012
149.	北海道経済と開発のプロセス	神﨑稔章	Dec.2012
150.	道内における地域経済の現状について	渡久地朝央 & Baljinnyam Maitsetseg	Dec.2012
151.	モンゴルにおける資本主義転換後の地域間経済格差に関するパネルデータ分析	渡久地朝央 & Baljinnyam Maitsetseg	Dec.2012
152.	モンゴルにおける食肉価格の動向に関するパネルデータ分析	Baljinnyam Maitsetseg & 渡久地朝央	Dec.2012
153.	付加価値率の動向と地方自治体による政策効果の関係について －北海道における製造産業を対象としたパネルデータ分析－	渡久地朝央	Dec.2012
154.	CRMにおける組織能力	近藤公彦	Feb.2013
155.	19世紀北西ドイツの農村ゲマインデ制の変革---自治参加資格と家屋・土地保有要件---	平井　進	Feb.2013
156.	北海道における産業クラスターに関する文献資料目録	加藤敬太	Mar.2013
157.	トヨタ自動車北海道のマネジメント	乙政佐吉	Mar.2013
158.	Mechanism design to the budget constrained buyer: a canonical mechanism approach	Naoki Kojima	May 2013
159.	First Price Package Auction with Many Traders	Yasuhiro Shirata	Jun.2013
160.	整数ナップサックの周期性についてあれこれ	飯田浩志	Jul.2013

161.	Non-cooperative versus Cooperative Family	Atsue Mizushima & Koichi Futagami	Oct.2013
162.	Perverse effects of a ban on child labour in an overlapping generations model	Kouki Sugawara & Atsue Mizushima & Koichi Futagami	Oct.2013
163.	Human Infrastructure, Child Labor, and Growth	Atsue Mizushima	Oct.2013
164.	18・19世紀前半北海沿岸農村社会の指導的地域役職者・領邦地方官吏と土地所有：Landschaft Norderdithmarschen	平井　進	Mar.2014
165.	ビジネスシステムの形成から見る6次産業化—パイオニアジャパングループの事例分析—	笹本香菜 & 加藤敬太	Mar.2014
166.	ナップサック問題への2近似算法について雑感	飯田浩志	Jul.2014
167.	A further addendum to "Some thoughts on the 2-approximation algorithm for knapsack problems: A survey"	Hiroshi Iida	Nov.2014
168.	ワイン産業における企業家コミュニティー北海道と山梨県の比較事例研究ー	長村知幸	Jan.2015
169.	サハリン住宅建設企業の動向—2013年9月調査に基づいて—	小田福男	Mar.2015
170.	Asymptotic cumulants of some information criteria	Haruhiko Ogasawara	Apr.2015
171.	*k*KPをめぐる雑感	飯田浩志	Jun.2015
172.	Should income inequality be praised? Multiple public goods Provision, income distribution and social welfare	Jun-ichi Itaya & Atsue Mizushima	Jun.2015
173.	三種の*k*KP→□→0-1KPまとめ	飯田浩志	Nov.2015
174.	Asymptotic cumulants of some information criteria (2nd version)	Haruhiko Ogasawara	Nov.2015

Discussion Paper Series Department of Economics, Otaru University of Commerce No.1-16 Feb.1985-Oct.1991

Discussion Paper Series Department of Commerce, Otaru University of Commerce No.1-2 Apr.1985-May 1989

Center for Business Creation, National University Corporation Otaru University of Commerce
3-5-21, Midori, Otaru, Hokkaido 047-8501, Japan Tel +81-134-27-5290 Fax +81-134-27-5293
E-mail:cbcjimu@office.otaru-uc.ac.jp

国立大学法人小樽商科大学ビジネス創造センター
〒047-8501 北海道小樽市緑３丁目５番２１号 Tel 0134-27-5290 Fax 0134-27-5293
E-mail:cbcjimu@office.otaru-uc.ac.jp