HIROSHIMA MATH. J.
50 (2020), 223–251

Asymptotic bias of C_p type criterion for model selection in the GEE when the sample size and the cluster sizes are large

Tomoharu SATO

(Received March 22, 2019)
(Revised December 23, 2019)

ABSTRACT. In this paper, we evaluate the asymptotic bias of C_p type criterion for model selection in the GEE (generalized estimating equation) method when the sample and cluster sizes are large. We present the asymptotic properties of GEE estimator and the model selection criterion. Then, we present the order of the asymptotic bias of PMSEG (the prediction mean squared error in the GEE).

1. Introduction

Longitudinal data in which the observations are correlated are widely used in many fields. Generalized estimating equation (GEE) proposed by Liang and Zeger [6] is a representative method for analyzing such data. In the GEE method, we use a working correlation matrix to estimate the regression coefficients without specifying the joint distribution of observations. We can choose a working correlation matrix freely, which is one of the reasons that the GEE method is widely used. The asymptotic properties of the GEE estimator were derived by Xie and Yang [11]. They ensured the existence, consistency and asymptotic normality of the GEE estimator under some conditions.

Model selection is an important issue in the GEE framework. A widely used model selection criterion is the Akaike's Information Criterion (AIC) (Akaike [1], [2]). The AIC is based on the likelihood function of responses and the asymptotic properties of the maximum likelihood estimator. Furthermore, Generalized Information Criterion (GIC) proposed by Nishii [8] and Rao [10] which is a generalization of AIC is also widely used.

Since we get the GEE estimator without specifying a joint distribution, there is no likelihood. Thus, the likelihood based information criterion cannot be used in the GEE. Pan [9] proposed QIC (quasi-likelihood under the independence model criterion for GEE) by using an independent quasi-likelihood. The Mallows's C_p (Mallows [7]) based on the prediction mean squared error is

2010 *Mathematics Subject Classification*. Primary 62H12; Secondary 62F07.

Key words and phrases. generalized estimating equation, longitudinal data, prediction mean squared error, model selection, prediction error.

also widely used. Since the Mallows's C_p is based on the prediction mean squared error without using a likelihood, we can use this kind of criterion for GEE. Inatsu and Imori [4] proposed the new model selection criterion PMSEG (the prediction mean squared error in the GEE) by using the risk function based on the prediction mean squared error normalized by the covariance matrix. PMSEG can reflect the correlation between responses. Inatsu and Sato [5] evaluated the influence of estimation of the correlation parameters included in the working correlation matrix and the scale parameter included in the marginal distribution of responses. They mentioned that we can get an asymptotic unbiased estimator of the risk function when the sample size is large by using the moment estimators of the correlation parameters and the scale parameter. They also mentioned that by using PMSEG, we can select an optimal subset of variables and a working correlation matrix simultaneously. By selecting both the subset and the working correlation matrix, we may be able to improve the prediction accuracy. Inatsu and Imori [4], and Inatsu and Sato [5] proposed the C_p type criterion when the sample size is large and the maximum cluster size is bounded.

In this paper, we evaluate the asymptotic bias of PMSEG when the maximum cluster size goes to infinity as the sample size goes to infinity. The present paper is organized as follows: In section 2, we introduce the GEE framework and PMSEG. After that, we introduce the asymptotic property of the estimator of a regression coefficient vector. In section 3, we evaluate the asymptotic bias. In section 4, we perform a numerical study. In Appendix, we prove the theorems in section 3.

2. Model selection in the GEE

Let $(y_{ij}, \boldsymbol{x}_{f,ij})$ be observations for the jth measurement on the ith subject, where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m_i$ and m_i is the cluster size of the ith subject. Here, y_{ij} is a scalar response and $\boldsymbol{x}_{f,ij}$ is an l-dimensional explanatory vector. Assume that the observations from different subjects are independent and the observations from the same subject are correlated. For each $i = 1, \ldots, n$, let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im_i})'$ be a response variable vector from the ith subject and $\boldsymbol{X}_{f,i} = (\boldsymbol{x}_{f,i1}, \ldots, \boldsymbol{x}_{f,im_i})'$ be a full explanatory matrix from the ith subject. Moreover let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im_i})'$ be a $m_i \times p$ submatrix of the matrix $\boldsymbol{X}_{f,i}$, where $l \geq p$. Liang and Zeger [6] used the generalized linear model (GLM) to model the marginal density of y_{ij}:

$$f(y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{\beta}, \phi) = \exp[\{y_{ij}\theta_{ij} - a(\theta_{ij})\}/\phi + b(y_{ij}, \phi)], \tag{2.1}$$

where $a(\cdot)$ and $b(\cdot)$ are known functions, θ_{ij} is an unknown location parameter defined by $\theta_{ij} = u(\eta_{ij}) = \theta_{ij}(\boldsymbol{\beta})$ with known injective function $u(\cdot)$, and ϕ is a

nuisance scale parameter. Here, $\boldsymbol{\beta}$ is a p-dimensional unknown regression coefficient vector and $\eta_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta}$ is called the linear predictor. We call that the model with $\boldsymbol{x}_{f,ij}$ and $\boldsymbol{x}_{ij}$ as the full model and the candidate model, respectively. In the present paper, let Θ be the *natural parameter space* (see, Xie and Yang [11]) of the exponential family distributions presented in (2.1), and the interior of Θ is denoted as Θ°. Then, it is known that Θ is convex and all the derivatives of $a(\cdot)$ and all the moments of y_{ij} exist in Θ°. We denote the derivative and the second derivative of a function $f(x)$ as $\dot{f}(x)$ and $\ddot{f}(x)$, respectively. Under this model specification, the first two moments of y_{ij} are given by

$$\mu_{ij}(\boldsymbol{\beta}) = \mathrm{E}[y_{ij}] = \dot{a}(\theta_{ij}), \qquad \sigma_{ij}^2(\boldsymbol{\beta}) = \mathrm{Var}[y_{ij}] = \ddot{a}(\theta_{ij})\phi \equiv v(\mu_{ij}(\boldsymbol{\beta})).$$

Denote $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), \ldots, \mu_{im_i}(\boldsymbol{\beta}))'$, $\boldsymbol{D}_i(\boldsymbol{\beta}) = \partial\boldsymbol{\mu}_i(\boldsymbol{\beta})/\partial\boldsymbol{\beta}' = \boldsymbol{A}_i(\boldsymbol{\beta})\boldsymbol{\Delta}_i(\boldsymbol{\beta})\boldsymbol{X}_i$, $\boldsymbol{A}_i(\boldsymbol{\beta}) = \mathrm{diag}(\sigma_{i1}^2(\boldsymbol{\beta}), \ldots, \sigma_{im_i}^2(\boldsymbol{\beta}))$, $\boldsymbol{\Delta}_i(\boldsymbol{\beta}) = \mathrm{diag}(\partial\theta_{i1}/\partial\eta_{i1}, \ldots, \partial\theta_{im_i}/\partial\eta_{im_i})$ and $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}_w(\boldsymbol{\alpha})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\phi$, where $\boldsymbol{\alpha}$ is a nuisance correlation parameter. Here, $\boldsymbol{R}_w(\boldsymbol{\alpha})$ is called a "working correlation matrix" that one can choose freely. Typical working correlation matrices are follows:

$$\begin{aligned}
&\text{Independence} : (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = 0 \ \ (j \neq k),\\
&\text{Exchangeable} : (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = \alpha \ \ (j \neq k),\\
&\text{Autoregressive} : (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{kj} = \alpha^{j-k} \ \ (j > k),\\
&\text{1-dependence} : (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{kj} = \begin{cases} \alpha & (j = k+1) \\ 0 & (j \neq k+1,\ j \neq k) \end{cases},\\
&\text{Unstructured} : (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{kj} = \alpha_{jk} \ \ (j > k).
\end{aligned}$$

Note that the diagonal elements of $\boldsymbol{R}_w(\boldsymbol{\alpha})$ are ones, since it is a correlation matrix. Assume that the cluster sizes m_i on ith subject is the common m. Denote $\boldsymbol{\Sigma}_i(\boldsymbol{\beta}) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}_0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\phi$, where $\boldsymbol{R}_0$ is the true correlation matrix of $\boldsymbol{y}_i$, here we assume that for $i = 1, \ldots, n$, the true correlation matrix is the common. Moreover, $\boldsymbol{R}_w(\boldsymbol{\alpha})$ includes nuisance parameter $\boldsymbol{\alpha}$. If $\boldsymbol{R}_w(\boldsymbol{\alpha})$ is equal to the true correlation matrix $\boldsymbol{R}_0$, then $\boldsymbol{V}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}) = \boldsymbol{\Sigma}_i(\boldsymbol{\beta}_0) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}_0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\phi = \mathrm{Cov}[\boldsymbol{y}_i]$ at the true regression coefficient $\boldsymbol{\beta}_0$.

In the GEE method, when the correlation parameters and the scale parameter are known, we solve the following equation:

$$\boldsymbol{q}_{nm}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{D}_i'(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = \boldsymbol{0}_p, \tag{2.2}$$

where $\mathbf{0}_p$ is the p-dimensional vector of zeros. Here, the parameter space of the correlation parameters is defined as follows:

$$\mathscr{A} = \{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s)' \in \mathbb{R}^s \mid \boldsymbol{R}_w(\boldsymbol{\alpha}) \text{ is positive definite}\}.$$

When the correlation parameters and the scale parameter are unknown, we estimate $\boldsymbol{\alpha}$ by $\boldsymbol{\beta}$ and an estimator of ϕ. Furthermore, we estimate ϕ by $\boldsymbol{\beta}$. Let $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})) = (\hat{\alpha}_1(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})), \ldots, \hat{\alpha}_s(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})))$ be an estimator of $\boldsymbol{\alpha}$, where $\hat{\phi}(\boldsymbol{\beta})$ is an estimator of ϕ. We replace (2.2) with the following equation:

$$\boldsymbol{s}_{nm}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{D}_i'(\boldsymbol{\beta}) \boldsymbol{\Gamma}_i^{-1}(\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = \mathbf{0}_p, \tag{2.3}$$

where $\boldsymbol{\Gamma}_i(\boldsymbol{\beta}) = \boldsymbol{V}_i(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})))$. The solution of (2.3) denoted as $\hat{\boldsymbol{\beta}}$ is the estimator of true regression coefficient $\boldsymbol{\beta}_0$. We call $\hat{\boldsymbol{\beta}}$ the GEE estimator.

When the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and ϕ are unknown, we estimate them by the following iterative method (see, Inatsu and Sato [5]):

Algorithm (Estimation method for parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and ϕ)

Step 1 Set an initial value of $\boldsymbol{\alpha}$ denoted as $\hat{\boldsymbol{\alpha}}^{\langle 0 \rangle}$.
Step 2 Solve the GEE with $\hat{\boldsymbol{\alpha}}^{\langle k \rangle}$, and the solution of the GEE is denoted as $\hat{\boldsymbol{\beta}}^{\langle k \rangle} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}^{\langle k \rangle})$.
Step 3 Estimate $\hat{\phi}^{\langle k+1 \rangle}$ by $\hat{\boldsymbol{\beta}}^{\langle k \rangle}$.
Step 4 Estimate $\hat{\boldsymbol{\alpha}}^{\langle k+1 \rangle}$ by $\hat{\boldsymbol{\beta}}^{\langle k \rangle}$ and $\hat{\phi}^{\langle k+1 \rangle}$.
Step 5 Iterate from step 2 to 4 until a certain condition about the convergence holds.

We use the moment estimators of ϕ and $\boldsymbol{\alpha}$, for example:

$$\text{Scale parameter} : \hat{\phi}(\hat{\boldsymbol{\beta}}) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(y_{ij} - \mu_{ij}(\hat{\boldsymbol{\beta}}))^2}{\ddot{a}(\theta_{ij}(\hat{\boldsymbol{\beta}}))},$$

$$\text{Exchangeable} : \hat{\alpha}(\hat{\boldsymbol{\beta}}, \hat{\phi}(\hat{\boldsymbol{\beta}})) = \frac{1}{nm(m-1)} \sum_{i=1}^{n} \sum_{j>k} \hat{r}_{i,j}(\hat{\boldsymbol{\beta}}) \hat{r}_{i,k}(\hat{\boldsymbol{\beta}}) / \hat{\phi}(\hat{\boldsymbol{\beta}}),$$

$$\text{Autoregressive} : \hat{\alpha}(\hat{\boldsymbol{\beta}}, \hat{\phi}(\hat{\boldsymbol{\beta}})) = \frac{1}{n(m-1)} \sum_{i=1}^{n} \sum_{j=1}^{m-1} \hat{r}_{i,j}(\hat{\boldsymbol{\beta}}) \hat{r}_{i,j+1}(\hat{\boldsymbol{\beta}}) / \hat{\phi}(\hat{\boldsymbol{\beta}}),$$

$$\text{1-dependence} : \hat{\alpha}(\hat{\boldsymbol{\beta}}, \hat{\phi}(\hat{\boldsymbol{\beta}})) = \frac{1}{(n-p)(m-1)} \sum_{i=1}^{n} \sum_{j=1}^{m-1} \hat{r}_{i,j}(\hat{\boldsymbol{\beta}}) \hat{r}_{i,j+1}(\hat{\boldsymbol{\beta}}) / \hat{\phi}(\hat{\boldsymbol{\beta}}),$$

$$\text{Unstructured} : \hat{\alpha}_{jk}(\hat{\boldsymbol{\beta}}, \hat{\phi}(\hat{\boldsymbol{\beta}})) = \frac{1}{n} \sum_{i=1}^{n} \hat{r}_{i,j}(\hat{\boldsymbol{\beta}}) \hat{r}_{i,k}(\hat{\boldsymbol{\beta}}) / \hat{\phi}(\hat{\boldsymbol{\beta}}),$$

where $\hat{r}_{i,j}(\hat{\boldsymbol{\beta}}) = y_{ij} - \mu_{ij}(\hat{\boldsymbol{\beta}})$. Then, we assume that $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \phi_0) \xrightarrow{p} \boldsymbol{\alpha}_0 \in \mathscr{A}^\circ$ and $\hat{\boldsymbol{\phi}}(\boldsymbol{\beta}_0) \xrightarrow{p} \phi_0$, where $\boldsymbol{\alpha}_0$ is the convergence value of $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \phi_0)$ and ϕ_0 is the convergence value of $\hat{\boldsymbol{\phi}}(\boldsymbol{\beta}_0)$.

Inatsu and Imori [4], and Inatsu and Sato [5] proposed the following risk function:

$$\mathrm{Risk}_P = \mathrm{PMSE} - mn = \mathrm{E}_y\left[\mathrm{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)\right]\right] - mn,$$

where $\hat{\boldsymbol{\mu}}_i = (\hat{\mu}_{i1}, \ldots, \hat{\mu}_{im})' = \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})$, $\boldsymbol{\Sigma}_{i,0} = \boldsymbol{\Sigma}_i(\boldsymbol{\beta}_0)$ and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})'$ is an m-dimensional random vector that is independent of $\boldsymbol{y}_i$ and has the same distribution as $\boldsymbol{y}_i$. Note that if the estimator of regression coefficient equals to the true regression coefficient, Risk_P has the minimum value zero.

Denote

$$\hat{\boldsymbol{R}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))'\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})/\hat{\phi}(\boldsymbol{\beta}),$$

$$\mathscr{L}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_1))'\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_2)\hat{\boldsymbol{R}}^{-1}(\boldsymbol{\beta}_2)\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_2)(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_1))\hat{\phi}^{-1}(\boldsymbol{\beta}_2),$$

$$\mathscr{L}^*(\boldsymbol{\beta}) = \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})).$$

Since the PMSE is typically unknown, Inatsu and Sato [5] estimate the PMSE by $\mathscr{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$, where $\hat{\boldsymbol{\beta}}_f$ is the GEE estimator from the full model. Then, by correcting the asymptotic bias of the estimator of the risk, they propose the model selection criterion PMSEG as follows:

$$\mathrm{PMSEG} = \mathscr{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f) + 2p.$$

Xie and Yang [11] presented asymptotic properties of the GEE estimator of the regression coefficient $\hat{\boldsymbol{\beta}}$ when the cluster size m goes to infinity as the sample size n goes to infinity. Let $\lambda_{\min}(\boldsymbol{A})$ ($\lambda_{\max}(\boldsymbol{A})$) denotes the smallest (largest) eigenvalue of a matrix $\boldsymbol{A}$, and

$$\boldsymbol{H}_{nm}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\boldsymbol{D}_i'(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)\boldsymbol{D}_i(\boldsymbol{\beta}),$$

$$\boldsymbol{M}_{nm}(\boldsymbol{\beta}) = \mathrm{Cov}[\boldsymbol{q}_{nm}(\boldsymbol{\beta})] = \sum_{i=1}^{n}\boldsymbol{D}_i'(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)\boldsymbol{\Sigma}_i(\boldsymbol{\beta})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha}_0)\boldsymbol{D}_i(\boldsymbol{\beta}),$$

$$\boldsymbol{F}_{nm}(\boldsymbol{\beta}) = \boldsymbol{H}_{nm}(\boldsymbol{\beta})\boldsymbol{M}_{nm}^{-1}(\boldsymbol{\beta})\boldsymbol{H}_{nm}(\boldsymbol{\beta}).$$

We consider the following regularity conditions (see, e.g., Xie and Yang [11], Inatsu and Sato [5]):

C1. The set $\mathscr{X}$ is compact. For all sequence $\{\boldsymbol{x}_{ij}\}$, it is established that $u(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) \in \Theta^\circ$ and $\boldsymbol{x}_{ij} \in \mathscr{X}$.

C2. The true regression coefficient $\boldsymbol{\beta}_0$ is in $\mathscr{B}^\circ$, where $\mathscr{B}^\circ$ is the interior of an admissible set $\mathscr{B}$, i.e., $\boldsymbol{\beta}_0 \in \mathscr{B}^\circ$, $\mathscr{B} = \{\boldsymbol{\beta} \mid u^{-1}(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) \in \Theta \text{ if } \boldsymbol{x}_{ij} \in \mathscr{X}\}$.

C3. For any $\boldsymbol{\beta} \in \mathscr{B}$, it is established that $\boldsymbol{x}_{ij}'\boldsymbol{\beta}$ is included in $g(\mathscr{M})$, where $\mathscr{M}$ is the image $\dot{\boldsymbol{a}}(\Theta^\circ)$ of Θ°.

C4. The function $u(\eta_{ij})$ is four times continuously differentiable and $\dot{\boldsymbol{u}}(\eta_{ij}) > 0$ in $g(\mathscr{M}^\circ)$, where $\mathscr{M}^\circ$ is the interior of $\mathscr{M}$.

C5. The matrix $\boldsymbol{M}_{nm,0}$ is positive definite when n or m is sufficiently large, denoted by

$$\boldsymbol{M}_{nm,0} = \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \boldsymbol{V}_{i,0}^{-1} \boldsymbol{\Sigma}_{i,0} \boldsymbol{V}_{i,0}^{-1} \boldsymbol{D}_{i,0},$$

where $\boldsymbol{D}_{i,0} = \boldsymbol{D}_i(\boldsymbol{\beta}_0)$ and $\boldsymbol{V}_{i,0} = \boldsymbol{V}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)$.

C6. It is established that $\liminf_{n\to\infty, m\to\infty} \lambda_{\min}(\boldsymbol{H}_{nm,0}/nm) > 0$, where $\boldsymbol{H}_{nm,0} = \boldsymbol{H}_{nm}(\boldsymbol{\beta}_0)$.

C7. It holds that $\tau_{nm}\lambda_{\max}(\boldsymbol{H}_{nm,0}^{-1}) \to 0$, where $\tau_{nm} = \lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{R}_0)$.

C8. It holds that $\pi_{nm}^2 \tau_{nm} m \gamma_{nm}^{(0)} \to 0$, where

$$\pi_{nm} = \frac{\lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))}{\lambda_{\min}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))},$$

$$\gamma_{nm}^{(0)} = \max_{1\le i\le n} \max_{1\le j\le m} \boldsymbol{x}_{ij}' \boldsymbol{H}_{nm,0}^{-1} \boldsymbol{x}_{ij}.$$

C9. It holds that $(c_{nm})^{1+\delta}(\tilde{\lambda}_{nm} m)^{2+\delta} \gamma_{nm}^{(0)} \to 0$ for some $\delta > 0$, where

$$c_{nm} = \lambda_{\max}(\boldsymbol{M}_{nm,0}^{-1} \boldsymbol{H}_{nm,0}),$$

$$\tilde{\lambda}_{nm} = \lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)).$$

The conditions C1–C9 are the modifications of the conditions proposed by Xie and Yang [11]. Here, to evaluate the asymptotic bias of PMSEG, we present the following lemma:

LEMMA 1. *Suppose the conditions* C1–C9 *hold.*

(a) *There exists a sequence of random variable* $\hat{\boldsymbol{\beta}}$ *such that* $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$ *in probability, and* $\boldsymbol{M}_{nm,0}^{-1/2} \boldsymbol{H}_{nm,0}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *and* $\boldsymbol{M}_{nm,0}^{-1/2} \boldsymbol{q}_{nm}(\boldsymbol{\beta}_0)$ *have the same asymptotic distributions.*

(b) *When* $n \to \infty$,

$$\boldsymbol{M}_{nm,0}^{-1/2} \boldsymbol{H}_{nm,0}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(\boldsymbol{0}_p, \boldsymbol{I}_p) \text{ in distribution.}$$

Lemma 1 is the same as that of Corollary 1 of Xie and Yang [11], so we omit the proof. Here, $\boldsymbol{I}_p$ is the p-dimensional identity matrix.

3. The asymptotic bias of PMSEG

In this section, we evaluate the asymptotic bias of PMSEG. We denote the derivatives of a matrix $\boldsymbol{W}$ whose elements w_{ij}'s are functions of $\boldsymbol{\beta}$, by $\boldsymbol{\beta}$ and β_k as follows:

$$\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{W} = \left(\frac{\partial \boldsymbol{W}}{\partial \beta_1}, \ldots, \frac{\partial \boldsymbol{W}}{\partial \beta_p} \right), \qquad \frac{\partial \boldsymbol{W}}{\partial \beta_k} = \left(\frac{\partial w_{ij}}{\partial \beta_k} \right),$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$.

We consider the following assumptions (see, Inatsu and Sato [5]):

C10. There exists a compact neighborhood of $\boldsymbol{\alpha}_0$, say $U_{\boldsymbol{\alpha}_0}$, such that $\mathrm{vec}\{\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha})\}$ is three times continuously differentiable in the interior of $U_{\boldsymbol{\alpha}_0}$.

C11. There exists a compact neighborhood of $\boldsymbol{\beta}_0$, say $U_{\boldsymbol{\beta}_0}$, such that $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\boldsymbol{\phi}}(\boldsymbol{\beta}))$ is three times continuously differentiable in the interior of $U_{\boldsymbol{\beta}_0}$.

C12. For all $\boldsymbol{\beta} \in U_{\boldsymbol{\beta}_0}$, it is established that $\hat{\boldsymbol{\alpha}}^{(k)}(\boldsymbol{\beta}) = O_p(1)$ $(k = 1, 2, 3)$, where

$$\hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}) = \frac{\partial \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\boldsymbol{\phi}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}'}, \qquad \hat{\boldsymbol{\alpha}}^{(2)}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}), \qquad \hat{\boldsymbol{\alpha}}^{(3)}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \hat{\boldsymbol{\alpha}}^{(2)}(\boldsymbol{\beta}).$$

C13. The estimator $\hat{\boldsymbol{\alpha}}_0 = \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \hat{\boldsymbol{\phi}}(\boldsymbol{\beta}_0))$ satisfies $(\sqrt{n}/m)(\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0) = O_p(1)$, and there exists an $s \times p$ nonstochastic matrix $\mathscr{H}$ such that $\hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}_0) - \mathscr{H} = O_p(m/\sqrt{n})$.

C14. The following equations hold:

$$\mathrm{E}\left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{\Sigma}_{i,0}^{-1} \boldsymbol{D}_{i,0} \boldsymbol{h}_{1,0} \right] = O(m^4/n),$$

$$\mathrm{E}\left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{\Sigma}_{i,0}^{-1} \boldsymbol{D}_{i,0} \boldsymbol{j}_{1,0} \right] = O(m^4/n),$$

$$\mathrm{E}\left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \, \mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*} \boldsymbol{b}_{f,0}) \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{D}_{i,0} \boldsymbol{h}_{1,0} \right] = O(m^4/n),$$

$$\mathrm{E}\left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \, \mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*} \boldsymbol{b}_{f,0}) \boldsymbol{D}_{i,0} \boldsymbol{h}_{1,0} \right] = O(m^4/n),$$

$$\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\,\mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{j}_{1,0}\right]=O(m^4/n),$$

$$\mathrm{E}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\,\mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})\boldsymbol{D}_{i,0}\boldsymbol{j}_{1,0}\right]=O(m^4/n),$$

where $\boldsymbol{\mu}_{i,0}=\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)$ and $\boldsymbol{A}_{i,0}=\boldsymbol{A}_i(\boldsymbol{\beta}_0)$.

C15. $\liminf_{n\to\infty}\lambda_{\min}(\boldsymbol{B}_{nm,0}/nm)>0$, where

$$\boldsymbol{B}_{nm,0}=\sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{\Delta}_{i,0}\boldsymbol{A}_{i,0}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_i,$$

and $\boldsymbol{\Delta}_{i,0}=\boldsymbol{\Delta}_i(\boldsymbol{\beta}_0)$.

We write the definitions of $\boldsymbol{h}_{1,0}$, $\boldsymbol{j}_{1,0}$, $\boldsymbol{A}^{*}_{f,i,0}$ and $\boldsymbol{b}_{f,0}$ in the proof of Theorem 1 in Appendix. By using the moment estimator of the correlation parameters and the scale parameter, the conditions C10, C11, C12, C13 and C14 are fulfilled. Condition C15 is necessary to prove following Lemma 2:

LEMMA 2. *Suppose the conditions* C1–C15 *hold. Even if the working correlation matrix is misspecified, we have*

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0=O_p(m/\sqrt{n}).$$

PROOF. Suppose the conditions C1–C15 hold, we have

$$\begin{aligned}\boldsymbol{H}_{nm,0}&=\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{\Delta}_{i,0}\boldsymbol{A}_{i,0}^{1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{1/2}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_i\\&\geq\lambda_{\min}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))\boldsymbol{B}_{nm,0}\\&=\frac{1}{\lambda_{\max}(\boldsymbol{R}_w(\boldsymbol{\alpha}_0))}\boldsymbol{B}_{nm,0},\\\boldsymbol{M}_{nm,0}&=\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{\Delta}_{i,0}\boldsymbol{A}_{i,0}^{1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{R}_0\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{1/2}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_i\\&\leq m\{\lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))\}^2\boldsymbol{B}_{nm,0}.\end{aligned}$$

According to Lemma 1, $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\to N(\boldsymbol{0},\boldsymbol{F}_{nm,0}^{-1})$ in distribution. We calculate $\boldsymbol{F}_{nm,0}$ as follows:

$$\begin{aligned}\boldsymbol{F}_{nm,0}^{-1}&=\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{M}_{nm,0}\boldsymbol{H}_{nm,0}^{-1}\\&\leq m\{\lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))\}^2\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{nm,0}\boldsymbol{H}_{nm,0}^{-1}.\end{aligned}$$

Then, we can get the following inequality:

$$\begin{aligned}
&\boldsymbol{B}_{nm,0}^{1/2}\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{M}_{nm,0}\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{nm,0}^{1/2} \\
&\quad \leq m\{\lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))\}^2\boldsymbol{B}_{nm,0}^{1/2}\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{nm,0}\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{nm,0}^{1/2} \\
&\quad = m\{\lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))\}^2(\boldsymbol{B}_{nm,0}^{1/2}\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{nm,0}^{1/2})^2 \\
&\quad \leq m\{\lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))\}^2\{\lambda_{\max}(\boldsymbol{R}_w(\boldsymbol{\alpha}_0))\boldsymbol{I}_p\}^2.
\end{aligned}$$

Thus, we calculate $\boldsymbol{F}_{nm,0}^{-1}$ as follows:

$$\begin{aligned}
\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{M}_{nm,0}\boldsymbol{H}_{nm,0}^{-1} &\leq m\{\lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))\}^2\{\lambda_{\max}(\boldsymbol{R}_w(\boldsymbol{\alpha}_0))\}^2\boldsymbol{B}_{nm,0}^{-1} \\
&= O(m^2/n).
\end{aligned}$$

Hence, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(m/\sqrt{n}).$$

Furthermore, we evaluate the asymptotic bias of PMSEG.

THEOREM 1. *Suppose the conditions* C1–C15 *hold. The variance of the asymptotic bias of PMSEG excluding the bias independent of a candidate model goes to* 0 *with the rate of* m^4/n *or faster even if we use a wrong correlation structure as a working correlation.*

We prove Theorem 1 in Appendix.

Furthermore, to evaluate the case that we use the true correlation matrix as a working correlation matrix, we present the following lemma:

LEMMA 3. *Suppose the conditions* C1–C15 *hold. If* $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, *we have*

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(1/\sqrt{n}).$$

PROOF. Suppose that $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, we have

$$\begin{aligned}
\boldsymbol{M}_{nm,0} &= \sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{\Delta}_{i,0}\boldsymbol{A}_{i,0}^{1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{R}_0\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{1/2}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_i \\
&= \sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{\Delta}_{i,0}\boldsymbol{A}_{i,0}^{1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{1/2}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_i \\
&= \boldsymbol{H}_{nm,0}.
\end{aligned}$$

Thus, we have

$$\boldsymbol{F}_{nm,0}^{-1} = \boldsymbol{H}_{nm,0}^{-1} \leq \lambda_{\max}(\boldsymbol{R}_w(\boldsymbol{\alpha}_0))\boldsymbol{B}_{nm,0}^{-1} = O(1/n).$$

By the above, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(1/\sqrt{n}).$$

Then, we evaluate the asymptotic bias of PMSEG when we use true correlation matrix.

THEOREM 2. *Suppose the conditions* C1–C15 *hold. The variance of the asymptotic bias of PMSEG excluding the bias independent of a candidate model goes to* 0 *with the rate of* m^2/n *or faster if we use the true correlation structure as a working correlation.*

We prove Theorem 2 in Appendix.

4. Numerical study

In this section, we perform a numerical study and discuss the result. The purpose of this simulation is to compare the results by using the correct correlation structure and the results by using a wrong correlation structure. The targets of comparison are the values of each bias and the prediction errors. In this simulation, we got data from gamma distributions which have scale parameter included in exponential family. In this simulation we supposed that there are two groups (e.g., male and female). To create data distributed according to the gamma distributions with correlation, we used copula method. We set $m = 10, 20$. When $m = 10$, we set $n = 20, 50, 100$. For each $i = 1, 2, \ldots, n$, we construct a 10×8 explanatory matrix $\boldsymbol{X}_{f,i} = (\boldsymbol{x}_{f,i1}, \boldsymbol{x}_{f,i2}, \ldots, \boldsymbol{x}_{f,i10})'$. Here, for each $i = 1, \ldots, (n/2)$, the first column of $\boldsymbol{X}_{f,i}$ is $\mathbf{1}_{10}$, where $\mathbf{1}_p$ is the p-dimensional vector of ones. The second column of $\boldsymbol{X}_{f,i}$ is $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$. The third and forth columns of $\boldsymbol{X}_{f,i}$ are $\mathbf{0}_{10}$. Furthermore, all the elements of the fifth, sixth, seventh and eighth columns are independent and identically distributed according to the uniform distribution on the interval $[-1, 1]$. For each $i = (n/2) + 1, \ldots, n$, the first column of $\boldsymbol{X}_{f,i}$ is $\mathbf{1}_{10}$. The second column of $\boldsymbol{X}_{f,i}$ is $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$. The third column of $\boldsymbol{X}_{f,i}$ is $\mathbf{1}_{10}$, and the forth column of $\boldsymbol{X}_{f,i}$ is $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$. Furthermore, all the elements of the fifth, sixth, seventh and eighth columns are independent and identically distributed according to the uniform distribution on the interval $[-1, 1]$. When $m = 20$, we set $n = 80, 200, 400$. For each $i = 1, 2, \ldots, n$, we construct a 20×8 explanatory matrix $\boldsymbol{X}_{f,i} = (\boldsymbol{x}_{f,i1}, \boldsymbol{x}_{f,i2}, \ldots, \boldsymbol{x}_{f,i20})'$. Here, for each $i = 1, \ldots, (n/2)$, the first column of $\boldsymbol{X}_{f,i}$ is $\mathbf{1}_{20}$. The second column of $\boldsymbol{X}_{f,i}$ is $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0)$. The third and forth columns of $\boldsymbol{X}_{f,i}$ are $\mathbf{0}_{20}$. Furthermore, all the elements of the fifth, sixth, seventh and eighth columns are independent and

identically distributed according to the uniform distribution on the interval $[-1, 1]$. For each $i = (n/2) + 1, \ldots, n$, the first column of $\boldsymbol{X}_{f,i}$ is $\mathbf{1}_{20}$. The second column of $\boldsymbol{X}_{f,i}$ is $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0)$. The third column of $\boldsymbol{X}_{f,i}$ is $\mathbf{1}_{20}$, and the forth column of $\boldsymbol{X}_{f,i}$ is $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0)$. Furthermore, all the elements of the fifth, sixth, seventh and eighth columns are independent and identically distributed according to the uniform distribution on the interval $[-1, 1]$.

Let $\boldsymbol{\beta}_0 = (0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0, 0)'$ be the true value of regression coefficient. The explanatory matrix for the ith subject in the kth model $(k = 1, 2, \ldots, 8)$ consists of the first k columns of $\boldsymbol{X}_{f,i}$. Let the true correlation structure be the exchangeable structure, i.e., $\boldsymbol{R}_0 = (1 - \alpha)\boldsymbol{I}_m + \alpha \mathbf{1}_m \mathbf{1}_m'$. Furthermore, we set $\alpha = 0.3$. We simulate 10,000 realizations of $\boldsymbol{y} = (y_{11}, \ldots, y_{1m}, \ldots, y_{n1}, \ldots, y_{nm})$, where each y_{ij} is distributed according to the gamma distribution with the mean $\mu_{ij} = \exp(\boldsymbol{x}_{f,ij}'\boldsymbol{\beta}_0)$. Here, in order to obtain $\hat{\boldsymbol{\beta}}_f$, we used the independence working correlation matrix in this simulation.

First, we considered the case that we use the correct correlation structure. Since the bias includes Bias3 in proof of Theorem 1, to ignore Bias3, we evaluate (the bias of the 8th model) − (the bias of the each model). The frequencies of selecting models and the prediction errors are given in Table 1. In Table 1, the frequency of selecting the 6th model tends to be large when m^2/n goes to 0. Furthermore, the frequencies of selecting of the 1–5th models tend to 0. In Table 2, (the bias of the 8th model) − (the bias of the 6th model) seems to go to 0 as m^2/n goes to 0 when $m = 10$ and $m = 20$.

Next, we consider the case that we use a wrong correlation structure as a working correlation structure. We use the autoregressive structure as one of such structures. The frequency of selection of each model and prediction error are given in Table 3. Table 3 indicates that in the case of the working correlation structure is misspecified, the frequency of selecting the 6th model

Table 1. Frequencies of selecting models (%) and prediction errors

n	m	1	2	3	4	5	6	7	8	Prediction Error
20	10	10.1	6.9	5.8	4.2	15.0	25.5	15.5	17.0	7.9230 (0.04)
50	10	3.1	0.8	0.8	0.8	2.1	56.7	17.5	18.2	7.3248 (0.04)
100	10	0.1	0.0	0.2	0.2	0.3	62.1	18.9	17.2	6.9307 (0.04)
80	20	6.1	1.8	1.9	0.0	3.2	51.3	18.7	17.0	9.4235 (0.05)
200	20	0.0	0.0	0.0	0.0	0.0	58.8	22.3	18.9	9.1930 (0.05)
400	20	0.0	0.0	0.0	0.0	0.0	76.1	12.1	11.8	8.5806 (0.04)

Table 2. (The bias of the 8th model) – (The bias of the each model)

n	m	1	2	3	4	5	6	7	8
20	10	14.83	12.65	11.69	9.705	2.416	6.025	2.640	0.0
50	10	39.28	25.26	27.31	26.27	11.62	−3.193	0.993	0.0
100	10	5.483	1.858	−1.464	−2.077	0.378	1.028	0.522	0.0
80	20	177.1	186.2	172.8	179.7	48.62	12.96	2.024	0.0
200	20	−245.9	−184.0	−97.02	−85.64	−35.13	2.705	1.111	0.0
400	20	−766.9	−497.2	−314.7	−273.1	−138.2	1.024	0.490	0.0

Table 3. Frequencies of selecting models (%) and prediction errors

n	m	1	2	3	4	5	6	7	8	Prediction Error
20	10	14.5	7.5	6.9	3.8	12.5	29.2	12.5	13.1	8.0385 (0.04)
50	10	1.7	0.6	0.9	1.6	1.8	55.8	20.3	17.3	7.8406 (0.04)
100	10	0.1	0.1	0.0	0.0	0.1	69.2	19.3	11.2	7.6361 (0.04)
80	20	6.6	1.7	1.1	3.5	4.8	44.5	16.4	21.4	9.8726 (0.05)
200	20	0.2	0.0	0.0	0.1	0.2	72.2	14.9	12.4	9.9280 (0.05)
400	20	0.0	0.0	0.0	0.0	0.0	73.1	15.8	11.1	10.1993 (0.05)

Table 4. (The bias of the 8th model) – (The bias of the each model)

n	m	1	2	3	4	5	6	7	8
20	10	64.84	139.3	107.3	39.11	41.90	−172.1	−67.21	0.0
50	10	18.13	10.39	9.038	6.802	0.834	−1.311	1.238	0.0
100	10	3.335	1.409	1.878	1.288	0.471	1.135	0.4865	0.0
80	20	−649.7	−339.8	−266.8	−199.1	−71.66	−5.982	−7.344	0.0
200	20	−279.5	−187.5	−104.3	−88.66	−45.36	1.404	0.5944	0.0
400	20	−763.7	−500.4	−322.1	−283.0	−141.2	1.278	0.5243	0.0

tends to large as m^4/n is small, and the frequencies of selecting of the 1–5 models tend to 0. In Table 4, the differences between the bias of the 8th model and the bias of the 6th model and the 7th model go to 0. Furthermore, Table 4 indicates that the rate of the asymptotic bias of PMSEG m^4/n is overestimate, so we may not need so many samples.

Appendix

We prove the Theorem 1 and Theorem 2, simultaneously.

PROOF. By applying Taylor's expansion around $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$ to the equation $\boldsymbol{s}_{nm}(\hat{\boldsymbol{\beta}}) = \mathbf{0}_p$, $\boldsymbol{s}_{nm}(\hat{\boldsymbol{\beta}})$ is expanded as follows:

$$\begin{aligned}
&\boldsymbol{s}_{nm,0} + \frac{\partial \boldsymbol{s}_{nm}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad + \frac{1}{2}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_p\}\left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_{nm}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right)\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= \boldsymbol{s}_{nm,0} - \mathscr{D}_{nm,0}(\boldsymbol{I}_p + \mathscr{D}_{1,0} + \mathscr{D}_{2,0})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad + \frac{1}{2}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_p\}\boldsymbol{L}_1(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= \mathbf{0}_p,
\end{aligned}$$

where $\boldsymbol{\beta}^*$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$, and $\boldsymbol{s}_{nm,0} = \boldsymbol{s}_{nm}(\boldsymbol{\beta}_0)$. Here, $\boldsymbol{L}_1(\boldsymbol{\beta}^*)$, $\mathscr{D}_{nm,0}$, $\mathscr{D}_{1,0}$ and $\mathscr{D}_{2,0}$ are defined as follows:

$$\boldsymbol{L}_1(\boldsymbol{\beta}^*) = \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_{nm}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right)\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}, \qquad \mathscr{D}_{nm,0} = \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \boldsymbol{\Gamma}_{i,0}^{-1} \boldsymbol{D}_{i,0},$$

$$\mathscr{D}_{1,0} = -\mathscr{D}_{nm,0}^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{\Gamma}_i^{-1}(\boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\{\boldsymbol{I}_p \otimes (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\},$$

$$\mathscr{D}_{2,0} = -\mathscr{D}_{nm,0}^{-1} \sum_{i=1}^{n} \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{D}_i'(\boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)[\boldsymbol{I}_p \otimes \{\boldsymbol{\Gamma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}],$$

where $\boldsymbol{\Gamma}_{i,0} = \boldsymbol{\Gamma}_i(\boldsymbol{\beta}_0)$. By Lindberg central limit theorem, it holds that $\boldsymbol{L}_1(\boldsymbol{\beta}^*) = O_p(nm)$ and $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(m/\sqrt{n})$. Furthermore, if $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, we have $\boldsymbol{L}_1(\boldsymbol{\beta}^*) = O_p(nm^{1/2})$ and $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(1/\sqrt{n})$. Moreover, $\boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0)$ is expanded as follows:

$$\boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) = \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) + \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) + O_p(m^2/n). \quad \text{(A.4)}$$

By Taylor's theorem, since $\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0 = O_p(m/\sqrt{n})$, it holds that

$$\|\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\| \le \left\|\frac{\partial}{\partial \boldsymbol{\alpha}} \otimes \boldsymbol{R}_w(\boldsymbol{\alpha})\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}\right\| \|\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0\| = O_p(m/\sqrt{n}),$$

i.e., $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0) = O_p(m/\sqrt{n})$, where $\boldsymbol{\alpha}^*$ lies between $\boldsymbol{\alpha}_0$ and $\hat{\boldsymbol{\alpha}}_0$. If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, we have $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0) = O_p(1/\sqrt{n})$ and the third term of

(A.4) is $O_p(1/n)$. Hence, it holds that

$$\begin{aligned}\mathscr{D}_{nm,0} &= \sum_{i=1}^{n} \boldsymbol{D}'_{i,0}\boldsymbol{\Gamma}_{i,0}^{-1}\boldsymbol{D}_{i,0} \\ &= \sum_{i=1}^{n} \boldsymbol{D}'_{i,0}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0)\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\boldsymbol{D}_{i,0} \\ &= \boldsymbol{H}_{nm,0} + O_p(m^2 n^{1/2}).\end{aligned}$$

Thus, by using the fact that $\boldsymbol{s}_{nm,0} = \boldsymbol{q}_{nm,0} + O_p(m^2)$, $\hat{\boldsymbol{\beta}}$ is expanded as follows:

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{H}_{nm,0}^{-1}\boldsymbol{q}_{nm,0} + O_p(m^3/n) = \boldsymbol{b}_{1,0} + O_p(m^3/n),$$

where $\boldsymbol{q}_{nm,0} = \boldsymbol{q}_{nm}(\boldsymbol{\beta}_0)$. Also, since

$$\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\boldsymbol{\phi}}(\boldsymbol{\beta})))\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right) - \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\boldsymbol{\phi}}(\boldsymbol{\beta})))\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] = O_p(m/\sqrt{n}),$$

the GEE substituted in $\boldsymbol{\beta}_0$ is expanded as follows:

$$\begin{aligned}\boldsymbol{s}_{nm,0} = &-\sum_{i=1}^{n} \boldsymbol{D}'_{i,0}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \\ &+ \boldsymbol{H}_{nm,0}(\boldsymbol{I}_p + \boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{1,0} + \boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{2,0} + \boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{3,0})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &+ \sum_{i=1}^{n} \boldsymbol{D}'_{i,0}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &- \frac{1}{2}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_p\}\{\mathscr{S}_{1,0} + (\boldsymbol{L}_1(\boldsymbol{\beta}_0) - \mathscr{S}_{1,0})\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &- \frac{1}{6}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_p\}\left\{\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_{nm}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right)\right\}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{**}} \\ &\cdot \{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}, \qquad \text{(A.5)}\end{aligned}$$

where $\boldsymbol{\beta}^{**}$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$, and $\mathscr{S}_{1,0} = \mathrm{E}[\boldsymbol{L}_1(\boldsymbol{\beta}_0)]$. We define $\boldsymbol{B}_{1,0}$, $\boldsymbol{B}_{2,0}$ and $\boldsymbol{B}_{3,0}$ as follows:

$$\boldsymbol{B}_{1,0} = \sum_{i=1}^{n} \boldsymbol{X}_i' \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{\Delta}_i(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right) \{ \boldsymbol{I}_p \otimes \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) \boldsymbol{A}_{i,0}^{-1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \},$$

$$\boldsymbol{B}_{2,0} = \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{\Delta}_{i,0} \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right) \{ \boldsymbol{I}_p \otimes \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) \boldsymbol{A}_{i,0}^{-1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \},$$

$$\boldsymbol{B}_{3,0} = \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{\Delta}_{i,0} \boldsymbol{A}_{i,0}^{1/2} \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right) \{ \boldsymbol{I}_p \otimes (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \}.$$

Here, we calculate the rate of $\boldsymbol{B}_{1,0}$.

$$\begin{aligned}
\mathrm{E}[\boldsymbol{B}_{1,0} \boldsymbol{B}_{1,0}'] &= \mathrm{E}\left[\sum_{i=1}^{n} \sum_{k=1}^{p} \boldsymbol{X}_i' \frac{\partial \boldsymbol{\Delta}_i(\boldsymbol{\beta})}{\partial \beta_k} \boldsymbol{A}_{i,0}^{1/2} \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) \boldsymbol{A}_{i,0}^{-1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \right. \\
&\qquad \left. \cdot (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) \boldsymbol{A}_{i,0}^{1/2} \frac{\partial \boldsymbol{\Delta}_i(\boldsymbol{\beta})}{\partial \beta_k} \boldsymbol{X}_i \right] \\
&= \sum_{i=1}^{n} \sum_{k=1}^{p} \boldsymbol{X}_i' \frac{\partial \boldsymbol{\Delta}_i(\boldsymbol{\beta})}{\partial \beta_k} \boldsymbol{A}_{i,0}^{1/2} \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) \boldsymbol{R}_0 \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) \boldsymbol{A}_{i,0}^{1/2} \frac{\partial \boldsymbol{\Delta}_i(\boldsymbol{\beta})}{\partial \beta_k} \boldsymbol{X}_i \\
&\le m \{ \lambda_{\max}(\boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0)) \}^2 \sum_{i=1}^{n} \sum_{k=1}^{p} \boldsymbol{X}_i' \frac{\partial \boldsymbol{\Delta}_i(\boldsymbol{\beta})}{\partial \beta_k} \boldsymbol{A}_{i,0} \frac{\partial \boldsymbol{\Delta}_i(\boldsymbol{\beta})}{\partial \beta_k} \boldsymbol{X}_i \\
&= O(nm^2).
\end{aligned}$$

Thus, $\boldsymbol{B}_{1,0} = O_p(n^{1/2}m)$. Similarly, we calculate $\boldsymbol{B}_{2,0} = O_p(n^{1/2}m)$ and $\boldsymbol{B}_{3,0} = O_p(n^{1/2}m)$. Furthermore, if $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, we have $\boldsymbol{B}_{1,0} = O_p(n^{1/2}m^{1/2})$, $\boldsymbol{B}_{2,0} = O_p(n^{1/2}m^{1/2})$ and $\boldsymbol{B}_{3,0} = O_p(n^{1/2}m^{1/2})$. By (A.5), we have

$$\begin{aligned}
&\boldsymbol{H}_{nm,0}(\boldsymbol{I}_p + \boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{1,0} + \boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{2,0} + \boldsymbol{H}_{nm,0}^{-1}\boldsymbol{B}_{3,0})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad = \boldsymbol{q}_{nm,0} + \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \{ \boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0) \} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \boldsymbol{A}_{i,0}^{-1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \\
&\qquad - \sum_{i=1}^{n} \boldsymbol{D}_{i,0}' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \{ \boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0) \} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \\
&\qquad + \frac{1}{2} (\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_p) \mathscr{S}_{1,0} \boldsymbol{b}_{1,0} \\
&\qquad + O_p(m^4/\sqrt{n}).
\end{aligned}$$

From the above, we expand $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ as follows:

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{H}_{nm,0}^{-1}\boldsymbol{q}_{nm,0} + \frac{1}{2}\boldsymbol{H}_{nm,0}^{-1}(\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_p)\mathscr{S}_{1,0}\boldsymbol{b}_{1,0} + \boldsymbol{H}_{nm,0}^{-1}(\boldsymbol{B}_{1,0} + \boldsymbol{B}_{2,0} + \boldsymbol{B}_{3,0})\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{q}_{nm,0} + \boldsymbol{h}_{1,0} + \boldsymbol{j}_{1,0} + O_p(m^4/n^{3/2}),$$

where

$$\boldsymbol{j}_{1,0} = \boldsymbol{H}_{nm,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}),$$

$$\boldsymbol{h}_{1,0} = \boldsymbol{H}_{nm,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}.$$

Denote

$$\boldsymbol{b}_{2,0} = \boldsymbol{H}_{nm,0}^{-1}(\boldsymbol{B}_{1,0} + \boldsymbol{B}_{2,0} + \boldsymbol{B}_{3,0})\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{q}_{nm,0},$$

$$\boldsymbol{b}_{3,0} = \boldsymbol{H}_{nm,0}^{-1}(\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_p)\mathscr{S}_{1,0}\boldsymbol{b}_{1,0}/2 + \boldsymbol{h}_{1,0} + \boldsymbol{j}_{1,0}.$$

Hence, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{b}_{1,0} + \boldsymbol{b}_{2,0} + \boldsymbol{b}_{3,0} + O_p(m^4/n^{3/2}). \tag{A.6}$$

Note that, $\boldsymbol{b}_{1,0} = O_p(m/\sqrt{n})$, $\boldsymbol{b}_{2,0} = O_p(m^2/n)$ and $\boldsymbol{b}_{3,0} = O_p(m^3/n)$. Furthermore, if $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{b}_{1,0} + \boldsymbol{b}_{2,0} + \boldsymbol{b}_{3,0} + O_p(m^2/n^{3/2}),$$

where $\boldsymbol{b}_{1,0} = O_p(1/\sqrt{n})$, $\boldsymbol{b}_{2,0} = O_p(m/n)$ and $\boldsymbol{b}_{3,0} = O_p(m/n)$.

We calculated the asymptotic bias of PMSEG as follows:

$$\begin{aligned}\text{Bias} &= \text{PMSE} - \text{E}_y[\mathscr{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)] \\ &= \{\text{Risk}_P - \text{E}_y[\mathscr{L}^*(\hat{\boldsymbol{\beta}})]\} + \{\text{E}_y[\mathscr{L}^*(\hat{\boldsymbol{\beta}})] - \text{E}_y[\mathscr{L}^*]\} \\ &\quad + \{\text{E}_y[\mathscr{L}^*] - \text{E}_y[\mathscr{L}(\hat{\boldsymbol{\beta}}_f)]\} + \{\text{E}_y[\mathscr{L}^*(\hat{\boldsymbol{\beta}}_f)] - \text{E}_y[\mathscr{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)]\} \\ &= \text{Bias1} + \text{Bias2} + \text{Bias3} + \text{Bias4}.\end{aligned}$$

We evaluate Bias1, Bias2, Bias3 and Bias4 separately.

Bias1 is expanded as follows:

$$\begin{aligned}
\text{Bias1} &= \mathrm{E}_y\left[\mathrm{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i-\hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i-\hat{\boldsymbol{\mu}}_i)\right]-\sum_{i=0}^{n}(\boldsymbol{y}_i-\hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i-\hat{\boldsymbol{\mu}}_i)\right] \\
&= \mathrm{E}_y\left[\mathrm{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_{i,0}+\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_{i,0}+\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]\right. \\
&\qquad \left.-\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0}+\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0}+\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right] \\
&= \mathrm{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_{i,0})\right]+\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right] \\
&\quad -\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\right]-2\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right] \\
&\quad -\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right] \\
&= 2\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_{i,0})\right]. \qquad \text{(A.7)}
\end{aligned}$$

Since $\hat{\boldsymbol{\mu}}_i$ is the function of $\hat{\boldsymbol{\beta}}$, by applying Taylor's expansion around $\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\mu}}_i$ is expanded as follows:

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_{i,0} &= \left.\frac{\partial\boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\
&\quad +\frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\otimes\boldsymbol{I}_m\}\left.\left(\frac{\partial}{\partial\boldsymbol{\beta}}\otimes\frac{\partial\boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}\right)\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\
&\quad +\frac{1}{6}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\otimes\boldsymbol{I}_m\}\left.\left\{\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\left(\frac{\partial}{\partial\boldsymbol{\beta}}\otimes\frac{\partial\boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}\right)\right\}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{***}} \\
&\quad \cdot\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\otimes(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\} \\
&= \boldsymbol{D}_{i,0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\otimes\boldsymbol{I}_m\}\boldsymbol{D}_{i,0}^{(1)}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\
&\quad +O_p(m^{7/2}/n^{3/2}), \qquad \text{(A.8)}
\end{aligned}$$

where $\boldsymbol{\beta}^{***}$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$, and $\boldsymbol{D}_{i,0}^{(1)}$ is defined by

$$\boldsymbol{D}_{i,0}^{(1)}=\left.\left(\frac{\partial}{\partial\boldsymbol{\beta}}\otimes\boldsymbol{D}_i(\boldsymbol{\beta})\right)\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.$$

By substituting (A.6) for (A.8), we can expand $\hat{\boldsymbol{\mu}}_i$ as follows:

$$\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{i,0}(\boldsymbol{b}_{1,0} + \boldsymbol{b}_{2,0} + \boldsymbol{b}_{3,0}) + \frac{1}{2}(\boldsymbol{b}'_{1,0} \otimes \boldsymbol{I}_m)\boldsymbol{D}^{(1)}_{i,0}\boldsymbol{b}_{1,0} + O_p(m^{7/2}/n^{3/2}). \tag{A.9}$$

By using (A.7) and (A.9), we get the following expansion:

$$\begin{aligned}
\frac{1}{2}\text{Bias1} &= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}^{-1}_{i,0}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0})\right] \\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}^{-1}_{i,0}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right] \\
&\quad + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}^{-1}_{i,0}\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0}\right] \\
&\quad + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}^{-1}_{i,0}\boldsymbol{D}_{i,0}\boldsymbol{b}_{3,0}\right] \\
&\quad + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}^{-1}_{i,0}(\boldsymbol{b}'_{1,0} \otimes \boldsymbol{I}_m)\boldsymbol{D}^{(1)}_{i,0}\boldsymbol{b}_{1,0}\right] \\
&\quad + \mathrm{E}_y[O_p(n^{-1/2}m^{7/2})].
\end{aligned} \tag{A.10}$$

Same as Inatsu and Sato [5], the first term of (A.10) is calculated as follows:

$$\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}^{-1}_{i,0}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right] = p.$$

Since the data from different two subjects are independent, we calculate the second term of (A.10) as follows:

$$\begin{aligned}
&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}^{-1}_{i,0}\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0}\right] \\
&\quad = \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}^{-1}_{i,0}\boldsymbol{D}_{i,0}\boldsymbol{H}^{-1}_{nm,0}\boldsymbol{G}_0\boldsymbol{H}^{-1}_{nm,0}\boldsymbol{D}'_{i,0}\boldsymbol{V}^{-1}_{i,0}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] \\
&\quad = \mathrm{tr}\left(\sum_{i=1}^{n}\boldsymbol{H}^{-1}_{nm,0}\boldsymbol{G}_0\boldsymbol{H}^{-1}_{nm,0}\boldsymbol{D}'_{i,0}\boldsymbol{V}^{-1}_{i,0}\boldsymbol{D}_{i,0}\right) \\
&\quad = O(m^2/n),
\end{aligned}$$

where $\boldsymbol{G}_0 = \boldsymbol{B}_{1,0} + \boldsymbol{B}_{2,0} + \boldsymbol{B}_{3,0}$. If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, the second term of (A.10) is $O(m^2/n)$. Similarly, the orders of the third and the forth term of (A.10) are evaluated as follows:

$$\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{3,0}\right] = O(m^{7/2}/n),$$

$$\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right] = O(m^{5/2}/n).$$

Furthermore, if $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, the order of the third term of (A.10) is $O(m^{3/2}/n)$ and the order of the forth term of (A.10) is $O(\sqrt{m}/n)$. Under the regularity conditions, the limit of expectation is equal to the expectation of limit. Furthermore, in many cases, a moment of statistic can be expanded as power series in n^{-1} (e.g., Hall [3]). Therefore, we obtain

$$\mathrm{Bias1} = 2p + O(m^{7/2}/n).$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, we have

$$\mathrm{Bias1} = 2p + O(m^{3/2}/n).$$

Similarly, we calculate Bias2 + Bias4. Now, Bias2 and Bias4 are expressed as follows:

$$\begin{aligned}
\mathrm{Bias2} &= \mathrm{E}_y[\mathscr{L}^*(\hat{\boldsymbol{\beta}})] - \mathrm{E}_y[\mathscr{L}^*(\boldsymbol{\beta}_0)] \\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i) - \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] \\
&= \mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
&\quad + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right],
\end{aligned}$$

$$\begin{aligned}
\mathrm{Bias4} &= \mathrm{E}_y[\mathscr{L}(\boldsymbol{\beta}_0, \hat{\boldsymbol{\beta}}_f)] - \mathrm{E}_y[\mathscr{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)] \\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right] \\
&\quad - \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right]
\end{aligned}$$

$$= -\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f)\right]$$

$$-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)'\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f)\right].$$

Hence, Bias2 + Bias4 is

$$\text{Bias2} + \text{Bias4} = \mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\{\boldsymbol{\Sigma}_{i,0}^{-1}-\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f)\}\right.$$

$$\left.\cdot(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right] \tag{A.11}$$

$$+\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)'\{\boldsymbol{\Sigma}_{i,0}^{-1}-\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f)\}\right.$$

$$\left.\cdot(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]. \tag{A.12}$$

Then, we perform the stochastic expansion of $\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)$, $\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)$, $\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_f)$, $\hat{\boldsymbol{\beta}}_f$ and $\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\beta}}_f)$. The expansion of $\hat{\boldsymbol{\beta}}_f$ is as follows:

$$\hat{\boldsymbol{\beta}}_f-\boldsymbol{\beta}_{f,0}=\boldsymbol{H}_{f,nm,0}^{-1}\boldsymbol{q}_{f,nm}(\boldsymbol{\beta}_{f,0})+O_p(m^3/n)=\boldsymbol{b}_{f,0}+O_p(m^3/n),$$

where $\boldsymbol{\beta}_{f,0}$ is the true value of $\boldsymbol{\beta}_f$, $\boldsymbol{b}_{f,0}=\boldsymbol{H}_{f,nm,0}^{-1}\boldsymbol{q}_{f,nm}(\boldsymbol{\beta}_{f,0})$,

$$\boldsymbol{q}_{f,nm}(\boldsymbol{\beta}_f)=\sum_{i=1}^{n}\boldsymbol{D}_{f,i}'(\boldsymbol{\beta}_f)\boldsymbol{V}_i^{-1}(\boldsymbol{\beta}_f,\boldsymbol{\alpha}_{f,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_f)),$$

$\boldsymbol{\alpha}_{f,0}$ is the convergence value of a correlation parameter in the full model and $\boldsymbol{D}_{f,i}(\boldsymbol{\beta}_f)=\boldsymbol{A}_i(\boldsymbol{\beta}_f)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_f)\boldsymbol{X}_{f,i}$. Here, $\boldsymbol{H}_{f,nm,0}$ is

$$\boldsymbol{H}_{f,nm,0}=\sum_{i=1}^{n}\boldsymbol{D}_{f,i,0}'\boldsymbol{A}_{i,0}^{-1/2}\overline{\boldsymbol{R}}_i^{-1}(\boldsymbol{\alpha}_f)\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{f,i,0},$$

where $\boldsymbol{D}_{f,i,0}=\boldsymbol{A}_{i,0}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_{f,i}$ and $\overline{\boldsymbol{R}}_i^{-1}(\boldsymbol{\alpha}_f)$ is a working correlation matrix which can be chosen freely including a nuisance correlation parameter $\boldsymbol{\alpha}_f$. Furthermore, if $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, we have

$$\hat{\boldsymbol{\beta}}_f-\boldsymbol{\beta}_{f,0}=\boldsymbol{H}_{f,nm,0}^{-1}\boldsymbol{q}_{f,nm}(\boldsymbol{\beta}_{f,0})+O_p(m/n)=\boldsymbol{b}_{f,0}+O_p(m/n).$$

Thus, we can expand $\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_f)$ as follows:

$$\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_f) - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{f,i,0}\boldsymbol{b}_{f,0} + O_p(m^{7/2}/n).$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, we have

$$\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_f) - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{f,i,0}\boldsymbol{b}_{f,0} + O_p(m^{3/2}/n).$$

Furthermore, $\boldsymbol{a}_{f,i}(\boldsymbol{\beta}_f)$ is the m-dimensional vector consisting of the diagonal components of $\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{\beta}_f)$, i.e., $\mathrm{diag}(\boldsymbol{a}_{f,i}(\boldsymbol{\beta}_f)) = \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_f)$. Then, we can perform Taylor's expansion of $\boldsymbol{a}_{f,i}(\hat{\boldsymbol{\beta}}_f)$ around $\hat{\boldsymbol{\beta}}_f = \boldsymbol{\beta}_{f,0}$ as follows:

$$\boldsymbol{a}_{f,i}(\hat{\boldsymbol{\beta}}_f) = \boldsymbol{a}_{f,i}(\boldsymbol{\beta}_{f,0}) + \boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0} + O_p(m^3/n),$$

where

$$\boldsymbol{A}^*_{f,i,0} = \frac{\partial}{\partial \boldsymbol{\beta}_f'}\boldsymbol{a}_{f,i}(\boldsymbol{\beta}_f)\bigg|_{\boldsymbol{\beta}_f=\boldsymbol{\beta}_{f,0}}.$$

Therefore, we can expand $\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)$ as follows:

$$\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f) = \mathrm{diag}(\boldsymbol{a}_{f,i}(\hat{\boldsymbol{\beta}}_f)) = \boldsymbol{A}_{i,0}^{-1/2} + \mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0}) + O_p(m^3/n).$$

Note that $\boldsymbol{b}_{f,0} = O_p(m/\sqrt{n})$, $\boldsymbol{D}_{f,i,0}\boldsymbol{b}_{f,0} = O_p(m^{3/2}/\sqrt{n})$ and $\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0}) = O_p(m/\sqrt{n})$. If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, we have $\boldsymbol{b}_{f,0} = O_p(1/\sqrt{n})$, $\boldsymbol{D}_{f,i,0}\boldsymbol{b}_{f,0} = O_p(\sqrt{m}/\sqrt{n})$ and $\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0}) = O_p(1/\sqrt{n})$. Moreover, we can expand $\hat{\phi}(\hat{\boldsymbol{\beta}}_f)$ as follows:

$$\hat{\phi}(\hat{\boldsymbol{\beta}}_f) = \phi_0 + O_p(m/\sqrt{n}).$$

Furthermore, same as Inatsu and Sato [5], $\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)$ is expanded as follows:

$$\begin{aligned}
\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f) = \boldsymbol{R}_0^{-1} + \boldsymbol{R}_0^{-1}\Bigg\{ & \boldsymbol{R}_0 - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
& - \frac{1}{n}\sum_{i=1}^{n}\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
& - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\,\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
& - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}(\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1})\Bigg\}\boldsymbol{R}_0^{-1} \\
& + O_p(m^3/n). \qquad \text{(A.13)}
\end{aligned}$$

Note that the second term of (A.13) is $O_p(m/\sqrt{n})$. Then, we have

$$
\begin{aligned}
&\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f) \\
&\quad = \boldsymbol{\Sigma}_{i,0}^{-1} - \{\boldsymbol{A}_{i,0}^{-1/2} + \operatorname{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\} \\
&\qquad \cdot \left[\boldsymbol{R}_0^{-1} + \boldsymbol{R}_0^{-1}\left\{\boldsymbol{R}_0 - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\right.\right. \\
&\qquad - \frac{1}{n}\sum_{i=1}^{n}\operatorname{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&\qquad - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\operatorname{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
&\qquad \left.\left. - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}(\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1})\right\}\boldsymbol{R}_0^{-1}\right] \\
&\qquad \cdot \{\boldsymbol{A}_{i,0}^{-1/2} + \operatorname{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\}\{\phi_0^{-1} + (\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1})\} \\
&\qquad + O_p(m^3/n) \\
&\quad = -\operatorname{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} - \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\operatorname{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
&\qquad - \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\left\{\boldsymbol{R}_0 - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\right. \\
&\qquad - \frac{1}{n}\sum_{i=1}^{n}\operatorname{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&\qquad - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\operatorname{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
&\qquad \left. - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}(\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1})\right\}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&\qquad - \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_i^{-1/2}\{\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1}\} \\
&\qquad + O_p(m^3/n),
\end{aligned}
$$

where $\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\beta}}_f) = O_p(m/\sqrt{n})$ and $\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0} = O_p(m^{3/2}/\sqrt{n})$. Then, (A.12) is calculated as follows:

$$\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)'\{\boldsymbol{\Sigma}_{i,0}^{-1}-\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f)\}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]$$
$$=\mathrm{E}_y[O_p(m^4/\sqrt{n})]$$
$$=O(m^4/n).$$

Furthermore, if $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, we have

$$\boldsymbol{\Sigma}_{i,0}^{-1}-\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\beta}}_f)=O_p(1/\sqrt{n}),$$

and $\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_{i,0}=\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}=O_p(\sqrt{m}/\sqrt{n})$. Thus, the order of (A.12) is $O(m/n)$. In addition, we calculate (A.11).

$$\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\{\boldsymbol{\Sigma}_{i,0}^{-1}-\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f)\}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]$$
$$=\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\{\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\right.$$
$$\left.+\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\,\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\phi_0^{-1}\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$
$$-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$
$$\left.\cdot\,\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$
$$-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\mathrm{diag}(\boldsymbol{A}_{f,j,0}^{*}\boldsymbol{b}_{f,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$
$$\left.\cdot\,\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$
$$-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$
$$\left.\cdot\,\mathrm{diag}(\boldsymbol{A}_{f,j,0}^{*}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$

$$
\begin{aligned}
&- \mathrm{E}_y \left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \frac{2}{n} \sum_{j=1}^{n} \boldsymbol{A}_{j,0}^{-1/2} (\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})' \boldsymbol{A}_{j,0}^{-1/2} \right. \\
&\qquad \left. \cdot \{ \hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1} \} \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \phi_0^{-1} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \right] \\
&+ \mathrm{E}_y \left[2 \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \{ \hat{\boldsymbol{\phi}}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1} \} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \right] \\
&+ \mathrm{E}_y \left[2 \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i}^{-1/2} \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \phi_0^{-1} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \right] + O(m^4/n). \qquad \text{(A.14)}
\end{aligned}
$$

Note that $\mathrm{E}[(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \otimes (\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})'(\boldsymbol{y}_k - \boldsymbol{\mu}_{k,0})] = \boldsymbol{0}_m$ (not $i = j = k$), so we can expand the first term of (A.14) as follows:

$$
\begin{aligned}
&\mathrm{E}_y \left[2 \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \{ \mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*} \boldsymbol{b}_{f,0}) \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \phi_0^{-1} \right. \\
&\qquad \left. + \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \, \mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*} \boldsymbol{b}_{f,0}) \phi_0^{-1} \} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \right] \\
&= \mathrm{E}_y \left[2 \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \{ \mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*} \boldsymbol{b}_{f,i,0}) \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \phi_0^{-1} \right. \\
&\qquad \left. + \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \, \mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*} \boldsymbol{b}_{f,i,0}) \phi_0^{-1} \} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \right] \\
&= O(m^4/n), \qquad \text{(A.15)}
\end{aligned}
$$

where $\boldsymbol{b}_{f,i,0} = \boldsymbol{H}_{f,nm,0}^{-1} \boldsymbol{D}_{f,i}'(\boldsymbol{\beta}_{f,0}) \boldsymbol{V}_i^{-1}(\boldsymbol{\beta}_{f,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{f,0}))$. Moreover, if $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, the order of the first term of (A.14) is $O(m^2/n)$. Similarly, since $\mathrm{E}_y[(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})'(\boldsymbol{y}_k - \boldsymbol{\mu}_{k,0})] = 0$ (unless $i = k$), the second term of (A.14) is expanded as follows:

$$
\begin{aligned}
&- \mathrm{E}_y \left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \frac{2}{n} \sum_{j=1}^{n} \boldsymbol{A}_{j,0}^{-1/2} (\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})' \right. \\
&\qquad \left. \cdot \boldsymbol{A}_{j,0}^{-1/2} \boldsymbol{R}_0^{-1} \phi_0^{-2} \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \right]
\end{aligned}
$$

$$= -\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1,i\neq j}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$

$$\left.\cdot\,\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]$$

$$+\,O(m^4/n)$$

$$= -\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]+O(m^4/n)$$

$$= -2p+O(m^4/n), \tag{A.16}$$

where $\boldsymbol{b}_{1i,0}=\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{D}_{i,0}'\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})=O_p(m/n)$. If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, we have

$$-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$

$$\left.\cdot\,\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$

$$= -2p+O(m^2/n).$$

Here, we will use a kind of abbreviation for summations such as:

$$\sum_{i,j}=\sum_{i=1}^{n}\sum_{j=1}^{n},$$

$$\sum_{i\neq j}=\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}.$$

It holds that $\mathrm{E}_y[(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0}\otimes\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0}')(\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0}\otimes\boldsymbol{y}_l-\boldsymbol{\mu}_{l,0})]=0$ unless the following condition:

$i=j=l$ or $i=j\neq k=l$ or $i=l\neq k=j$ or $j=l\neq k=i$.

Thus, the third term of (A.14) is calculated as follows:

$$-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\mathrm{diag}(\boldsymbol{A}_{f,j,0}^{*}\boldsymbol{b}_{f,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$

$$\left.\cdot\,\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$

$$= -\mathrm{E}_y\left[\sum_{i,j}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\,\mathrm{diag}(\boldsymbol{A}_{f,j,0}^{*}\boldsymbol{b}_{f,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$

$$\left.\cdot\,\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$

$$= -\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\,\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\right.$$

$$\left.\cdot\,\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$

$$-\mathrm{E}_y\left[\sum_{i\neq j}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\,\mathrm{diag}(\boldsymbol{A}_{f,j,0}^{*}\boldsymbol{b}_{f,i,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$

$$\left.\cdot\,\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1j,0}\right]$$

$$-\mathrm{E}_y\left[\sum_{i\neq j}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\,\mathrm{diag}(\boldsymbol{A}_{f,j,0}^{*}\boldsymbol{b}_{f,j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$

$$\left.\cdot\,\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]$$

$$+\,O(m^4/n)$$

$$= O(m^4/n). \tag{A.17}$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, the order of the third term of (A.14) is $O(m^2/n)$. Similarly, the forth term of (A.14) is expanded as follows:

$$-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\right.$$

$$\left.\cdot\,\mathrm{diag}(\boldsymbol{A}_{f,j,0}^{*}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]$$

$$= O(m^4/n). \tag{A.18}$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, the order of the forth term of (A.14) is $O(m^2/n)$. The fifth term of (A.14) is calculated as follows:

$$
\begin{aligned}
&-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\boldsymbol{A}_{j,0}^{-1/2}\right.\\
&\qquad\left.\cdot\{\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1}\}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&=-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\right.\\
&\qquad\left.\cdot\frac{\partial\hat{\boldsymbol{\phi}}(\boldsymbol{\beta}_f)}{\partial\boldsymbol{\beta}_f}\bigg|_{\boldsymbol{\beta}_f=\boldsymbol{\beta}_{f,0}}\boldsymbol{b}_{f,j,0}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1j,0}\right]\\
&\quad-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\boldsymbol{A}_{j,0}^{-1/2}\right.\\
&\qquad\left.\cdot\frac{\partial\hat{\boldsymbol{\phi}}(\boldsymbol{\beta}_f)}{\partial\boldsymbol{\beta}_f}\bigg|_{\boldsymbol{\beta}_f=\boldsymbol{\beta}_{f,0}}\boldsymbol{b}_{f,i,0}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]\\
&\quad-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\boldsymbol{A}_{j,0}^{-1/2}\right.\\
&\qquad\left.\cdot\frac{\partial\hat{\boldsymbol{\phi}}(\boldsymbol{\beta}_f)}{\partial\boldsymbol{\beta}_f}\bigg|_{\boldsymbol{\beta}_f=\boldsymbol{\beta}_{f,0}}\boldsymbol{b}_{f,i,0}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1j,0}\right]\\
&\quad-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\boldsymbol{A}_{j,0}^{-1/2}\right.\\
&\qquad\left.\cdot\frac{\partial\hat{\boldsymbol{\phi}}(\boldsymbol{\beta}_f)}{\partial\boldsymbol{\beta}_f}\bigg|_{\boldsymbol{\beta}_f=\boldsymbol{\beta}_{f,0}}\boldsymbol{b}_{f,j,0}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]\\
&=O(m^4/n).
\end{aligned}
\tag{A.19}
$$

The sixth term of (A.14) is calculated as follows:

$$
\begin{aligned}
&\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\{\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1}\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&\quad=\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\frac{\partial\hat{\boldsymbol{\phi}}(\boldsymbol{\beta}_f)}{\partial\boldsymbol{\beta}_f}\bigg|_{\boldsymbol{\beta}_f=\boldsymbol{\beta}_{f,0}}\boldsymbol{b}_{f,i,0}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]\\
&\quad=O(m^3/n).
\end{aligned}
\tag{A.20}
$$

Furthermore, the seventh term of (A.14) is calculated as follows:

$$\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]=2p. \tag{A.21}$$

By (A.15)–(A.21), (A.11) is calculated as follows:

$$\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\{\boldsymbol{\Sigma}_{i,0}^{-1}-\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]$$
$$=O(m^4/n).$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, the order of (A.11) is $O(m^2/n)$. Thus, we have

$$\mathrm{Bias2}+\mathrm{Bias4}=O(m^4/n).$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, we have $\mathrm{Bias2}+\mathrm{Bias4}=O(m^2/n)$. From the above, the bias is expanded as follows:

$$\mathrm{Bias}=2p+\mathrm{Bias3}+O(m^4/n).$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, the bias is expanded as follows:

$$\mathrm{Bias}=2p+\mathrm{Bias3}+O(m^2/n).$$

Note that Bias3 does not depend on the candidate model. If we ignore Bias3, the asymptotic bias of PMSEG goes to 0 with the rate of m^4/n or faster. Furthermore, if we use the true correlation structure as a working correlation, the asymptotic bias of PMSEG goes to 0 with the rate of m^2/n or faster.

Acknowledgement

The author is deeply grateful to Professor Hirofumi Wakaki of Hiroshima University for their helpful comments and suggestions.

References

[1] H. Akaike, Information theory and an extension of the maximum likelihood principle, In 2nd International Symposium on Information Theory (eds. B. N. Petrov & F. Csáki), (1973), 267–281, Akadémiai Kiadó, Budapest.

[2] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automatic Control, **AC-19** (1974), 716–723.

[3] P. Hall, The Bootstrap and Edgeworth Expansion, Springer-Verlag, New York, 1992.

[4] Y. Inatsu and S. Imori, Model selection criterion based on the prediction mean squared error in generalized estimating equations, Hiroshima Math. J., **48** (2018), 307–334.

[5] Y. Inatsu and T. Sato, A C_p type criterion for model selection in the GEE method when both scale and correlation parameters are unknown, TR17-08, Statistical Research Group, Hiroshima University, Hiroshima.
[6] K. Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika, **73** (1986), 13–22.
[7] C. L. Mallows, Some comments on C_p, Technometrics, **15** (1973), 661–675.
[8] R. Nishii, Asymptotic properties of criteria for selecting of variables in multiple regression, Ann. Statist., **12** (1984), 758–765.
[9] W. Pan, Akaike's information criterion in generalized estimating equations, Biometrics, **57** (2001), 120–125.
[10] C. R. Rao and Y. Wu, A strongly consistent procedure for model selection in a regression problem, Biometrika, **76** (1989), 369–374.
[11] M. Xie and Y. Yang, Asymptotics for generalized estimating equations with large cluster sizes, Ann. Statist., **31** (2003), 310–347.

Tomoharu Sato
Department of Mathematics
Graduate School of Science
Hiroshima University
Higashi-Hiroshima 739-8526 *Japan*
E-mail: d171681@hiroshima-u.ac.jp