広島大学学位請求論文

C_p type criterion for model selection in the generalized estimating equation method

(一般化推定方程式における C_p 型モデル選択規準)

2019年

広島大学大学院理学研究科
数学専攻

佐藤 倫治

目次

主論文

C_p type criterion for model selection in the generalized estimating equation method

Tomoharu Sato

Graduate School of Science
Hiroshima University
Higashi-Hiroshima 739-8526
Japan

Contents

Summary

In this paper, we propose a model selection criterion in the generalized estimating equation method when the scale and correlation parameters are unknown. This model selection criterion is derived with reflecting the influence of the estimation of these unknown parameters. Furthermore, we evaluated the asymptotic bias of this criterion when the maximum cluster size goes to infinity as the sample size goes to infinity.

1 Introduction

Recently, in real data analysis, we treat data with correlation in many fields, for example medical science, economics and many other fields. Especially, data that are measured repeatedly over times from the same subjects, named longitudinal data, are widely used in those fields. In general, the data from the same subject have a correlation, whereas the data from different subjects are independent. Nelder and Wedderburn [13] proposed the generalized linear model (GLM), and after that Liang and Zeger [11] introduced an extension of the GLM, named generalized estimating equation (GEE). The GEE method is one of the methods to analyze the data with correlation. Defining features of the GEE method are that we use a working correlation matrix which can be chosen freely. We can get the consistent estimators of parameters whether the working correlation matrix is correct or not. It is worthy to say that we do not need a full specification of the joint distribution. In those reasons, the GEE method is widely used.

As with other statistical frameworks, the model selection problem in the GEE method is also important. In general, in the model selection, we measure the goodness of models by a certain risk. Then, by using some asymptotically unbiased estimators of the risk, we obtain a model selection criterion. For example, the most famous Akaike's information criterion (AIC) (Akaike, [1], [2]) was defined as an asymptotic unbiased estimator of the expected Kullback-Leibler divergence (Kullback and Leibler [10]). The AIC is calculated by AIC $=$ $-2\times$(the maximum log likelihood)$+2\times$(the number of parameters). Furthermore, the generalized information criterion (GIC) proposed by Nishii [14] and Rao [16] which is a generalization of the AIC is also applied to many fields. However, we cannot use model selection criteria based on the likelihood function such as the AIC or GIC for the GEE because we do not specify the joint distribution. Some model selection criteria like the AIC and GIC in the GEE method have been already proposed. For example, Pan [15] proposed the QIC (quasi-likelihood under the independence model criterion) based on the quasi-likelihood defined by Wedderburn [17]. Moreover, the GC_p (generalized version of Mallows's C_p) proposed by Cantoni *et al.* [3] is a generalization of Mallows's C_p (Mallows [12]). The correlation information criterion (CIC) proposed by Hin and Wang [6] and Gosho *et al.* [4] is a criterion for selecting the correlation structure. In the GEE method, we can get the smallest asymptotic variance of the GEE estimator by using the true correlation matrix as a working correlation matrix. It seems that the estimation accuracy can be improved by simultaneously selecting explanatory variables and a correlation structure, and the efficiency will be improved. Therefore, it is important to simultaneously select explanatory variables and a working correlation structure using one risk function. Unfortunately, the risk function of the QIC is based on the independent quasi likelihood, so the risk function does not reflect the correlation. Moreover, the CIC is focused on the working correlation structure modeling, on the other hand, the CIC is not focused on the variable selection. The Mallows's C_p is based on the prediction mean squared error so we can use these type of criteria in the GEE method. From this background, Inatsu and Imori [8] proposed the new model selection criterion, named PMSEG (the prediction mean squared error in the GEE) using the risk function based on the prediction mean squared error (PMSE) normalized by the covariance matrix. Inatsu and Imori [8] proposed this criterion when both the correlation parameters included in a working correlation matrix and the scale parameters are known, but the correlation and scale parameters are generally unknown in practice, so we consider to modify this criterion for the case that they are unknown.

In this paper, there are two purposes. One purpose is to propose a model selection criterion taking account of the correlation structure when both the correlation and scale parameters are unknown. In order to propose our model selection criterion, we evaluate the asymptotic bias of the estimator of a risk function and investigate the influences of the estimations of the correlation and scale parameters. We focus on the variable selection and the working correlation structure selection. The other purpose is to evaluate the asymptotic bias of the 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 propose an estimation method for parameters. In section 3, we perform the stochastic expansion of the GEE estimator and propose our model selection criterion when the scale and correlation parameters are unknown, in the case of the sample size goes to infinity and the cluster sizes are bounded. After that, we perform a numerical study. In section 4, we introduce asymptotic properties of the GEE estimator and the asymptotic bias of the PMSEG. After that, we perform a numerical study. In Appendix, we provide the calculation process of the asymptotic bias of the PMSEG, and the proofs of two theorems given in section 4.

2 Preliminaries

2.1 GEE estimator

Let y_{ij} be a scalar response variable from the ith subject at the jth observation time and $\boldsymbol{x}_{f,ij}$ be an l-dimensional nonstochastic vector consisting of possible explanatory variables, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Assume that the response variables from different subjects are independent and the response variables from the same subject are correlated. For each $i = 1, \ldots, n$, let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im})^{'}$ be the response vector from the ith subject and $\boldsymbol{X}_{f,i} = (\boldsymbol{x}_{f,i1}, \ldots, \boldsymbol{x}_{f,im})^{'}$ be the explanatory matrix from the ith subject. Moreover, let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im})^{'}$ be an $m \times p$ submatrix of the matrix $\boldsymbol{X}_{f,i}$. All the observed data for the ith subject are $(\boldsymbol{y}_i, \boldsymbol{X}_{f,i})$. Liang and Zeger [11] used the GLM as the marginal density of y_{ij},

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

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 a known function $u(\cdot)$ and ϕ is a scale parameter. Here, $\boldsymbol{\beta}$ is a p-dimensional unknown parameter and $\eta_{ij} = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}$ is called the linear predictor. In the present paper, we assume that the scale parameter ϕ is unknown, and let Θ be the *natural parameter space* (see, Xie and Yang [18]) of the exponential family of 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 these conditions, the expectation and variance of y_{ij} are given by

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

In the GLM framework, the expectation of y_{ij} is represented by the link function $g(\cdot)$ as $g(\mu_{ij}) = \eta_{ij} = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}$, where $g(t) = (\dot{a} \circ u)^{-1}(t)$. We call that the model with $\boldsymbol{x}_{f,ij}$ and $\boldsymbol{x}_{ij}$ as the full model and the candidate model, respectively. We assume that the true density function of y_{ij} can be written as (2.1), i.e., the true model is one of the candidate models. When the correlation and scale parameters are known, GEE proposed by Liang and Zeger [11] is as follows:

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

where $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), \ldots, \mu_{im}(\boldsymbol{\beta}))^{'}$, $\boldsymbol{D}_i(\boldsymbol{\beta}) = \partial\boldsymbol{\mu}_i(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^{'} = \boldsymbol{A}_i(\boldsymbol{\beta})\boldsymbol{\Delta}_i(\boldsymbol{\beta})\boldsymbol{X}_i$, $\boldsymbol{A}_i(\boldsymbol{\beta}) = \mathrm{diag}(\sigma_{i1}^2(\boldsymbol{\beta}), \ldots, \sigma_{im}^2(\boldsymbol{\beta}))$, $\boldsymbol{\Delta}_i(\boldsymbol{\beta}) = \mathrm{diag}(\partial\theta_{i1}/\partial\eta_{i1}, \ldots, \partial\theta_{im}/\partial\eta_{im})$ and $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}_w(\boldsymbol{\alpha})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\phi$. Here, $\boldsymbol{R}_w(\boldsymbol{\alpha})$ is called a *working correlation matrix* which can be chosen freely. Moreover, $\boldsymbol{R}_w(\boldsymbol{\alpha})$ includes nuisance parameter $\boldsymbol{\alpha}$. The nuisance parameter space is defined as follows:

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

We can use different working correlation matrices depending on each situation. Typical working correlation matrices are as follows:

(1) independence: $(\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = 0 \ (j \neq k)$,

(2) exchangeable: $(\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = \alpha \ (j \neq k)$,

(3) autoregressive: $(\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{kj} = \alpha^{j-k} \ (j > k)$,

(4) 1-dependence: $(\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{kj} = \begin{cases} \alpha \ (j = k+1) \\ 0 \ (j \neq k+1, \ j \neq k) \end{cases}$,

(5) unstructured: $(\boldsymbol{R}_w(\boldsymbol{\alpha}))_{jk} = (\boldsymbol{R}_w(\boldsymbol{\alpha}))_{kj} = \alpha_{jk} \ (j > k)$.

Note that the diagonal elements of $\boldsymbol{R}_w(\boldsymbol{\alpha})$ are ones, since it is a correlation matrix. The dimension of $\boldsymbol{\alpha}$ depends on the working correlation matrix. In many cases, $\boldsymbol{\alpha}$ is unknown. Although $\boldsymbol{\alpha}$ is the nuisance parameter, we must estimate $\boldsymbol{\alpha}$ in order to estimate $\boldsymbol{\beta}$. In practice, we estimate $\boldsymbol{\alpha}$ by real data. When both the correlation and scale parameters are unknown, we estimate $\boldsymbol{\alpha}$ by $\boldsymbol{\beta}$ and $\hat{\phi}$, where $\hat{\phi}$ is an estimator of ϕ. Denote $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}) = (\hat{\alpha}_1(\boldsymbol{\beta}, \hat{\phi}), \ldots, \hat{\alpha}_s(\boldsymbol{\beta}, \hat{\phi}))^{'}$, and assume that $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \phi_0) \xrightarrow{p} \boldsymbol{\alpha}_0 \in \mathcal{A}^{\circ}$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\alpha}}$ is the estimator of $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_0$ is the limiting value of $\hat{\boldsymbol{\alpha}}$, $\mathcal{A}^{\circ}$ is the interior of $\mathcal{A}$ and ϕ_0 is the limiting value of $\hat{\phi}$. 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. Assume that for $i = 1, \dots, n$, the true correlation matrix is the common matrix $\boldsymbol{R}_0$. If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, $\boldsymbol{V}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0) = \boldsymbol{\Sigma}_i(\boldsymbol{\beta}_0) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}_0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\phi_0 = \mathrm{Cov}[\boldsymbol{y}_i]$.

In this paper, we assume that $\boldsymbol{\alpha}$ and ϕ are unknown, so we replace $\boldsymbol{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ in (2.2) with $\boldsymbol{\Gamma}_i^{-1}(\boldsymbol{\beta})$ including the estimator of the correlation parameter $\hat{\boldsymbol{\alpha}}$, where $\boldsymbol{\Gamma}_i(\boldsymbol{\beta}) = \boldsymbol{V}_i(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})))$. Then, we obtain the following equation:

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

The solution of (2.3) denoted as $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}_0$. We call $\hat{\boldsymbol{\beta}}$ the GEE estimator.

2.2 Estimation method

The parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and ϕ are unknown, so we estimate them by the following iterative method:

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

Step 1 Set an initial value of $\boldsymbol{\alpha}$ denoted as $\hat{\boldsymbol{\alpha}}^{<0>}$

Step 2 Solve the GEE with $\hat{\boldsymbol{\alpha}}^{<k>}$, and the solution of the GEE is denoted as $\hat{\boldsymbol{\beta}}^{<k>} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}^{<k>})$.

Step 3 Estimate $\hat{\phi}^{<k+1>}$ by $\hat{\boldsymbol{\beta}}^{<k>}$.

Step 4 Estimate $\hat{\boldsymbol{\alpha}}^{<k+1>}$ by $\hat{\boldsymbol{\beta}}^{<k>}$ and $\hat{\phi}^{<k+1>}$.

Step 5 Iterate from step 2 to 4 until a certain condition about the convergence holds.

In the present paper, we estimate the scale parameter ϕ as follows:

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

and assume that $\hat{\phi} \xrightarrow{p} \phi_0$. In addition, the estimator $\hat{\boldsymbol{\alpha}}$ differs depending on each working correlation structure, and we give the following examples:

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

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

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

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

where $\hat{r}_{ij}(\hat{\boldsymbol{\beta}}) = y_{ij} - \mu_{ij}(\hat{\boldsymbol{\beta}})$. A moment estimation is popular. In fact, $\hat{\boldsymbol{\alpha}}$ is calculated by using the moment method in many statistical softwares. Empirically, by using the moment method, the above algorithm usually converges. However, the moment assumption does not necessarily imply that $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)$ is positive definite. Nevertheless, in many working assumptions (e.g., "Exchangeable" or "AR-1"), the positive definiteness of $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)$ mostly holds.

3 A C_p type criterion for model selection in the GEE method when both scale and correlation parameters are unknown

3.1 Stochastic expansion of GEE estimator

In this subsection, we perform the stochastic expansion of $\hat{\boldsymbol{\beta}}$. Furthermore, in order to evaluate the asymptotic properties of the GEE estimator, we assume the following conditions (Xie and Yang [18]):

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

C2. The true regression coefficient $\boldsymbol{\beta}_0$ is in an admissible set $\mathcal{B}$, and $\mathcal{B}$ is an open set of $\mathbb{R}^p$.

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

C4. The function $u(\eta_{ij})$ is four times continuously differentiable and $\dot{u}(\eta_{ij}) > 0$ in $g(\mathcal{M}^{\circ})$.

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

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

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

C6. It is established that $\liminf_{n\to\infty} \lambda_{\min}(\boldsymbol{H}_{n,0}/n) > 0$, where $\boldsymbol{H}_{n,0} = \sum_{i=1}^{n} \boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}\boldsymbol{D}_{i,0}$ and $\lambda_{\min}(\boldsymbol{A})$ is the minimum eigenvalue of a matrix $\boldsymbol{A}$.

C7. There exist a constant $c_0 > 0$ and n_0, such that for all $n \geq n_0$ and for any p-dimensional vector $\boldsymbol{\lambda}$ satisfying $||\boldsymbol{\lambda}|| = 1$, it holds that

$$P\left(-\boldsymbol{\lambda}^{'}\frac{\partial \boldsymbol{s}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{'}}\boldsymbol{\lambda} \geq nc_0\right) = 1 \ (\boldsymbol{\beta} \in N_0),$$

where N_0 is a neighborhood of $\boldsymbol{\beta}_0$.

C8. The GEE has a unique solution when n is large.

Conditions C1-C8 are modifications of the conditions proposed by Xie and Yang [18]. Conditions C1, C2 and C3 are necessary to consider the GLM framework. Conditions C4 and C5 are necessary to calculate the asymptotic bias of the estimator of the risk. In addition, Conditions C1, C6, C7 and C8 are necessary to have the strong consistency, asymptotic normality and uniqueness of the GEE estimator. Furthermore, in order to evaluate the asymptotic bias of the model selection criterion, we assume the following additional conditions.

C9. There exists a compact neighborhood of $\boldsymbol{\alpha}_0$, say $U_{\boldsymbol{\alpha}_0}$, and $\text{vec}\{\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha})\}$ is three times continuously differentiable in the interior of $U_{\boldsymbol{\alpha}_0}$.

C10. There exists a compact neighborhood of $\boldsymbol{\beta}_0$, say $U_{\boldsymbol{\beta}_0}$, and $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))$ is three times continuously differentiable in the interior of $U_{\boldsymbol{\beta}_0}$.

C11. For all $\boldsymbol{\beta} \in U_{\boldsymbol{\beta}_0}$, it is established that $\hat{\boldsymbol{\alpha}}^{(k)} = O_p(1)$ $(k = 1, 2, 3)$, where

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

C12. The estimator $\hat{\boldsymbol{\alpha}}_0 = \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \hat{\phi}(\boldsymbol{\beta}_0))$ satisfies $\sqrt{n}(\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0) = O_p(1)$, and there exists an $s \times p$ nonstochastic matrix $\boldsymbol{\mathcal{H}}$ such that $\hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{H}} = O_p(n^{-1/2})$.

C13. The following equations hold:

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

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

Note that for a matrix $\boldsymbol{W} = (w_{ij})$, the derivatives of $\boldsymbol{W}$ by $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and β_k are defined as follows:

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

We define $\boldsymbol{h}_{1,0}, \boldsymbol{j}_{1,0}, \boldsymbol{A}^*_{f,i,0}$ and $\boldsymbol{b}_{f,0}$ at the end of this section. Conditions C9, C10, C11, C12 and C13 are necessary for ignoring the influence of estimating the nuisance parameter $\boldsymbol{\alpha}$. Furthermore, by Condition C5, it is established that $\boldsymbol{H}_{n,0} = O(n)$. Furthermore, by Condition C12, $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \phi_0) \xrightarrow{p} \boldsymbol{\alpha}_0 \in \mathcal{A}^\circ$ holds.

Theorem 1. *Suppose that Conditions C1, C2, C3, C4, C7 and C8 hold. Furthermore, suppose that $\hat{\boldsymbol{\alpha}}$ is a moment estimator. If the matrix $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)$ is positive definite, Conditions C9, C10, C11, C12 and C13 hold.*

The moment estimator is defined by a continuous function of $\boldsymbol{\beta}$. By using properties of continuous functions, it is easy to show that Theorem 2 holds. Hence, we omit the proof of Theorem 2.

Based on the above conditions, to perform the stochastic expansion of $\hat{\boldsymbol{\beta}}$, we focus on the equation $\hat{\boldsymbol{s}}_n = \boldsymbol{s}_n(\hat{\boldsymbol{\beta}}) = \boldsymbol{0}_p$. By applying Taylor's expansion around $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$ to this equation, $\hat{\boldsymbol{s}}_n$ is expanded as follows:

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

where $\boldsymbol{\beta}^*$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$, $\boldsymbol{I}_p$ is the p-dimensional identity matrix and $\boldsymbol{s}_{n,0} = \boldsymbol{s}_n(\boldsymbol{\beta}_0)$. Here, $\boldsymbol{L}_1(\boldsymbol{\beta}^*)$, $\boldsymbol{\mathcal{D}}_{n,0}$, $\boldsymbol{\mathcal{D}}_{1,0}$ and $\boldsymbol{\mathcal{D}}_{2,0}$ are follows:

$$
\begin{aligned}
\boldsymbol{L}_1(\boldsymbol{\beta}^*) &= \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}, \boldsymbol{\mathcal{D}}_{n,0} = \sum_{i=1}^n \boldsymbol{D}_{i,0}' \boldsymbol{\Gamma}_{i,0}^{-1} \boldsymbol{D}_{i,0}, \\
\boldsymbol{\mathcal{D}}_{1,0} &= -\boldsymbol{\mathcal{D}}_{n,0}^{-1} \sum_{i=1}^n \boldsymbol{D}_{i,0}' \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{\Gamma}_i^{-1}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\{\boldsymbol{I}_p \otimes (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}, \\
\boldsymbol{\mathcal{D}}_{2,0} &= -\boldsymbol{\mathcal{D}}_{n,0}^{-1} \sum_{i=1}^n \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{D}_i'(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)[\boldsymbol{I}_p \otimes \{\boldsymbol{\Gamma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\},
\end{aligned}
$$

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

$$
\boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0) = \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) + \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) + O_p(n^{-1}).
$$

By Taylor's theorem, since $\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0 = O_p(n^{-1/2})$, it holds that

$$
||\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)|| \leq \left|\left|\frac{\partial}{\partial \boldsymbol{\alpha}} \otimes \boldsymbol{R}_w(\boldsymbol{\alpha})\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}\right|\right| \, ||\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0|| = O_p(n^{-1/2}),
$$

i.e., $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0) = O_p(n^{-1/2})$, where $\boldsymbol{\alpha}^*$ lies between $\boldsymbol{\alpha}_0$ and $\hat{\boldsymbol{\alpha}}$. Hence, it holds that

$$
\boldsymbol{\mathcal{D}}_{n,0} = \sum_{i=1}^n \boldsymbol{D}_{i,0}' \boldsymbol{\Gamma}_{i,0}^{-1} \boldsymbol{D}_{i,0}
$$

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

By this result and the fact that $\boldsymbol{s}_{n,0} = \boldsymbol{q}_{n,0} + O_p(1)$, $\hat{\boldsymbol{\beta}}$ is expanded as follows:

$$
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{H}_{n,0}^{-1} \boldsymbol{q}_{n,0} + O_p(n^{-1}) = \boldsymbol{b}_{1,0} + O_p(n^{-1}) \text{ (say)},
$$

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

$$
\begin{aligned}
&\left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right) - \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right] \\
&= O_p(n^{-1/2}),
\end{aligned}
$$

and above these results, (2.3) is expanded as follows:

$$
\begin{aligned}
&\boldsymbol{s}_{n,0} \\
&= \left[\boldsymbol{H}_{n,0} + \sum_{i=1}^{n} \boldsymbol{D}_{i,0}^{'} \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \{ \boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0) \} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{D}_{i,0} \right] \\
&\quad \cdot (\boldsymbol{I}_p + \boldsymbol{G}_{1,0} + \boldsymbol{G}_{2,0} + \boldsymbol{G}_{3,0})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad - \frac{1}{2} \{ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{'} \otimes \boldsymbol{I}_p \} \{ \boldsymbol{\mathcal{S}}_{1,0} + (\boldsymbol{L}_{1,0} - \boldsymbol{\mathcal{S}}_{1,0}) \} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad - \frac{1}{6} \{ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{'} \otimes \boldsymbol{I}_p \} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right) \right\} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{**}} \\
&\qquad \cdot \{ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \},
\end{aligned} \tag{3.4}
$$

where $\boldsymbol{\beta}^{**}$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$. Denote $\boldsymbol{\mathcal{S}}_{1,0} = \mathrm{E}[\boldsymbol{L}_{1,0}]$. Then, $\boldsymbol{\mathcal{S}}_{1,0} = O(n)$ and $\boldsymbol{L}_{1,0} - \boldsymbol{\mathcal{S}}_{1,0} = O_p(n^{1/2})$, where

$$
\boldsymbol{L}_{1,0} = \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.
$$

Note that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$ and

$$
\left\{ \frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right) \right\} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{**}} = O_p(n).
$$

Hence, the last term of (3.4) is $O_p(n^{-1/2})$. We define $\boldsymbol{\mathcal{C}}_{1i}$, $\boldsymbol{\mathcal{C}}_{2i}$, $\boldsymbol{\mathcal{C}}_{3i}$, $\boldsymbol{G}_{1,0}$, $\boldsymbol{G}_{2,0}$, $\boldsymbol{G}_{3,0}$, $\boldsymbol{h}_{1,0}$ and $\boldsymbol{j}_{1,0}$ as follows:

$$
\begin{aligned}
&\boldsymbol{\mathcal{C}}_{1i}(\boldsymbol{\beta}) = \boldsymbol{D}_i^{'}(\boldsymbol{\beta}) \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}) \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0), \ \boldsymbol{\mathcal{C}}_{2i}(\boldsymbol{\beta}) = \boldsymbol{D}_i^{'}(\boldsymbol{\beta}) \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}), \\
&\boldsymbol{\mathcal{C}}_{3i}(\boldsymbol{\beta}) = \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}), \\
&\boldsymbol{G}_{1,0} = -\boldsymbol{H}_{n,0}^{-1} \sum_{i=1}^{n} \mathcal{C}_{1i,0} \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right) \{ \boldsymbol{I}_p \otimes (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \}, \\
&\boldsymbol{G}_{2,0} = -\boldsymbol{H}_{n,0}^{-1} \sum_{i=1}^{n} \left(\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{\mathcal{C}}_{2i}(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right) [\boldsymbol{I}_p \otimes \{ \boldsymbol{\mathcal{C}}_{3i,0} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \}],
\end{aligned}
$$

$$\boldsymbol{G}_{3,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{2i,0}\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\hat{\phi}(\boldsymbol{\beta}))\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]$$
$$\cdot[\boldsymbol{I}_p\otimes\{\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\}],$$
$$\boldsymbol{h}_{1,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{1i,0}\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0)-\boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{\mathcal{C}}_{1i,0}'\boldsymbol{b}_{1,0},$$
$$\boldsymbol{j}_{1,0} = \boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{1i\cdot,0}\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0)-\boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{\mathcal{C}}_{3i,0}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0}),$$

where $\boldsymbol{\mathcal{C}}_{1i,0}=\boldsymbol{\mathcal{C}}_{1i}(\boldsymbol{\beta}_0)$, $\boldsymbol{\mathcal{C}}_{2i,0}=\boldsymbol{\mathcal{C}}_{2i}(\boldsymbol{\beta}_0)$ and $\boldsymbol{\mathcal{C}}_{3i,0}=\boldsymbol{\mathcal{C}}_{3i}(\boldsymbol{\beta}_0)$. Note that $\boldsymbol{G}_{1,0}=O_p(n^{-1/2})$, $\boldsymbol{G}_{2,0}=O_p(n^{-1/2})$, $\boldsymbol{G}_{3,0}=O_p(n^{-1/2})$, $\boldsymbol{h}_{1,0}=O_p(n^{-1})$ and $\boldsymbol{j}_{1,0}=O_p(n^{-1})$. By using the above equations, $\hat{\boldsymbol{\beta}}$ is expanded as follows:

$$\begin{aligned}
&\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\\
&=(\boldsymbol{I}_p-\boldsymbol{G}_{1,0}+\boldsymbol{G}_{2,0}+\boldsymbol{G}_{3,0})\Big[\boldsymbol{I}_p\\
&\quad-\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0)-\boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\Big]\\
&\quad\cdot\boldsymbol{H}_{n,0}^{-1}\left[\boldsymbol{s}_{n,0}+\frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\otimes\boldsymbol{I}_p\}\{\boldsymbol{\mathcal{S}}_{1,0}+(\boldsymbol{L}_{1,0}-\boldsymbol{\mathcal{S}}_{1,0})\}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\right]\\
&=\boldsymbol{b}_{1,0}+\boldsymbol{b}_{2,0}+O_p(n^{-3/2}),
\end{aligned}\tag{3.5}$$

where $\boldsymbol{b}_{1,0}=\boldsymbol{H}_{n,0}^{-1}\boldsymbol{q}_{n,0}=O_p(n^{-1/2})$ and $\boldsymbol{b}_{2,0}=\boldsymbol{H}_{n,0}^{-1}(\boldsymbol{b}_{1,0}'\otimes\boldsymbol{I}_p)\boldsymbol{\mathcal{S}}_{1,0}\boldsymbol{b}_{1,0}/2-\boldsymbol{G}_{1,0}\boldsymbol{b}_{1,0}-\boldsymbol{G}_{2,0}\boldsymbol{b}_{1,0}-\boldsymbol{G}_{3,0}\boldsymbol{b}_{1,0}+\boldsymbol{h}_{1,0}+\boldsymbol{j}_{1,0}=O_p(n^{-1})$.

3.2 Main result

In this section, we propose a model selection criterion. We measure the goodness of fit of the model by the risk function based on the PMSE normalized by the covariance matrix. The risk function is as follows:

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

where $\hat{\boldsymbol{\mu}}_i=\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})$ 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 $\boldsymbol{y}_i$. If $\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}_0$, Risk_P has the minimum value zero, i.e., PMSE has the minimum value mn. We consider that the model which has minimum PMSE is the optimum model, and we want to select this model. Since the PMSE is typically unknown, we must estimate it.

We define $\hat{\boldsymbol{R}}(\boldsymbol{\beta})$, $\boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$ and $\boldsymbol{\mathcal{L}}^*(\boldsymbol{\beta})$ as follows:

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

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

Then, we estimate the PMSE by $\mathcal{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$, where $\hat{\boldsymbol{\beta}}_f$ is the GEE estimator from the full model, namely, we obtain $\hat{\boldsymbol{\beta}}_f$ as the solution of the following equation:

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

where $\boldsymbol{D}_i(\boldsymbol{\beta}_f) = \boldsymbol{A}_i(\boldsymbol{\beta}_f)\boldsymbol{\Delta}(\boldsymbol{\beta}_f)\boldsymbol{X}_{f,i}, \boldsymbol{V}_i(\boldsymbol{\beta}_f, \boldsymbol{\alpha}_f) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_f)\bar{\boldsymbol{R}}_i(\boldsymbol{\alpha}_f)\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_f)$ and $\bar{\boldsymbol{R}}_i(\boldsymbol{\alpha}_f)$ is a positive definite working correlation matrix which can be chosen freely. Also, $\bar{\boldsymbol{R}}_i(\boldsymbol{\alpha}_f)$ is the same for all the candidate models. For simplicity, we denote $\mathcal{L}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_2) = \mathcal{L}(\boldsymbol{\beta}_2)$ and $\mathcal{L}^*(\boldsymbol{\beta}_0) = \mathcal{L}^*$.

We construct a model selection criterion by correcting the asymptotic bias of the estimator $\mathcal{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ as an estimator of PMSE like as the Mallows's C_p. The bias of $\mathcal{L}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ is given by

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

We evaluate Bias1, Bias2, Bias3 and Bias4 separately.

At first, Bias3 is as follows:

$$\begin{aligned}
&\text{Bias3} \\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}(\hat{\boldsymbol{\beta}}_f)\}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] \\
&= mn - \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}(\hat{\boldsymbol{\beta}}_f)(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right].
\end{aligned}$$

Hence, Bias3 depends on only the full model, so we can ignore Bias3 for model selection.

Second, Bias1 is expanded as follows:

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

$$
\begin{aligned}
&- \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\right] - 2\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]\\
&- \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]\\
=\ & 2\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_{i,0})\right]. \qquad (3.6)
\end{aligned}
$$

For expanding Bias1, we must expand $\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0}$. 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:

$$
\begin{aligned}
&\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0}\\
&= \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&\quad + \frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^{'}\otimes \boldsymbol{I}_m\}\left(\frac{\partial}{\partial \boldsymbol{\beta}}\otimes\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&\quad + \frac{1}{6}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^{'}\otimes \boldsymbol{I}_m\}\left\{\frac{\partial}{\partial \boldsymbol{\beta}'}\otimes\left(\frac{\partial}{\partial \boldsymbol{\beta}}\otimes\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right)\right\}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{***}}\\
&\qquad \cdot\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\otimes(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}\\
&= \boldsymbol{D}_{i,0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^{'}\otimes \boldsymbol{I}_m\}\boldsymbol{D}_{i,0}^{(1)}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + O_p(n^{-3/2}), \qquad (3.7)
\end{aligned}
$$

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

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

By substituting (3.5) for (3.7), we can expand $\hat{\boldsymbol{\mu}}_i$ as follows:

$$
\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0} + \left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}^{'}\otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\} + O_p(n^{-3/2}). \qquad (3.8)
$$

By using (3.6) and (3.8), we get the following expansion:

$$
\begin{aligned}
\frac{1}{2}\mathrm{Bias1} &= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_{i,0})\right]\\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&\quad + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}^{'}\otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\}\right]\\
&\quad + \mathrm{E}_y[O_p(n^{-1/2})]. \qquad (3.9)
\end{aligned}
$$

Since the data from different two subjects are independent, we can get $\mathrm{E}[(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})] = 0$ $(i \neq j)$. The first term of (3.9) is calculated as follows:

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

$$
\begin{aligned}
&= \mathrm{E}_y\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{j,0}^{'}\boldsymbol{V}_{j,0}^{-1}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})\right]\\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\right]\\
&= \mathrm{E}_y\left[\mathrm{tr}\left\{\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\right\}\right]\\
&= \mathrm{E}_y\left[\mathrm{tr}\left\{\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right\}\right]\\
&= \mathrm{tr}\left\{\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}\mathrm{E}\left[(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\right]\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right\}\\
&= \mathrm{tr}\left(\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right)\\
&= \mathrm{tr}\,(\boldsymbol{I}_p)\\
&= p. \qquad (3.10)
\end{aligned}
$$

Also, since for all i, j, k (not $i = j = k$),

$$
\mathrm{E}\left[(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\otimes(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})^{'}(\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0})\right]=\boldsymbol{0}_m,
$$

the second term of (3.9) is calculated as follows:

$$
\begin{aligned}
&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0}+\frac{1}{2}(\boldsymbol{b}_{1,0}^{'}\otimes\boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\}\right]\\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2i,0}+\frac{1}{2}(\boldsymbol{b}_{1i,0}^{'}\otimes\boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1i,0}\right\}\right]\\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}(\boldsymbol{b}_{2i,0}-\boldsymbol{h}_{1,0}-\boldsymbol{j}_{1,0})\right.\right.\\
&\qquad\left.\left.+\frac{1}{2}(\boldsymbol{b}_{1i,0}^{'}\otimes\boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1i,0}\right\}\right]\\
&\quad+\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\{\boldsymbol{D}_{i,0}(\boldsymbol{h}_{1,0}+\boldsymbol{j}_{1,0})\}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{b}_{1i,0} &= \boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0}),\\
\boldsymbol{b}_{2i,0} &= \boldsymbol{H}_{n,0}^{-1}(\boldsymbol{b}_{1i,0}^{'}\otimes\boldsymbol{I}_p)\boldsymbol{\mathcal{S}}_{1,0}\boldsymbol{b}_{1i,0}/2-\boldsymbol{G}_{1i,0}\boldsymbol{b}_{1i,0}-\boldsymbol{G}_{2i,0}\boldsymbol{b}_{1i,0}-\boldsymbol{G}_{3i,0}\boldsymbol{b}_{1i,0}\\
&\quad+\boldsymbol{h}_{1,0}+\boldsymbol{j}_{1,0},\\
\boldsymbol{G}_{1i,0} &= -\boldsymbol{H}_{n,0}^{-1}\mathcal{C}_{1i,0}\left(\frac{\partial}{\partial\boldsymbol{\beta}^{'}}\otimes\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\{\boldsymbol{I}_p\otimes(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\},\\
\boldsymbol{G}_{2i,0} &= -\boldsymbol{H}_{n,0}^{-1}\left(\frac{\partial}{\partial\boldsymbol{\beta}^{'}}\otimes\boldsymbol{\mathcal{C}}_{2i}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)[\boldsymbol{I}_p\otimes\{\boldsymbol{\mathcal{C}}_{3i,0}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\}],
\end{aligned}
$$

$$\boldsymbol{G}_{3i,0} = -\boldsymbol{H}_{n,0}^{-1}\boldsymbol{\mathcal{C}}_{2i,0}\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\beta}'} \otimes \boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})))\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right]$$
$$\cdot [\boldsymbol{I}_p \otimes \{\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\}].$$

Under Condition C13, we have

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

so the second term of (3.9) is calculated as follows:

$$\mathrm{E}_y\left[\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\}\right] = O(n^{-1}). \tag{3.11}$$

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 [5]). Therefore, by substituting (3.10) and (3.11) for (3.9), we obtain

$$\mathrm{Bias1} = 2p + O(n^{-1}).$$

Similarly, we obtain

$$\mathrm{Bias2} + \mathrm{Bias4} = O(n^{-1}). \tag{3.12}$$

The derivation of (3.12) is shown in Appendix.

From the above, the bias is expanded as follows:

$$\mathrm{Bias} = 2p + \mathrm{Bias3} + O(n^{-1}).$$

Note that Bias3 does not depend on all the candidate models so we propose the model selection criterion as

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

This criterion is the same as the criterion proposed by Inatsu and Imori [8].

3.3 Numerical study

In this section, we perform a numerical study and discuss the result. There are two aims to perform this simulation. One is to compare the frequencies of selecting models in the case of we use the correct correlation structure as a working correlation and in the case of we use the wrong correlation structure as a working correlation. The other is to compare the prediction errors in the same situation with estimating the correlation and scale parameters. The QIC proposed by Pan [15] and modified QIC proposed by Imori [7] are representative model selection criteria in the GEE method, and Inatsu and Imori [8] confirmed a usefulness of the PMSEG through comparisons with the QIC and modified QIC. Similar results of the comparisons can be expected in the framework of this paper. Therefore, the comparisons with the QIC and modified QIC are not performed in this numerical study.

In this simulation, we got data from the gamma distributions which have the scale parameter included in the exponential family. Then, we supposed that there are two groups (e.g., male and female). Furthermore, we supposed that the distribution of observations from one group is different from the other one. To create data distributed according to the gamma distributions with correlation, we used the copula method. We set $n = 50, 100, 150, 200$ and $m = 3$. For each $i = 1, 2, \ldots, n$, we constructed the 3×8 explanatory matrix $\boldsymbol{X}_{f,i} = (\boldsymbol{x}_{f,i1}, \boldsymbol{x}_{f,i2}, \boldsymbol{x}_{f,i3})^{'} = (\boldsymbol{X}_{1i}, \boldsymbol{X}_{2i})$. Here, for each $i = 1, \ldots, (n/2)$,

$$\boldsymbol{X}_{1i} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 \end{pmatrix},$$

and for each $i = (n/2) + 1, \ldots, n$,

$$\boldsymbol{X}_{1i} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Furthermore, all the elements of $\boldsymbol{X}_{2i}$ $(i = 1, \ldots, n)$ are independent and identically distributed according to the uniform distribution on the interval $[-1, 1]$. Let the true correlation structure be the exchangeable structure, i.e., $\boldsymbol{R}_0 = (1-\alpha)\boldsymbol{I}_m + \alpha \boldsymbol{1}_m \boldsymbol{1}_m^{'}$, where α is the correlation parameter. Furthermore, in this simulation, we prepare two situations, as follows:

Case 1: $\alpha = 0.3$, $\boldsymbol{\beta}_0 = (0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0, 0)^{'}$,

Case 2: $\alpha = 0.8$, $\boldsymbol{\beta}_0 = (0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0, 0)^{'}$.

The explanatory matrix for the ith subject in the kth model $(k = 1, 2, \ldots, 8)$ consists of the first k columns of $\boldsymbol{X}_{f,i}$. We simulate 10,000 realizations of $\boldsymbol{y} = (y_{11}, \ldots, y_{13}, \ldots, y_{n1}, \ldots, y_{n3})^{'}$, 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 consider the situation we use the exchangeable structure as a working correlation structure. The frequencies of selecting models and the prediction errors in Case 1 and Case 2 are given in Table 9 and Table 10, respectively. The values in parentheses are the standard errors of the prediction error of each situation. In the both situations, the frequency of selecting the 6th model tends to be large as n is large. Furthermore, the frequencies of selecting the 1-5th models tend to 0.

Table 1: Frequencies of selecting models (%) and prediction errors when $\alpha = 0.3$ using exchangeable working correlation matrix

n	1	2	3	4	5	6	7	8	Prediction Error
50	3.4	1.4	3.8	0.7	12.6	53.1	14.2	10.8	6.573 (0.03)
100	0.1	0.0	0.2	0.1	3.3	71.8	13.1	11.4	6.512 (0.03)
150	0.0	0.0	0.0	0.0	0.3	75.4	13.6	10.7	6.641 (0.03)
200	0.0	0.0	0.0	0.0	0.0	75.5	15.6	8.9	6.494 (0.03)

Next, we consider the situation we use a wrong correlation structure as a working correlation structure. We use the autoregressive structure as one of

Table 2: Frequencies of selecting models (%) and prediction errors when $\alpha = 0.8$ using exchangeable working correlation matrix

n	1	2	3	4	5	6	7	8	Prediction Error
50	0.6	0.5	0.3	0.2	0.9	67.4	17.6	12.5	7.089(0.04)
100	0.0	0.0	0.0	0.0	0.0	71.7	17.4	10.9	6.533(0.03)
150	0.0	0.0	0.0	0.0	0.0	73.7	15.5	10.8	6.455(0.03)
200	0.0	0.0	0.0	0.0	0.0	75.4	14.9	9.7	6.688(0.03)

such structures. The frequencies of selecting models and the prediction errors in Case 1 and in Case 2 are given in Table 11 and Table 12, respectively. In the case of using the different correlation structure as well as using the true correlation structure, the frequency of selecting the 6th model tends to large as n is large, and the frequencies of selecting the 1-5 models tend to 0. In Case 1, the prediction error in Table 9 is not much different from that in Table 11 for each n, on the other hand, in Case 2, the prediction error in Table 10 is different from that in Table 12 for each n. From this, it is considered that the larger the true correlation value, the greater the influence of the working correlation structure on the prediction error.

Table 3: Frequencies of selecting models (%) and prediction errors when $\alpha = 0.3$ using autoregressive working correlation matrix

n	1	2	3	4	5	6	7	8	Prediction Error
50	8.2	0.9	4.2	0.7	6.7	58.0	11.2	10.1	6.660 (0.03)
100	0.2	0.0	0.6	0.0	2.1	73.8	14.9	8.4	6.810 (0.04)
150	0.0	0.0	0.0	0.0	0.5	74.8	13.4	11.3	6.767 (0.03)
200	0.0	0.0	0.0	0.0	0.0	78.2	12.8	9.0	6.990 (0.04)

Table 4: Frequencies of selecting models (%) and prediction errors when $\alpha = 0.8$ using autoregressive working correlation matrix

n	1	2	3	4	5	6	7	8	Prediction Error
50	1.2	0.6	0.4	0.2	2.9	65.5	17.0	12.2	7.268 (0.04)
100	0.1	0.1	0.0	0.0	0.0	74.2	16.8	8.8	7.158 (0.04)
150	0.0	0.0	0.0	0.0	0.0	78.2	13.3	8.5	7.017 (0.04)
200	0.0	0.0	0.0	0.0	0.0	79.6	12.9	7.5	7.402 (0.04)

Next, we consider the situation we use the independence structure as a working correlation structure, namely, we assume the GLM. The frequencies of selecting models and the prediction errors in Case 1 and in Case 2 are given in Table 5 and table 6, respectively. In this situation, the frequency of selecting the 6th model is the largest of three situations, but the prediction error is the largest.

Finally, we consider selecting the explanatory variables and the working correlation structure simultaneously. We use three working correlation structures, i.e., exchangeable (Ex.), autoregressive (AR) and independence (Ind.). Then, the number of models is $8 \times 3 = 24$. The frequencies of selecting models and the prediction errors in Case 1 and in Case 2 are given in Table 7 and Table 8,

Table 5: Frequencies of selecting models (%) and prediction errors when $\alpha = 0.3$ using independence working correlation matrix

n	1	2	3	4	5	6	7	8	Prediction Error
50	8.7	1.9	3.2	0.9	9.2	52.6	13.8	9.7	6.829 (0.04)
100	0.3	0.0	1.5	0.0	3.2	69.4	15.3	10.3	7.135 (0.04)
150	0.0	0.0	0.0	0.0	0.3	75.8	14.5	9.4	7.069 (0.04)
200	0.0	0.0	0.0	0.0	0.0	78.6	13.1	8.3	7.199 (0.04)

Table 6: Frequencies of selecting models (%) and prediction errors when $\alpha = 0.8$ using independence working correlation matrix

n	1	2	3	4	5	6	7	8	Prediction Error
50	2.2	2.0	1.0	0.3	5.4	69.3	12.1	7.7	11.600(0.04)
100	0.1	0.0	0.0	0.0	0.7	83.6	11.1	4.5	11.276(0.04)
150	0.0	0.1	0.0	0.0	0.2	84.0	10.6	5.1	11.833(0.04)
200	0.0	0.0	0.0	0.0	0.0	87.8	7.6	4.6	11.585(0.04)

respectively. By comparing Table 7 with Table 9 and Table 8 with Table 10, it shows that the prediction errors in Table 7 and Table 8 are significantly smaller than the prediction errors in the case of we use the true correlation structure as a working correlation for each n. Similarly, by comparing Table 7 with Table 11 and Table 8 with Table 12, it shows that the prediction errors in Table 7 and Table 8 are significantly smaller than the prediction errors in the case of we use the wrong correlation structure as a working correlation. Table 7 and Table 8 indicate that by selecting both variables and a working correlation, we may be able to improve the prediction accuracy. Note that if we use a specific correlation structure, the prediction error might be large.

Table 7: Frequencies of selecting models (%) and prediction errors when $\alpha = 0.3$ using tree types of correlation matrix

n	W-Cor.	1	2	3	4	5	6	7	8	Prediction Error
	Ex.	3.2	1.1	1.5	0.6	4.7	24.2	8.0	6.3	
50	AR	6.2	0.7	2.2	0.4	2.2	15.0	3.5	2.6	6.043 (0.03)
	Ind.	0.7	0.1	0.7	0.4	2.2	9.8	1.6	2.1	
	Ex.	0.0	0.0	0.2	0.2	0.8	41.2	8.5	6.0	
100	AR	0.1	0.1	0.1	0.0	0.5	17.1	3.4	2.9	6.147 (0.03)
	Ind.	0.0	0.0	0.2	0.0	0.8	13.8	2.7	1.4	
	Ex.	0.0	0.0	0.0	0.0	0.4	41.7	8.8	7.2	
150	AR	0.0	0.0	0.0	0.0	0.0	19.9	4.0	2.3	6.104 (0.03)
	Ind.	0.0	0.0	0.0	0.0	0.1	12.3	2.5	0.8	
	Ex.	0.0	0.0	0.0	0.0	0.1	41.8	8.1	6.2	
200	AR	0.0	0.0	0.0	0.0	0.1	21.5	3.6	2.2	6.028 (0.03)
	Ind.	0.0	0.0	0.0	0.0	0.0	13.7	1.4	1.3	

Table 8: Frequencies of selecting models (%) and prediction errors when $\alpha = 0.8$ using tree types of correlation matrix

n	W-Cor.	1	2	3	4	5	6	7	8	Prediction Error
	Ex.	0.5	0.4	0.1	0.2	0.7	40.7	10.9	7.9	
50	AR	0.7	0.0	0.0	0.0	0.5	19.2	5.8	6.1	6.098 (0.03)
	Ind.	0.0	0.1	0.1	0.0	0.2	4.8	0.6	0.5	
	Ex.	0.0	0.2	0.0	0.0	0.0	48.3	9.7	7.3	
100	AR	0.1	0.0	0.0	0.0	0.0	21.3	4.9	3.4	6.136 (0.03)
	Ind.	0.0	0.0	0.0	0.0	0.0	4.2	0.3	0.3	
	Ex.	0.0	0.0	0.0	0.0	0.0	47.7	10.1	8.4	
150	AR	0.0	0.0	0.0	0.0	0.0	19.2	4.6	2.3	5.949 (0.03)
	Ind.	0.0	0.0	0.0	0.0	0.0	6.4	0.8	0.5	
	Ex.	0.0	0.0	0.0	0.0	0.0	49.0	9.0	7.5	
200	AR	0.0	0.0	0.0	0.0	0.0	22.7	4.3	2.6	5.844 (0.03)
	Ind.	0.0	0.0	0.0	0.0	0.0	4.3	0.3	0.3	

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

4.1 Model selection in the GEE

In this section, since n and m go to infinity, we change the notations. The GEE is as follows:

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

and

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

Let

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

We consider the following regularity conditions (see, e.g., Xie and Yang [18], Inatsu and Sato [9]):

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

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

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

C14. It holds that $\tau_{nm}\lambda_{\max}(\boldsymbol{H}_{nm,0}^{-1}) \to 0$, where $\tau_{nm} = \lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{R}_0)$.

C15. It holds that $\pi_{nm}^2\tau_{nm}m\gamma_{nm}^{(0)} \to 0$, where

$$\begin{aligned}
\pi_{nm} &= \frac{\lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))}{\lambda_{\min}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0))},\\
\gamma_{nm}^{(0)} &= \max_{1\le i\le n}\max_{1\le j\le m} \boldsymbol{x}_{ij}'\boldsymbol{H}_{nm,0}^{-1}\boldsymbol{x}_{ij}.
\end{aligned}$$

C16. It holds that $(c_{nm})^{1+\delta}(\tilde{\lambda}_{nm}m)^{2+\delta}\gamma_{nm}^{(0)} \to 0$ for some $\delta > 0$, where

$$\begin{aligned}
c_{nm} &= \lambda_{\max}(\boldsymbol{M}_{nm,0}^{-1}\boldsymbol{H}_{nm,0}),\\
\tilde{\lambda}_{nm} &= \lambda_{\max}(\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)).
\end{aligned}$$

Conditions C5*, C6* and C14-C16 are the modifications of the conditions proposed by Xie and Yang [18]. Here, to evaluate the asymptotic bias of the PMSEG, we present the following lemma:

Lemma 1. *Suppose Conditions C1 - C4, C5*, C6* and C14-C16 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 distribution.

(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 Corollary 1 of Xie and Yang [18], so we omit the proof. Here, $\boldsymbol{I}_p$ is the p-dimensional identity matrix.

4.2 Asymptotic bias of PMSEG

In this section, we evaluate the asymptotic bias of the PMSEG. We consider the following assumptions (see, Inatsu and Sato [9]):

C12*. The estimator $\hat{\boldsymbol{\alpha}}_0 = \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \hat{\phi}(\boldsymbol{\beta}_0))$ satisfies $(\sqrt{n}/m)(\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0) = O_p(1)$, and there exists an $s \times p$ nonstochastic matrix $\boldsymbol{\mathcal{H}}$ such that $\hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{H}} = O_p(m/\sqrt{n})$.

C13*. The following equations hold:

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

C17. $\liminf_{n\to\infty}\lambda_{\min}(\boldsymbol{B}_{nm,0}/nm) > 0$, where

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

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 2 in Appendix. By using the moment estimator of the correlation parameters and the scale parameter, Conditions C9, C10, C11, C12* and C13* are fulfilled. Condition C17 is necessary to prove following Lemma 2:

Lemma 2. *Suppose Conditions C1 - C4, C5*, C6*, C8, C9-C11, C12*, C13* and C14-C17 hold. Even if the working correlation matrix is misspecified, we have*

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

Proof. Suppose Conditions C1 - C4, C5*, C6*, C8, C9-C11, C12*, C13* and C14-C17 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:

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

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

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

□

Then, we evaluate the asymptotic bias of the PMSEG.

Theorem 2. *Suppose Conditions C1 - C4, C5*, C6*, C8, C9-C11, C12*, C13* and C14-C17 hold. The variance of the asymptotic bias of the 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 2 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 Conditions C1 - C4, C5*, C6*, C8, C9-C11, C12*, C13* and C14-C17 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 the PMSEG when we use true correlation matrix.

Theorem 3. *Suppose Conditions C1 - C4, C5*, C6*, C8, C9-C11, C12*, C13* and C14-C17 hold. The variance of the asymptotic bias of the 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 3 in Appendix.

4.3 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 2, 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 9. In Table 9, 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 10, (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 11. Table 11 indicates that in the case of the working correlation structure is misspecified, the frequency of selecting the 6th model tends to large as m^4/n is small, and the frequencies of selecting of the 1-5 models tend to 0. In Table 12, 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 12 indicates that the rate of the asymptotic bias of the PMSEG m^4/n is overestimate, so we may not need so many samples.

Table 9: 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 10: (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

5 Conclusions and discussions

In this paper, we proposed a C_p type criterion for model selection in the GEE method when the both scale and correlation parameters are unknown. Furthermore, we discussed about the asymptotic bias of the PMSEG when the sample size and the cluster sizes are large.

According to Section 3, when n goes to infinity and m is bounded, the asymptotic bias of the PMSEG excluding the bias independent of a candidate model goes to 0. Furthermore, according to Section 4, when m goes to infinity as n goes to infinity, the asymptotic bias of the PMSEG excluding the bias independent of a candidate model goes to 0 if $m^4/n \to 0$. Furthermore, if we use the true correlation structure as a working correlation and m^2/n, the asymptotic bias of the PMSEG excluding the bias independent of a candidate model goes to 0.

The GEE method is widely used in many studies. The GEE method is packaged in statistical software "R" and "SAS", so this method is useful. Moreover, a moment estimation is used in these software. We can select the explanatory variables and the working correlation structure simultaneously by using the PMSEG and prove the prediction accuracy. Hence, we thought that this criterion

Table 11: 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 12: (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

is useful.

Acknowledgements

I would like to express the deepest gratitude to Professor Hirofumi Wakaki and Professor Hirokazu Yanagihara of Hiroshima University for their helpful comments and suggestions. I would also like to express my gratitude to Dr. Yu Inatsu of RIKEN Center for Advanced Intelligence Project for his helpful advise throughout my research.

A Appendix

A.1 Derivation of (3.12) in section 3.2

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

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

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

Hence, Bias2 + Bias4 is

$$
\begin{aligned}
&\text{Bias2} + \text{Bias4} \\
&= \mathrm{E}_y\Bigg[2\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\left\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\} \\
&\qquad \cdot (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\Bigg] \qquad \text{(A1.1)} \\
&\quad + \mathrm{E}_y\Bigg[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)^{'}\left\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\} \\
&\qquad \cdot (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\Bigg]. \qquad \text{(A1.2)}
\end{aligned}
$$

In order to evaluate these expectations, 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 $\hat{\phi}(\hat{\boldsymbol{\beta}}_f)$. We expand $\hat{\boldsymbol{\beta}}_f$ as with the expansion of $\hat{\boldsymbol{\beta}}$ in section 2. The expansion is as follows:

$$\hat{\boldsymbol{\beta}}_f - \boldsymbol{\beta}_{f,0} = \boldsymbol{H}_{f,n,0}^{-1}\boldsymbol{s}_{f,n}(\boldsymbol{\beta}_{f,0}) + O_p(n^{-1}) = \boldsymbol{b}_{f,0} + O_p(n^{-1}),$$

where $\boldsymbol{\beta}_{f,0}$ is the true value of $\boldsymbol{\beta}_f$. Here, $\boldsymbol{H}_{f,n,0}$ is

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

where $\boldsymbol{D}_{f,i} = \boldsymbol{A}_i(\boldsymbol{\beta}_f)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_f)\boldsymbol{X}_{f,i}$, $\boldsymbol{D}_{f,i,0} = \boldsymbol{A}_{i,0}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_{f,i}$ and $\bar{\boldsymbol{R}}_i$ is the working correlation matrix of the full model. In addition, as with the expansion of $\hat{\boldsymbol{\mu}}_i$ in section 3, we expand $\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_f)$ as follows:

$$\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_f) - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{f,i,0}\boldsymbol{b}_{f,0} + O_p(n^{-1}).$$

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 expansion of $\boldsymbol{a}_{f,i}(\hat{\boldsymbol{\beta}}_f)$ around $\hat{\boldsymbol{\beta}}_f = \boldsymbol{\beta}_{f,0}$ as follows:

$$\boldsymbol{a}_{f,i}(\hat{\boldsymbol{\beta}}_f) = \boldsymbol{a}_{f,i}(\boldsymbol{\beta}_{f,0}) + \boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0} + O_p(n^{-1}),$$

where

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

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

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

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

$$\hat{\phi}(\hat{\boldsymbol{\beta}}_f) = \phi_0 + O_p(n^{-1/2}).$$

Furthermore, $\hat{\boldsymbol{R}}(\hat{\boldsymbol{\beta}}_f)$ is expanded as follows:

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

$$
\begin{aligned}
&+ (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{D}_{f,i,0}\boldsymbol{b}_{f,0})'\}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}(\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1}) \\
&+ O_p(n^{-1}). \qquad \text{(A1.3)}
\end{aligned}
$$

By Lindberg central limit theorem, the first term of (A1.3) is $O_p(n^{-1})$. Then, we get the following expansion with using above expansions:

$$
\begin{aligned}
&\boldsymbol{R}_0^{-1/2}\hat{\boldsymbol{R}}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1/2} \\
&= \boldsymbol{I}_m - \boldsymbol{I}_m + \frac{1}{n}\boldsymbol{R}_0^{-1/2}\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}\boldsymbol{R}_0^{-1/2}\phi_0^{-1} \\
&\quad + \frac{1}{n}\boldsymbol{R}_0^{-1/2}\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}\boldsymbol{R}_0^{-1/2}\phi_0^{-1} \\
&\quad + \frac{1}{n}\boldsymbol{R}_0^{-1/2}\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})\boldsymbol{R}_0^{-1/2}\phi_0^{-1} \\
&\quad + \frac{1}{n}\boldsymbol{R}_0^{-1/2}\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}\boldsymbol{R}_0^{-1/2}(\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1}) \\
&\quad + O_p(n^{-1}) \\
&= \boldsymbol{I}_m - \boldsymbol{R}_0^{-1/2}\Big\{\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} \\
&\quad - \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} \\
&\quad - \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} \\
&\quad - \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})\Big\}\boldsymbol{R}_0^{-1/2} \\
&\quad + O_p(n^{-1}).
\end{aligned}
$$

Therefore, the inverse matrix of $\boldsymbol{R}_0^{-1/2}\hat{\boldsymbol{R}}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1/2}$ can be expanded as follows:

$$
\begin{aligned}
&\boldsymbol{R}_0^{1/2}\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{1/2} \\
&= \boldsymbol{I}_m + \boldsymbol{R}_0^{-1/2}\Big\{\boldsymbol{R}_0 - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{n}\sum_{i=1}^{n} \mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&- \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})' \mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
&- \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2}(\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1})\Big\}\boldsymbol{R}_0^{-1/2} \\
&+ O_p(n^{-1}). \qquad \text{(A1.4)}
\end{aligned}
$$

Therefore, $\hat{\boldsymbol{R}}^{-1}$ is expanded as follows:

$$
\begin{aligned}
&\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f) \\
&= \boldsymbol{R}_0^{-1} + \boldsymbol{R}_0^{-1}\Big\{\boldsymbol{R}_0 - \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&- \frac{1}{n}\sum_{i=1}^{n} \mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&- \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})' \mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
&- \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2}(\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1})\Big\}\boldsymbol{R}_0^{-1} \\
&+ O_p(n^{-1}). \qquad \text{(A1.5)}
\end{aligned}
$$

Note that the second term of (A1.5) is $O_p(n^{-1/2})$. 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{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f) \\
&= \boldsymbol{\Sigma}_{i,0}^{-1} - \{\boldsymbol{A}_{i,0}^{-1/2} + \mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})\} \\
&\cdot\Bigg[\boldsymbol{R}_0^{-1} + \boldsymbol{R}_0^{-1}\Big\{\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})\Big\}\boldsymbol{R}_0^{-1}\Bigg] \\
&\cdot\{\boldsymbol{A}_{i,0}^{-1/2} + \mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})\}\{\phi_0^{-1} + (\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1})\} \\
&+ O_p(n^{-1}) \\
&= -\mathrm{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}\mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
&- \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\Big\{\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}
\end{aligned}
$$

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

Note that $\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f) = O_p(n^{-1/2})$ and $\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0} = O_p(n^{-1/2})$. Then, (A1.2) is calculated as follows:

$$
\begin{aligned}
&\mathrm{E}_y\Bigg[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\Big\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\Big\} \\
&\qquad \cdot (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\Bigg] \\
&= O(n^{-1}).
\end{aligned}
$$

In addition, we calculate (A1.1) as follows:

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

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

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

$$\cdot\,\{\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1}\}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg]$$

$$+ \mathrm{E}_y\Bigg[2\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\{\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1}\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg]$$

$$+ \mathrm{E}_y\Bigg[2\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg] + O(n^{-1}). \tag{A1.6}$$

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 calculate the first term of (A1.6) as follows:

$$\begin{aligned}
&\mathrm{E}_y\Bigg[2\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\Bigg\{\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\\
&\qquad + \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\Bigg\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg]\\
&= \mathrm{E}_y\Bigg[2\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\Bigg\{\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,i,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\\
&\qquad + \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,i,0})\Bigg\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg]\\
&= O(n^{-1}),
\end{aligned} \tag{A1.7}$$

where $\boldsymbol{b}_{f,i,0}=\boldsymbol{H}^{-1}_{f,n,0}\boldsymbol{D}_i'(\boldsymbol{\beta}_{f,0})\boldsymbol{V}_i^{-1}(\boldsymbol{\beta}_{f,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_{f,0}))$. Similarly, because of $\mathrm{E}_y[(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'(\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0})]=0$ (unless $i=k$), the second term of (A1.6) is calculated as follows:

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

$$= -2p + O(n^{-1}). \tag{A1.8}$$

Here, we define notations of summation as follows:

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

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:

$$i = j = l \text{ or } i = j \neq k = l \text{ or } i = l \neq k = j \text{ or } j = l \neq k = i.$$

Thus, the third term of (A1.6) is calculated as follows:

$$\begin{aligned}
&-\mathrm{E}_y\Bigg[\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})' \\
&\qquad \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}\Bigg] \\
&= -\mathrm{E}_y\Bigg[\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})' \\
&\qquad \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}\Bigg] \\
&= -\mathrm{E}_y\Bigg[\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})' \\
&\qquad \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}\Bigg] \\
&\quad -\mathrm{E}_y\Bigg[\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})' \\
&\qquad \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}\Bigg] \\
&\quad -\mathrm{E}_y\Bigg[\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})' \\
&\qquad \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}\Bigg] \\
&\quad + O(n^{-1}) \\
&= O(n^{-1}).
\end{aligned} \tag{A1.9}$$

Similarly, the forth term of (A1.6) is calculated as follows:

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

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

$$
= O(n^{-1}). \tag{A1.10}
$$

The fifth term of (A1.6) is calculated as follows:

$$
\begin{aligned}
&-\mathrm{E}_y\Bigg[\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} \\
&\qquad \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}\Bigg] \\
&= -\mathrm{E}_y\Bigg[\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} \\
&\qquad \cdot\frac{\partial\hat{\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}\Bigg] \\
&\quad -\mathrm{E}_y\Bigg[\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} \\
&\qquad \cdot\frac{\partial\hat{\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}\Bigg] \\
&\quad -\mathrm{E}_y\Bigg[\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} \\
&\qquad \cdot\frac{\partial\hat{\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}\Bigg] \\
&\quad -\mathrm{E}_y\Bigg[\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} \\
&\qquad \cdot\frac{\partial\hat{\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}\Bigg] \\
&= O(n^{-1})
\end{aligned} \tag{A1.11}
$$

The sixth term of (A1.6) is calculated as follows:

$$
\begin{aligned}
&\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\{\hat{\phi}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1}\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right] \\
&= \mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\frac{\partial\hat{\phi}(\boldsymbol{\beta}_f)}{\partial\boldsymbol{\beta}_f}\bigg|_{\boldsymbol{\beta}_f=\boldsymbol{\beta}_{f,0}}\boldsymbol{b}_{f,i,0}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right] \\
&= O(n^{-1})
\end{aligned} \tag{A1.12}
$$

Furthermore, the seventh term of (A1.6) is calculated as with (3.10).

$$
\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right] = 2p. \tag{A1.13}
$$

By (A1.7)-(A1.13), (A1.1) is calculated as follows:

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

Thus, we have $\mathrm{Bias2}+\mathrm{Bias4}=O(n^{-1})$.

A.2 Proof of Theorem 2 and Theorem 3

We prove the Theorem 2 and Theorem 3, simultaneously. In the proof of this section, the notations used in section 3 and A.1 are used again, but with some changes, they are redefined.

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

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

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

$$\begin{aligned}
&\boldsymbol{L}_1(\boldsymbol{\beta}^*)=\left(\frac{\partial}{\partial\boldsymbol{\beta}}\otimes\frac{\partial\boldsymbol{s}_{nm}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*},\boldsymbol{\mathcal{D}}_{nm,0}=\sum_{i=1}^{n}\boldsymbol{D}_{i,0}'\boldsymbol{\Gamma}_{i,0}^{-1}\boldsymbol{D}_{i,0},\\
&\boldsymbol{\mathcal{D}}_{1,0}=-\boldsymbol{\mathcal{D}}_{nm,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}'\left(\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{\Gamma}_i^{-1}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\{\boldsymbol{I}_p\otimes(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\},\\
&\boldsymbol{\mathcal{D}}_{2,0}=-\boldsymbol{\mathcal{D}}_{nm,0}^{-1}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{D}_i'(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)[\boldsymbol{I}_p\otimes\{\boldsymbol{\Gamma}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\}],
\end{aligned}$$

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:

$$\boldsymbol{R}_w^{-1}(\hat{\boldsymbol{\alpha}}_0)=\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)+\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)\{\boldsymbol{R}_w(\boldsymbol{\alpha}_0)-\boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0)+O_p(m^2/n). \quad \text{(A2.1)}$$

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

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

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 (A2.1) is $O_p(1/n)$. Hence, it holds that

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

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

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

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

$$\begin{aligned}
&\boldsymbol{s}_{nm,0} \\
&= -\sum_{i=1}^{n} \boldsymbol{D}_{i,0}^{'} \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \boldsymbol{A}_{i,0}^{-1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \\
&\quad + \boldsymbol{H}_{nm,0}(\boldsymbol{I}_p + \boldsymbol{H}_{nm,0}^{-1} \boldsymbol{B}_{1,0} + \boldsymbol{H}_{nm,0}^{-1} \boldsymbol{B}_{2,0} + \boldsymbol{H}_{nm,0}^{-1} \boldsymbol{B}_{3,0})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad + \sum_{i=1}^{n} \boldsymbol{D}_{i,0}^{'} \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{D}_{i,0} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad - \frac{1}{2} \{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{'} \otimes \boldsymbol{I}_p\} \{\boldsymbol{\mathcal{S}}_{1,0} + (\boldsymbol{L}_1(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{S}}_{1,0})\} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad - \frac{1}{6} \{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{'} \otimes \boldsymbol{I}_p\} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_{nm}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right) \right\} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{**}} \\
&\quad \cdot \{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}, \qquad \text{(A2.2)}
\end{aligned}$$

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

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

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

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

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

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 (A2.2), we have

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

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

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

where

$$\begin{aligned}
&\boldsymbol{j}_{1,0} \\
&= \boldsymbol{H}_{nm,0}^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{i,0}^{'} \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \boldsymbol{A}_{i,0}^{-1/2} \\
&\quad \cdot (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}), \\
&\boldsymbol{h}_{1,0} \\
&= \boldsymbol{H}_{nm,0}^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{i,0}^{'} \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \{\boldsymbol{R}_w(\boldsymbol{\alpha}_0) - \boldsymbol{R}_w(\hat{\boldsymbol{\alpha}}_0)\} \boldsymbol{R}_w^{-1}(\boldsymbol{\alpha}_0) \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0}.
\end{aligned}$$

Denote

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

Note that above $\boldsymbol{b}_{2,0}$ is different from $\boldsymbol{b}_{2,0}$ in section 3, and the sum of above $\boldsymbol{b}_{2,0}$ and $\boldsymbol{b}_{3,0}$ is equal to $\boldsymbol{b}_{2,0}$ in section 3. Hence, we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{b}_{1,0} + \boldsymbol{b}_{2,0} + \boldsymbol{b}_{3,0} + O_p(m^4/n^{3/2}) \qquad \text{(A2.3)}$$

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 the PMSEG as follows:

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

We evaluate Bias1, Bias2, Bias3 and Bias4 separately.

Bias1 is expanded as follows:

$$\begin{aligned}
&\text{Bias1} \\
&= \text{E}_y\left[\text{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)\right] - \sum_{i=0}^{n}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)\right] \\
&= \text{E}_y\Bigg[\text{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
&\qquad - \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\Bigg] \\
&= \text{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0})\right] + \text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
&\quad - \text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] - 2\text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
&\quad - \text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
&= 2\text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0})\right]. \qquad \text{(A2.4)}
\end{aligned}$$

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:

$$\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0}$$

$$
\begin{aligned}
&= \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\
&\quad + \frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^{'} \otimes \boldsymbol{I}_m\}\left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \\
&\quad + \frac{1}{6}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^{'} \otimes \boldsymbol{I}_m\}\left\{\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right)\right\}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{***}} \\
&\qquad \cdot \{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \otimes (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\} \\
&= \boldsymbol{D}_{i,0}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \frac{1}{2}\{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)^{'} \otimes \boldsymbol{I}_m\}\boldsymbol{D}_{i,0}^{(1)}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + O_p(m^{7/2}/n^{3/2}) \quad \text{(A2.5)}
\end{aligned}
$$

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

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

By substituting (A2.3) for (A2.5), we can expand $\hat{\boldsymbol{\mu}}_i$ as follows:

$$
\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{i,0}(\boldsymbol{b}_{1,0}+\boldsymbol{b}_{2,0}+\boldsymbol{b}_{3,0}) + \frac{1}{2}(\boldsymbol{b}_{1,0}^{'} \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0} + O_p(m^{7/2}/n^{3/2}). \tag{A2.6}
$$

By using (A2.4) and (A2.6), we get the following expansion:

$$
\begin{aligned}
\frac{1}{2}\text{Bias1} &= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_{i,0})\right] \\
&= \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right] \\
&\quad + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0}\right] \\
&\quad + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{3,0}\right] \\
&\quad + \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{b}_{1,0}^{'} \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right] \\
&\quad + \mathrm{E}_y[O_p(n^{-1/2}m^{7/2})].
\end{aligned} \tag{A2.7}
$$

Same as Inatsu and Sato [9], the first term of (A2.7) is calculated as follows:

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

Since the data from different two subjects are independent, we calculate the second term of (A2.7) as follows:

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

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

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 (A2.7) is $O(m^2/n)$. Similarly, the orders of the third and the forth term of (A2.7) are evaluated as follows:

$$
\begin{aligned}
&\mathrm{E}_y\left[\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{\Sigma}_{i,0}^{-1} \boldsymbol{D}_{i,0} \boldsymbol{b}_{3,0}\right] = O(m^{7/2}/n) \\
&\mathrm{E}_y\left[\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{\Sigma}_{i,0}^{-1} (\boldsymbol{b}_{1,0}' \otimes \boldsymbol{I}_m) \boldsymbol{D}_{i,0}^{(1)} \boldsymbol{b}_{1,0}\right] = O(m^{5/2}/n).
\end{aligned}
$$

Furthermore, if $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, the order of the third term of (A2.7) is $O(m^{3/2}/n)$ and the order of the forth term of (A2.7) 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 [5]). 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:

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

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

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

Hence, Bias2 + Bias4 is

$$\begin{aligned}
&\text{Bias2} + \text{Bias4} \\
&= \mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\left\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\} \cdot (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \qquad \text{(A2.8)} \\
&+ \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)^{'}\left\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\} \cdot (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right]. \qquad \text{(A2.9)}
\end{aligned}$$

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 $\hat{\phi}(\hat{\boldsymbol{\beta}}_f)$. 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 limiting 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

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

where $\boldsymbol{D}_{f,i,0} = \boldsymbol{A}_{i,0}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_{f,i}$ and $\bar{\boldsymbol{R}}_i^{-1}(\boldsymbol{\alpha}_f)$ 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

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

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 [9], $\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)$ is expanded as follows:

$$\begin{aligned}
&\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\\
&= \boldsymbol{R}_0^{-1} + \boldsymbol{R}_0^{-1}\Big\{\boldsymbol{R}_0 - \frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\\
&\quad - \frac{1}{n}\sum_{i=1}^n \mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\\
&\quad - \frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\phi_0^{-1}\\
&\quad - \frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}(\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1})\Big\}\boldsymbol{R}_0^{-1}\\
&\quad + O_p(m^3/n). \qquad \text{(A2.10)}
\end{aligned}$$

Note that the second term of (A2.10) is $O_p(m/\sqrt{n})$. Then, we have

$$\begin{aligned}
&\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\\
&= \boldsymbol{\Sigma}_{i,0}^{-1} - \{\boldsymbol{A}_{i,0}^{-1/2} + \mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\}\\
&\quad\cdot\Bigg[\boldsymbol{R}_0^{-1} + \boldsymbol{R}_0^{-1}\Big\{\boldsymbol{R}_0 - \frac{1}{n}\sum_{i=1}^n \boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}
\end{aligned}$$

$$
\begin{aligned}
&- \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})\Big\}\boldsymbol{R}_0^{-1}\Bigg] \\
&\cdot \{\boldsymbol{A}_{i,0}^{-1/2} + \mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\}\{\phi_0^{-1} + (\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1})\} \\
&+ O_p(m^3/n) \\
= &-\mathrm{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}\mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\phi_0^{-1} \\
&- \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\Big\{\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})\Big\}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&- \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_i^{-1/2}\{\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1}\} \\
&+ O_p(m^3/n),
\end{aligned}
$$

where $\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}(\hat{\boldsymbol{\beta}}_f) = O_p(m/\sqrt{n})$ 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, (A2.9) is calculated as follows:

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

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

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

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 (A2.9) is $O(m/n)$. In addition, we calculate (A2.8).

$$
\begin{aligned}
&\mathrm{E}_y\Bigg[2\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \Bigg\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\Bigg\} \\
&\qquad \cdot (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathrm{E}_y\Bigg[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\Big\{\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&\qquad + \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\phi_0^{-1}\Big\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg] \\
&\quad - \mathrm{E}_y\Bigg[\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})' \\
&\qquad \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}\Bigg] \\
&\quad - \mathrm{E}_y\Bigg[\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})' \\
&\qquad \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}\Bigg] \\
&\quad - \mathrm{E}_y\Bigg[\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})' \\
&\qquad \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}\Bigg] \\
&\quad - \mathrm{E}_y\Bigg[\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} \\
&\qquad \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}\Bigg] \\
&\quad + \mathrm{E}_y\Bigg[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{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)-\phi_0^{-1}\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg] \\
&\quad + \mathrm{E}_y\Bigg[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}\Bigg] + O(m^4/n). \quad \text{(A2.11)}
\end{aligned}
$$

Note that $\mathrm{E}[(\boldsymbol{y}_{i,}-\boldsymbol{\mu}_{i,0})\otimes(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'(\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0})]=\mathbf{0}_m$ (not $i=j=k$), so we can expand the first term of (A2.11) as follows:

$$
\begin{aligned}
&\mathrm{E}_y\Bigg[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\Big\{\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&\qquad + \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})\phi_0^{-1}\Big\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg] \\
&= \mathrm{E}_y\Bigg[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\Big\{\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,i,0})\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} \\
&\qquad + \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,i,0})\phi_0^{-1}\Big\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg] \\
&= O(m^4/n), \qquad \text{(A2.12)}
\end{aligned}
$$

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 (A2.11) 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 (A2.11) is expanded as follows:

$$
\begin{aligned}
&- \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})^{'}\right.\\
&\qquad\left.\cdot\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&= -\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1,i\neq j}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})^{'}\right.\\
&\qquad\left.\cdot\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]\\
&\quad + O(m^4/n)\\
&= -\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right] + O(m^4/n)\\
&= -2p + O(m^4/n), \qquad\qquad (\mathrm{A2.13})
\end{aligned}
$$

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

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

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:

$$
i = j = l \text{ or } i = j \neq k = l \text{ or } i = l \neq k = j \text{ or } j = l \neq k = i.
$$

Thus, the third term of (A2.11) is calculated as follows:

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

$$
\begin{aligned}
&= -\mathrm{E}_y\Bigg[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\mathrm{diag}(\boldsymbol{A}^*_{f,i,0}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})' \\
&\qquad \cdot \boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg] \\
&\quad -\mathrm{E}_y\Bigg[\sum_{i\neq j}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\mathrm{diag}(\boldsymbol{A}^*_{f,j,0}\boldsymbol{b}_{f,i,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})' \\
&\qquad \cdot \boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1j,0}\Bigg] \\
&\quad -\mathrm{E}_y\Bigg[\sum_{i\neq j}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\mathrm{diag}(\boldsymbol{A}^*_{f,j,0}\boldsymbol{b}_{f,j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})' \\
&\qquad \cdot \boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\Bigg] \\
&\quad + O(m^4/n) \\
&= O(m^4/n). \qquad \text{(A2.14)}
\end{aligned}
$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, the order of the third term of (A2.11) is $O(m^2/n)$. Similarly, the forth term of (A2.11) is expanded as follows:

$$
\begin{aligned}
&-\mathrm{E}_y\Bigg[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})' \\
&\qquad \cdot \mathrm{diag}(\boldsymbol{A}^*_{f,j,0}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\Bigg] \\
&= O(m^4/n). \qquad \text{(A2.15)}
\end{aligned}
$$

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0)=\boldsymbol{R}_0$, the order of the forth term of (A2.11) is $O(m^2/n)$. The fifth term of (A2.11) is calculated as follows:

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

$$
\begin{aligned}
&- \mathrm{E}_y \left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \frac{2}{n} \sum_{j=1}^{n} \boldsymbol{A}_{j,0}^{-1/2} (\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})' \boldsymbol{A}_{j,0}^{-1/2} \right. \\
&\qquad \left. \cdot \left. \frac{\partial \hat{\phi}(\boldsymbol{\beta}_f)}{\partial \boldsymbol{\beta}_f} \right|_{\boldsymbol{\beta}_f = \boldsymbol{\beta}_{f,0}} \boldsymbol{b}_{f,i,0} \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \phi_0^{-1} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1j,0} \right] \\
&- \mathrm{E}_y \left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \frac{2}{n} \sum_{j=1}^{n} \boldsymbol{A}_{j,0}^{-1/2} (\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})' \boldsymbol{A}_{j,0}^{-1/2} \right. \\
&\qquad \left. \cdot \left. \frac{\partial \hat{\phi}(\boldsymbol{\beta}_f)}{\partial \boldsymbol{\beta}_f} \right|_{\boldsymbol{\beta}_f = \boldsymbol{\beta}_{f,0}} \boldsymbol{b}_{f,j,0} \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \phi_0^{-1} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1i,0} \right] \\
&= O(m^4/n) \qquad \text{(A2.16)}
\end{aligned}
$$

The sixth term of (A2.11) is calculated as follows:

$$
\begin{aligned}
&\mathrm{E}_y \left[2 \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \{ \hat{\phi}(\hat{\boldsymbol{\beta}}_f) - \phi_0^{-1} \} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \right] \\
&= \mathrm{E}_y \left[2 \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_{i,0}^{-1/2} \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \left. \frac{\partial \hat{\phi}(\boldsymbol{\beta}_f)}{\partial \boldsymbol{\beta}_f} \right|_{\boldsymbol{\beta}_f = \boldsymbol{\beta}_{f,0}} \boldsymbol{b}_{f,i,0} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1i,0} \right] \\
&= O(m^3/n) \qquad \text{(A2.17)}
\end{aligned}
$$

Furthermore, the seventh term of (A2.11) is calculated as follows:

$$
\mathrm{E}_y \left[2 \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \boldsymbol{A}_i^{-1/2} \boldsymbol{R}_0^{-1} \boldsymbol{A}_{i,0}^{-1/2} \phi_0^{-1} \boldsymbol{D}_{i,0} \boldsymbol{b}_{1,0} \right] = 2p. \qquad \text{(A2.18)}
$$

By (A2.12)-(A2.18), (A2.8) is calculated as follows:

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

If $\boldsymbol{R}_w(\boldsymbol{\alpha}_0) = \boldsymbol{R}_0$, the order of (A2.8) 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 the 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 the PMSEG goes to 0 with the rate of m^2/n or faster. □

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] E. Cantoni, J. M. Flemming and E. Ronchetti, Variable selection for marginal longitudinal generalized linear models, Biometrics, **61** (2005), 309-317.

[4] M. Gosho, C. Hamada and I. Yoshimura, Modifications of QIC and CIC for selecting a Working Correlation Structure in the Generalized Estimating Equation Method, Japanese Journal of Biometrics, **32** (2011), 1-12.

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

[6] L. Y. Hin and Y. G. Wang, Working-correlation-structure identification in generalized estimating equations, Statistics in Medicine, **28** (2009), 642-658.

[7] S. Imori, Model selection criterion based on the multivariate quasi-likelihood for generalized estimating equations, Scand. J. Stat., **42** (2015), 1214-1224.

[8] Y. Inatsu and S. Imori, Model selection criterion based on the prediction mean squared error in generalized estimating equations, Hiroshima Mathematical Journal, **48(3)** (2018), 307-334.

[9] 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.

[10] S. Kullback and R. Libler, On information and sufficiency, Ann. Math. Statist., **22** (1951), 79-86.

[11] K. Y. Liang, and S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika, **73** (1986), 13-22.

[12] C. L. Mallows, Some comments on C_p, Technometrics, **15** (1973), 661-675.

[13] J. A. Nelder and R. W. M. Wedderburn, Generalized linear models, J. R. Statist. Soc. ser. A, **135** (1972), 370-384.

[14] R. Nishii, Asymptotic Properties of Criteria for Selecting of Variables in Multiple Regression, Ann. Statist., **12** (1984), 758-765.

[15] W. Pan, Akaike's Information Criterion in Generalized Estimating Equations, Biometrics, **57** (2001), 120-125.

[16] C. R. Rao and Y. Wu, A strongly Consistent Procedure for Model Selection in a Regression Problem, Biometrika, **76** (1989), 369-374.

[17] R. W. M. Wedderburn, Quasi-likelihood functions, generalized linear models, and Gauss-Newton method, Biometrika, **61** (1974), 439-447.

[18] M. Xie and Y. Yang, Asymptotics for generalized estimating equations with large cluster sizes. Ann. Statist., **31** (2003), 310-347.