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

Yu Inatsu and Tomoharu Sato
Department of Mathematics, Graduate school of science, Hiroshima university

## 1 INTRODUCTION

Recently, in real data analysis, we consider the data with correlation for many fields, for example medical science, economics and many other fields. Especially, the data what is measured repeatedly over times from same subjects, named longitudinal data, is widely used in those fields. In general, the data from same subject have correlation, on the other hand, the data from different subjects are independent.. Liang and Zeger (1986) introduce an extension of generalized linear model (Nelder and Wedderburn, 1972), named generalized estimating equation (GEE). GEE method is one of the methods to analyze the data with correlation. Defining features of the GEE method are that we can use working correlation matrix one can choose freely. We can get good estimation of parameters if working correlation matrix is correct or not. It is important that we don't need a full specification of a joint distribution. In those reason, GEE method is widely used in many fields.

"Model selection" is also important problem, so we apply model selection to the GEE. In general, in model selection, we measure the goodness of fit by risk function, and choose the model with smallest risk function. Then, by using the asymptotically unbiased estimator of risk function, we consider the model selection criterion. For example, expected Kullback-Leibler information (Kullback and Leibler, 1951), and most famous Akaike's information criterion (AIC) (Akaike, 1973, 1974) are used. The AIC is calculated by $\mathrm{AIC} = -2 \times (\mathrm{maximumloglikelihood}) + 2 \times (\mathrm{thenumberofparameters})$. Furthermore, the GIC what is expansion of the AIC proposed by Nishii (1984) and Rao (1988) is also applied for many fields.

However, we can't use the model selection criterion based likelihood as AIC or GIC because of we don't specify joint distribution. Some model selection criteria like AIC and GIC in the GEE method have been already proposed. For example, Pan (2001) proposed the QIC based on the quasi-likelihood (defined by Wedderburn, 1974). Furthermore, the $GC_p$ proposed by Cantoni *et al.* (2005) is the generally extension of Mallow's $C_p$ (Mallows, 1973). The CIC proposed by Hin and Wang (2009) and Gosho *et al.* (2011) is criterion what select the correlation structure. Unfortunately, the above criteria are derived without consider the correlation structure so we regard to these criteria don't reflect the correlation.

From this background, in Inatsu and Imori (2013) proposed a new model selection criterion 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 (2013) proposed this criterion when both correlation and scale parameters are known, but correlation and scale parameters are generally unknown so we consider this criterion when both correlation and scale parameters are unknown.

In this paper, the main topic is to propose the model selection criterion considered correlation structure when both correlation and scale parameters are unknown. In order to propose the new model selection criterion, we evaluate the asymptotic bias of the estimator of risk function and consider the influence of estimation correlation parameter and scale parameter. We focus on the "variable selection" which selecting the optimum combination of variables.

The present paper organized as follows: In section 2, we introduce the GEE framework and propose the estimation method for parameters. After that, we perform the stochastic expansion of the GEE estimator. In section 3, we define the estimation of risk function, and evaluate the asymptotic bias by calculate the bias, and propose the new model selection criterion. In section 4, we perform numerical study. In section 5, we conclude our discussion. In appendix, we provide the calculation process for the bias.

# 2 STOCHASTIC EXPANSION OF THE GEE ESTIMATOR

## 2.1 GEE estimator

Let $y_{ij}$ be a scalar response variable, and $\boldsymbol{x}_{f,ij}$ be a l-dimensional nonstochastic vector consists of possible explanatory variables from the $i$th subject at the $j$th occasion, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Assume that the response variables from defferent subjects are independent and response variables from same subject are correlated. For each $i = 1, \ldots, n$, let response variable vector from $i$th subject be $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im})^{'}$ and explanatory variable matrix from $i$th subject be $\boldsymbol{X}_{f,i} = (\boldsymbol{x}_{f,i1}, \ldots, \boldsymbol{x}_{f,im})^{'}$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im})^{'}$ be a $m \times p$ submatrix of the matrix $\boldsymbol{X}_{*,i}$. Liang and Zeger (1986) used the generalized linear model (GLM) to the model of 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, $a(\cdot), b(\cdot)$ are known functions, $\theta_{ij}$ is an unknown location parameter and $\phi$ is a scale parameter. In the GLM framework, the location parameter $\theta_{ij} = u(\eta_{ij}) = \theta_{ij}(\boldsymbol{\beta})$, where $u(\cdot)$ is known function, and $\eta_{ij} = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is $p$-dementional unkown parameter. In the present paper, we assume that scale parameter $\phi$ is unkown parameter, and we also assume that $\boldsymbol{\Theta}$ is the *natural parameter space* (see, Xie and Yang, 2003) of the exponential family of distributions presented in (2.1), and the interior of $\Theta$ is denoted as $\Theta^0$. $\Theta$ is convex and in $\Theta^0$, all derivatives of $a(\cdot)$ and all moments of $y_{ij}$ exist. Under these conditions, mean and variance of $y_{ij}$ are given by

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

In the GLM framework, the expectation of $y_{ij}$ modeled by link function as $g(\mu ij) = \eta_{ij} = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}$. Then link function $g(t) = (\dot{a} \circ u)^{-1}(t)$ and linear predictor $\eta_{ij} = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}$.If $u(s) = s$, we say that $g(t) = \dot{a}^{-1}(t)$ is natural link function. We call that the model with $\boldsymbol{x}_{f,ij}$ or $\boldsymbol{x}_{ij}$ as full model or candidate model, respectively. The true density function of $y_{ij}$ can be written as (2.1), i.e. true model is one of candidate models.

GEE proposed by Liang and Zeger (1986) is as follows:

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

where $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), \ldots, \mu_{im}(\boldsymbol{\beta}))^{'}$, $\boldsymbol{D}_i(\boldsymbol{\beta}) = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta} = \boldsymbol{A}_i(\boldsymbol{\beta})\boldsymbol{\Delta}_i(\boldsymbol{\beta})\boldsymbol{X}_i$, $\boldsymbol{A}_i(\boldsymbol{\beta}) = \mathrm{diag}(\sigma_{i1}^2(\boldsymbol{\beta}), \ldots, \sigma_{im}^2(\boldsymbol{\beta}))$, $\boldsymbol{\Delta}_i(\boldsymbol{\beta}) = \mathrm{diag}(\partial\theta_{i1}/\partial\eta_{i1}, \ldots, \partial\theta_{im}/\partial\eta_{im})$ and $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\phi$. $\boldsymbol{R}(\boldsymbol{\alpha})$ is working correlation matrix one can chose freely. 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 true correlation matrix. Assume that for $i = 1, \ldots, n$, true correlation matrix is common $\boldsymbol{R}_0$. Working correlation $\boldsymbol{R}(\boldsymbol{\alpha})$ include nuisance parameter $\boldsymbol{\alpha}$. Nuisance parameter space is as follows:

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

We can use different working correlation depending on the situation. For example:
[1]independence: $(\boldsymbol{R})_{jk} = 0, (j \neq k)$.
[2]exchangeable: $(\boldsymbol{R})_{jk} = \alpha, (j \neq k)$.
[3]autoregressive: $(\boldsymbol{R})_{jk} = (\boldsymbol{R})_{kj} = \alpha^{j-k}, (j > k)$.
[4]1-dependence: $(\boldsymbol{R})_{jk} = (\boldsymbol{R})_{kj} = \alpha, (j = k + 1)$.
[5]unstructured: $(\boldsymbol{R})_{jk} = (\boldsymbol{R})_{kj} = \alpha_{jk}, (j > k)$.

Denote $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\phi(\boldsymbol{\beta})$. If $\boldsymbol{R}(\boldsymbol{\alpha}) = \boldsymbol{R}_0$, $\boldsymbol{V}_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}) = \boldsymbol{\Sigma}_i(\boldsymbol{\beta}_0) = \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}_0\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\phi_0 = \mathrm{Cov}[\boldsymbol{y}_i]$. Note that $\boldsymbol{\beta}_0$ is true parameter of $\boldsymbol{\beta}$. Dimension of $\boldsymbol{\alpha}$ depends on choose of working correlation. In many case, correlation parameter $\boldsymbol{\alpha}$ is unknown. Although $\boldsymbol{\alpha}$ is nuisance parameter, we must estimate $\boldsymbol{\alpha}$ so as to estimate $\boldsymbol{\beta}$. In practice, we estimate $\boldsymbol{\alpha}$ by real data. When both correlation and scale parameter are unknown, we estimate $\hat{\boldsymbol{\alpha}}$ by $\boldsymbol{\beta}$ and $\hat{\phi}$. 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 $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0,\phi_0) \xrightarrow{a.s.} \boldsymbol{\alpha}_0 \in \mathcal{A}^\circ$, where $\mathcal{A}^\circ$ is interior of $\mathcal{A}$. In present paper, we estimate scale parameter $\phi$ is as follows:

$$\hat{\phi} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(y_{ij}-\hat{\mu}_{ij})^2}{\ddot{a}(\hat{\theta}_{ij})}$$

and assume that $\hat{\phi} \xrightarrow{p} \phi_0$.

In this paper, we assume that $\boldsymbol{\alpha}$ and $\phi$ are unknown, so we consider the following equation:

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

where $\boldsymbol{\Gamma}(\boldsymbol{\beta}) = \boldsymbol{V}_i(\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\hat{\phi}))$. The solution of equation (2.2) denoted $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}_0$. We call $\hat{\boldsymbol{\beta}}$ the GEE estimator.

## 2.2 Estimation method

The true parameters $\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0$ and $\phi_0$ are unknown so we estimate parameters by following iterative method:

---

**Algorithm** (Estimation method for parameters)

---

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

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

Step 3 Estimate $\hat{\phi}^{<k+1>}$ by $\boldsymbol{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}^{<k>})$.

Step 4 Estimate $\hat{\boldsymbol{\alpha}}^{<k+1>} = \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}^{<k>}, \hat{\phi}^{<k+1>})$. We propose the estimation of $\hat{\boldsymbol{\alpha}}^{<k+1>}$ later.

Step 5 Iterate processes 2 to 4 until converge the value of parameters.

---

When one use the moment estimator for $\boldsymbol{\alpha}$, the fact that the condition C9 to C13 are fulfilled (Inatsu, 2013). In addition, the estimator $\hat{\boldsymbol{\alpha}}$ differ depending on the working correlation structure, and we give examples.

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

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

$$1-\text{dependence}: \hat{\alpha} = \frac{1}{n-1}\sum_{i=1}^{n-1}\hat{\alpha}_i, \hat{\alpha}_i = \frac{1}{n}\sum_{i=1}^{n}\hat{r}_{ij}\hat{r}_{i,j+1}/\hat{\phi}.$$

$$\text{Unstructured}: \hat{\alpha}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\hat{r}_{ij}\hat{r}_{ik}/\hat{\phi}.$$

## 2.3 STOCHASTIC EXPANSION OF GEE ESTIMATOR

In this section, in order to propose the new variable selection criterion, we perform the stochastic expansion of $\hat{\boldsymbol{\beta}}$. For simplicity, we omit $(\boldsymbol{\beta})$ from functions of $\boldsymbol{\beta}$, for example $\mu_{ij}(\boldsymbol{\beta})$=$\mu_{ij}$. In addition, In order to distinguish the function of $\boldsymbol{\beta}$ substituted $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$, we write them for example $\mu_{ij}(\boldsymbol{\beta}_0) = \mu_{ij,0}$ and $\mu_{ij}(\hat{\boldsymbol{\beta}}) = \hat{\mu}_{ij}$, respectively. Furthermore, in order to evaluate the asymptotic properties of GEE estimator, we assume that following conditions (Xie and Yang, 2003):

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

C2. $\boldsymbol{\beta}_0$ is in interior of admissible set $\mathcal{B}$, and $\mathcal{B}$ is an open set of $\mathbb{R}^p$,
i.e. $\boldsymbol{\beta}_0 \in \mathcal{B}^\circ$, $\mathcal{B} = \{\boldsymbol{\beta} | u^{-1}(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) \in \Theta, \boldsymbol{x}_{ij} \in \mathcal{X}\}$.

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

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

C5. $\boldsymbol{H}_{n,0}$ and $\boldsymbol{M}_{n,0}$ are both positive definite when $n$ is large, and $\boldsymbol{H}_n$ and $\boldsymbol{M}_n$ are defined as follows:

$$\boldsymbol{H}_n = \sum_{i=1}^{n} \boldsymbol{D}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{D}_i, \boldsymbol{M}_n = \sum_{i=1}^{n} \boldsymbol{D}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{\Sigma}_i\boldsymbol{V}_i^{-1}\boldsymbol{D}_i.$$

C6. $\liminf_{n\to\infty} \lambda_{\min}(\boldsymbol{H}_{n,0}/n) > 0$, where $\lambda_{\min}(\boldsymbol{A})$ is minimum eigenvalue of $\boldsymbol{A}$.

C7. In a neighborhood of $\boldsymbol{\beta}_0$, say $N_0$, there exists that consistent $c_0 > 0$ and $n_0$, for all $p$-dimensional vector $\boldsymbol{\lambda}$, where $|\boldsymbol{\lambda}| = 1$, when $n \geq n_0$, it is established follows:

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

C8. GEE has unique solution when $n$ is large.

C1, C2 and C3 are necessary to consider GLM framework. C4 and C5 are necessary to calculate the asymptotic bias of estimator of risk. In addition, C1, C6, C7 and C8 (modified Xie and Yang, 2003) are necessary to have the strong consistency and asymptotic normality, uniqueness of GEE estimator.

Furthermore, we assume following conditions by additions.

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

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

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

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

C12. $\sqrt{n}(\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0) = O_p(1)$. And there exists that bounded $s \times p$ nonstochastic matrix $\boldsymbol{\mathcal{H}}$ such that $(\hat{\boldsymbol{\alpha}}^{(1)}(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{H}}) = O_p(n^{-1/2})$.

C13.

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

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

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

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

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

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

We write about $\boldsymbol{h}_{1,0}, \boldsymbol{j}_{1,0}, \boldsymbol{A}_{f,i,0}^{*}, \boldsymbol{b}_{f,0}$ later.

C9, C10, C11, C12 and C13 are necessary that in order to ignore the influence of estimating nuisance parameter $\boldsymbol{\alpha}$. Furthermore, by condition C5, it is established $\boldsymbol{H}_{n,0} = O(n)$.

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

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

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

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

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

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

By Lindberg central limit theorem, $\boldsymbol{L}_1(\boldsymbol{\beta}^*) = O_p(n)$, $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0, \boldsymbol{\mathcal{D}}_{1,0}, \boldsymbol{\mathcal{D}}_{2,0} = O_p(n^{-1/2})$. And $\boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}}_0)$ is expanded as follows:

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

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

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

i.e. $\boldsymbol{R}(\boldsymbol{\alpha}_0) - \boldsymbol{R}(\hat{\boldsymbol{\alpha}}_0) = O_p(n^{-1/2})$. Hence, it follows that

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

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

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

Also, since

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

and above these conclusions, the GEE is expanded as follows:

$$\begin{aligned}\boldsymbol{s}_{n,0} =& \boldsymbol{H}_{n,0}(\boldsymbol{I}_p + \boldsymbol{G}_{1,0} + \boldsymbol{G}_{2,0} + \boldsymbol{G}_{3,0} + \boldsymbol{h}_{1,0})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\\ &- \frac{1}{2}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_p\}\{\boldsymbol{\mathcal{S}}_{1,0} + (\boldsymbol{L}_1(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{S}}_{1,0})\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\\ &- \frac{1}{6}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \otimes \boldsymbol{I}_p\}\left\{\frac{\partial}{\partial \boldsymbol{\beta}'} \otimes \left(\frac{\partial}{\partial \boldsymbol{\beta}} \otimes \frac{\partial \boldsymbol{s}_n}{\partial \boldsymbol{\beta}'}\right)\right\}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{**}}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \otimes (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}.\end{aligned} \tag{2.4}$$

where $\boldsymbol{\beta}^{**}$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$. Denote $\boldsymbol{\mathcal{S}}_{1,0} = \mathrm{E}[\boldsymbol{L}_1(\boldsymbol{\beta}_0)]$. Note that $\boldsymbol{\mathcal{S}}_{1,0} = O_p(n)$, $\boldsymbol{L}_1(\boldsymbol{\beta}_0) - \boldsymbol{\mathcal{S}}_{1,0} = O_p(n^{1/2})$. The last term of (2.4) 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:

$$
\begin{aligned}
&\boldsymbol{\mathcal{C}}_{1i} = \boldsymbol{D}_i^{'}\boldsymbol{A}_i^{-1/2}\boldsymbol{R}^{-1}(\boldsymbol{\alpha}_0), \boldsymbol{\mathcal{C}}_{2i} = \boldsymbol{D}_i^{'}\boldsymbol{A}_i^{-1/2}, \boldsymbol{\mathcal{C}}_{3i} = \boldsymbol{R}^{-1}(\boldsymbol{\alpha}_0)\boldsymbol{A}_i^{-1/2},\\
&\boldsymbol{G}_{1,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\mathcal{C}_{1i,0}\left(\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{A}_i^{-1/2}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\{\boldsymbol{I}_p\otimes(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\},\\
&\boldsymbol{G}_{2,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{\mathcal{C}}_{2i}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)[\boldsymbol{I}_p\otimes\{\boldsymbol{\mathcal{C}}_{3i,0}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\}],\\
&\boldsymbol{G}_{3,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{2i,0}\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{R}^{-1}(\hat{\boldsymbol{\alpha}})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right][\boldsymbol{I}_p\otimes\{\boldsymbol{A}_{i,0}^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})\}],\\
&\boldsymbol{h}_{1,0} = -\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{1i,0}\{\boldsymbol{R}(\boldsymbol{\alpha}_0)-\boldsymbol{R}(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{\mathcal{C}}_{1i,0}^{'}\boldsymbol{b}_{1,0},\\
&\boldsymbol{j}_{1,0} = \boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{\mathcal{C}}_{1i.,0}\{\boldsymbol{R}(\boldsymbol{\alpha}_0)-\boldsymbol{R}(\hat{\boldsymbol{\alpha}}_0)\}\boldsymbol{\mathcal{C}}_{3i,0}(\boldsymbol{y}_{i,0}-\boldsymbol{\mu}_{i,0}).
\end{aligned}
\tag{2.5}
$$

Note that $\boldsymbol{G}_{1,0}$, $\boldsymbol{G}_{2,0}$, $\boldsymbol{G}_{3,0} = O_p(n^{-1/2})$, $\boldsymbol{h}_{1,0}, \boldsymbol{j}_{1,0} = O_p(n^{-1})$. By (2.5), $\hat{\boldsymbol{\beta}}$ is expanded as follows:

$$
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \boldsymbol{b}_{1,0} + \boldsymbol{b}_{2,0} + O_p(n^{-3/2}). \tag{2.6}
$$

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

# 3 MAIN RESULT

In this section, we propose new variable selection criterion. We measured 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 $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})^{'}$ is $m$-dimensional random vector that is independent of $\boldsymbol{y}_i$ and has same distribution of $\boldsymbol{y}_i$. If $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$, $\mathrm{Risk}_P$ has the minimum value of zero, i.e. , PMSE has the minimum value of $mn$. We consider the model which has minimum PMSE is optimum model, and select this model. Since the PMSE is typically unknown, we must estimate it.

We define $\boldsymbol{R}_0$, $\boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and $\boldsymbol{\mathcal{L}}^*(\boldsymbol{\beta})$ as follows:

$$
\begin{aligned}
\boldsymbol{R}_0(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A}_i^{-1/2}(\boldsymbol{y}_i-\boldsymbol{\mu}_i)(\boldsymbol{y}_i-\boldsymbol{\mu}_i)^{'}\boldsymbol{A}_i^{-1/2}/\hat{\phi},\\
\boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= \sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_1))^{'}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_2)\boldsymbol{R}_0^{-1}(\boldsymbol{\beta}_2)\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_2)(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_1))\hat{\phi}^{-1}(\boldsymbol{\beta}_2),\\
\boldsymbol{\mathcal{L}}^*(\boldsymbol{\beta}) &= \sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_i)^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_i).
\end{aligned}
$$

Then, we estimate the PMSE by $\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ where $\hat{\boldsymbol{\beta}}_f$ is the GEE estimator from full model namely we obtain $\hat{\boldsymbol{\beta}}_f$ as the solution to 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 positive definite working correlation one can choose freely. Also $\bar{\boldsymbol{R}}_i(\boldsymbol{\alpha}_f)$ is the same for all candidate models. For simplicity, we denote $\boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_2) = \boldsymbol{\mathcal{L}}(\boldsymbol{\beta}_2)$ and $\boldsymbol{\mathcal{L}}^*(\boldsymbol{\beta}_0) = \boldsymbol{\mathcal{L}}^*$.

We need to evaluate the asymptotic bias of $\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ from PMSE in order to propose the new variable selection criterion because $\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ is not the asymptotic unbiased estimator of PMSE. The bias we estimate the PMSE by $\boldsymbol{\mathcal{L}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_f)$ is given as

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

We evaluate Bias1, Bias2, Bias3 and Bias4 separately.

At first, Bias3 is as follows

$$\begin{aligned}
\text{Bias3} =& \text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}(\hat{\boldsymbol{\beta}}_f)\}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] \\
=& mn - \text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}(\hat{\boldsymbol{\beta}}_f)(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right].
\end{aligned}$$

This term is not depending on the candidate model so we can ignore calculation of Bias3 for variable selection.

Second, Bias1 is expanded as follows:

$$\begin{aligned}
\text{Bias1} =& \text{E}_y\left[\text{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}}_i)\right] - \sum_{i=0}^{n}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)\right] \\
=& \text{E}_y\Bigg[\text{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
&- \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0} + \boldsymbol{\mu}_{i,0} - \boldsymbol{m}\hat{\boldsymbol{u}}_i)\Bigg] \\
=& \text{E}_z\left[\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_{i,0})\right] + \text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
&- \text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] - 2\text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \boldsymbol{m}\hat{\boldsymbol{u}}_i)\right] \\
&- \text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{\Sigma}_{i,0}^{-1}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] \\
=& 2\text{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0})\right]. \qquad (3.1)
\end{aligned}$$

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 the Taylor expansion around $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$, $\hat{\boldsymbol{\mu}}_i$ is expanded as follows:

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

where $\boldsymbol{\beta}^{***}$ lies between $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$. Substitute (2.6) for expansion of $\hat{\boldsymbol{\mu}}_i$, we can expand $\hat{\boldsymbol{\mu}}_i$ as follows:

$$\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0} + \left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}^{'} \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\} + O_p(n^{-3/2}). \tag{3.2}$$

By (3.1) and (3.2), we get the under conclusion

$$\begin{aligned}\frac{1}{2}\text{Bias1} =& \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}(\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0})\right] \\ =& \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{i,0}\right] \\ &+ \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}^{'} \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\}\right] \\ &+ \mathrm{E}_y[O_p(n^{-1/2})].\end{aligned} \tag{3.3}$$

Since the data from different two subjects are independent,
$\mathrm{E}[(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})] = 0, (i \neq j)$. The first term of (3.3) is calculated as follows

$$\begin{aligned}&\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{i,0}\right] \\ =& \mathrm{E}_y\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{j,0}^{'}\boldsymbol{V}_{j,0}^{-1}(\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})\right] \\ =& \mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right] \\ =& \mathrm{E}_y\left[\mathrm{tr}\left\{\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})\right\}\right] \\ =& \mathrm{E}_y\left[\mathrm{tr}\left\{\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right\}\right] \\ =& \mathrm{tr}\left\{\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}\mathrm{E}\left[(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\right]\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right\} \\ =& \mathrm{tr}\left\{\boldsymbol{H}_{n,0}^{-1}\sum_{i=1}^{n}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}\boldsymbol{D}_{i,0}\right\} \\ =& \mathrm{tr}\left\{\boldsymbol{I}_p\right\} \\ =& p,\end{aligned} \tag{3.4}$$

Also, for all $i, j, k$(not $i = j = k$), since $\mathrm{E}\left[(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0}) \otimes (\boldsymbol{y}_j - \boldsymbol{\mu}_{j,0})^{'}(\boldsymbol{y}_k - \boldsymbol{\mu}_{k,0})\right] = \boldsymbol{0}_m$, the second term of (3.3) is calculated as follows:

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

where $\boldsymbol{b}_{1i,0} = \boldsymbol{H}_{n,0}^{-1}\boldsymbol{D}_{i,0}^{'}\boldsymbol{V}_{i,0}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})$, $\boldsymbol{b}_{2i,0} = \boldsymbol{H}_{n,0}^{-1}(\boldsymbol{b}_{1i,0}^{'} \otimes \boldsymbol{I}_p)\boldsymbol{\mathcal{S}}_{1,0}\boldsymbol{b}_{1i,0}/2 - \boldsymbol{G}_{1i,0}\boldsymbol{b}_{1i,0} - \boldsymbol{G}_{2i,0}\boldsymbol{b}_{1i,0} - \boldsymbol{G}_{3i,0}\boldsymbol{b}_{1i,0} + \boldsymbol{h}_{1,0} + \boldsymbol{j}_{1,0}$ and

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

Under the condition C13,

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

so the second term of (3.3) is calculated as follows:

$$\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})^{'}\boldsymbol{\Sigma}_{i,0}^{-1}\left\{\boldsymbol{D}_{i,0}\boldsymbol{b}_{2,0} + \frac{1}{2}(\boldsymbol{b}_{1,0}^{'} \otimes \boldsymbol{I}_m)\boldsymbol{D}_{i,0}^{(1)}\boldsymbol{b}_{1,0}\right\}\right] = O(n^{-1}), \tag{3.5}$$

Under the regularity condition, 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 $n^{-1}$ (Hall, 1992). Therefore, by substitute (3.4) and (3.5) for (3.3), we obtain

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

Similarly, we obtain

$$\mathrm{Bias2} + \mathrm{Bias4} = O(n^{-1}), \tag{3.6}$$

The derivation of (3.6) is shown in Appendix.

From the above, the bias is expanded as follows:

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

Note that Bias3 is not depend on the candidate model so we propose the new variable selection criterion as

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

We call this criterion PMSEG (the prediction mean squared error in the GEE).

# 4 NUMERICAL STUDY

In this section, we perform the numerical study and discuss results. In this paper we perform numerical study in situation which very restrictive.

Let the number of models be 8, and $m = 3$. The number of subjects $n = 50, 100, 150, 200$. We perform Monte Carlo simulation with 10000 iterations.

First, explanatory matrix $\boldsymbol{X}_{f,i}$ is $8 \times 3$ matrix, let $\boldsymbol{X}_{f,i} = (\boldsymbol{x}_{f,i1}, \ldots, \boldsymbol{x}_{f,i8})^{'}$ and $\boldsymbol{x}_{f,i1} = (1,1,1)^{'}, \boldsymbol{x}_{f,i2} = (0,1,2)^{'}, \boldsymbol{x}_{f,i3} = (0,1,1)^{'}$. Furthermore, $\boldsymbol{x}_{f,i4} = (1,1,1)^{'}$, $\boldsymbol{x}_{f,i5} = (0,1,2)^{'}$, $\boldsymbol{x}_{f,i6} = (0,1,1)^{'}$ if male, and $\boldsymbol{x}_{f,i4} = (0,0,0)^{'}$, $\boldsymbol{x}_{f,i5} = (0,0,0)^{'}$, $\boldsymbol{x}_{f,i6} = (0,0,0)^{'}$ if female, and $\boldsymbol{x}_{f,i7}$ and $\boldsymbol{x}_{f,i8}$ have uniform distribution on the interval [-1, 1].

Let

$$\boldsymbol{R}_0 = \begin{pmatrix} 1 & \frac{1}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{6}} & 1 & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{6}} & 1 \end{pmatrix}, \boldsymbol{\beta}_0 = \left(\frac{1}{2}, \frac{1}{6}, -\frac{1}{3}, -\frac{1}{4}, -\frac{1}{12}, \frac{1}{6}, 0, 0\right)^{'},$$

and rink function is inverse link.

We prepare exchangeable(Ex.), Autoregressive(AR) and Independence(Ind.) correlation structure as working correlation. In this simulation, we divided $\boldsymbol{y} = (y_{11}, \ldots, y_{n3})'$ into $\boldsymbol{u} = (y_{11}, \ldots, y_{(n/2),3})'$ and $\boldsymbol{v} = (y_{(n/2)+1,1}, \ldots, y_{n3})'$, where $\boldsymbol{u} \sim \text{Gamma}(1,1)$, $\boldsymbol{v} \sim \text{Gamma}(2,1)$. For full model we choose independence correlation matrix for instance.

The figure 1 is the frequency of selection each models.

Table 1: Frequency of selection each models

| n | W-Cor. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 50 | Ex. | 0.0000 | 0.0000 | 0.0000 | 0.0310 | 0.0003 | 0.2132 | 0.0123 | 0.0048 |
| | AR | 0.0000 | 0.0000 | 0.0000 | 0.1456 | 0.0091 | 0.1310 | 0.0075 | 0.0026 |
| | Ind. | 0.0000 | 0.0000 | 0.0000 | 0.3254 | 0.0225 | 0.0863 | 0.0055 | 0.0029 |
| 100 | Ex. | 0.0000 | 0.0000 | 0.0000 | 0.0038 | 0.0001 | 0.3219 | 0.0170 | 0.0048 |
| | AR | 0.0000 | 0.0000 | 0.0000 | 0.0910 | 0.0044 | 0.2111 | 0.0071 | 0.0013 |
| | Ind. | 0.0000 | 0.0000 | 0.0000 | 0.1722 | 0.0074 | 0.1478 | 0.0072 | 0.0029 |
| 150 | Ex. | 0.0000 | 0.0000 | 0.0000 | 0.0007 | 0.0000 | 0.3481 | 0.0221 | 0.0063 |
| | AR | 0.0000 | 0.0000 | 0.0000 | 0.0543 | 0.0014 | 0.2735 | 0.0062 | 0.0019 |
| | Ind. | 0.0000 | 0.0000 | 0.0000 | 0.0686 | 0.0030 | 0.2057 | 0.0062 | 0.0020 |
| 200 | Ex. | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.3889 | 0.0254 | 0.0052 |
| | AR | 0.0000 | 0.0000 | 0.0000 | 0.0224 | 0.0009 | 0.2838 | 0.0106 | 0.0029 |
| | Ind. | 0.0000 | 0.0000 | 0.0000 | 0.0264 | 0.0006 | 0.2187 | 0.0115 | 0.0027 |

We can see frequency of selection of true model is large as $n$ is large. The frequency of the model which don't have true explanatory variables is decrease as $n$ is large.

Table 2: Risk and prediction error

| n | W-Cor. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | prediction error |
|---|---|---|---|---|---|---|---|---|---|---|
| 50 | Ex. | 98.3093 | 100.0501 | 45.5836 | 7.9348 | 9.2082 | 6.0060 | 7.4445 | 8.8537 | 5.6396 |
| | AR | 87.8126 | 88.9928 | 45.5836 | 7.8969 | 9.0771 | 6.0060 | 7.2654 | 8.4875 | |
| | Ind. | 98.3093 | 100.0501 | 45.5836 | 8.0313 | 9.2971 | 6.0060 | 7.6263 | 9.2654 | |
| 100 | Ex. | 195.1645 | 196.9164 | 87.7755 | 11.8481 | 13.1198 | 5.9952 | 7.3366 | 8.6811 | 5.9350 |
| | AR | 174.0422 | 175.2199 | 87.7755 | 11.7757 | 12.9520 | 5.9952 | 7.1585 | 8.3373 | |
| | Ind. | 195.1645 | 196.9164 | 87.7755 | 12.0015 | 13.2638 | 5.9952 | 7.5605 | 9.1274 | |
| 150 | Ex. | 292.2903 | 293.9994 | 130.0335 | 15.8707 | 17.1275 | 5.9782 | 7.3698 | 8.7719 | 5.9664 |
| | AR | 260.4902 | 261.6392 | 130.0335 | 15.7543 | 16.9207 | 5.9782 | 7.1743 | 8.3696 | |
| | Ind. | 292.2903 | 293.9994 | 130.0335 | 16.0921 | 17.3402 | 5.9782 | 7.6118 | 9.2310 | |
| 200 | Ex. | 389.0366 | 390.7803 | 172.1719 | 19.8464 | 21.1396 | 6.0095 | 7.3519 | 8.7245 | 6.0080 |
| | AR | 346.7245 | 347.8972 | 172.1719 | 19.6973 | 20.8954 | 6.0095 | 7.1701 | 8.3608 | |
| | Ind. | 389.0366 | 390.7803 | 172.1719 | 20.1195 | 21.4031 | 6.0095 | 7.5533 | 9.1635 | |

We can see the risks of model 6, 7, and 8 make no difference.

# A Appendix

In this section, we calculate Bias2 + Bias4. Bias2 and Bias are calculate respectively as follows:

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

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

Then Bias2 + Bias4 is calculated as follows:

$$\begin{aligned}
\text{Bias2} + \text{Bias4} =& \mathrm{E}_y\left[2\sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\mu}_{i,0})' \left\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f) \boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f) \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\} (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] && \text{(A.1)} \\
&+ \mathrm{E}_y\left[\sum_{i=1}^n (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)' \left\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f) \boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f) \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\} (\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right]. && \text{(A.2)}
\end{aligned}$$

In order to calculate this bias, we perform the stochastic expansion of $\boldsymbol{A}_i^{-1/2}, \boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f), \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_f), \hat{\boldsymbol{\beta}}_f$ and $\hat{\phi}(\hat{\boldsymbol{\beta}}_f)$. Then we denote $\boldsymbol{D}_{f,i} = \boldsymbol{A}_i(\boldsymbol{\beta}_f)\boldsymbol{\Delta}_i(\boldsymbol{\beta}_f)\boldsymbol{X}_{f,i}$, $\boldsymbol{D}_{f,i,0} = \boldsymbol{A}_{i,0}\boldsymbol{\Delta}_{i,0}\boldsymbol{X}_{f,i}$. We expand $\hat{\boldsymbol{\beta}}_f$ as with the expansion of $\hat{\boldsymbol{\beta}}$ in section 2.

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

where $\boldsymbol{\beta}_{f,0}$ is true value of $\boldsymbol{\beta}_f$, and define $\boldsymbol{H}_{f,n,0}$ be as follows:

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

In addition, we expand $\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_f)$ as with the expansion of $\hat{\boldsymbol{\mu}}_i$ in section 3.

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

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

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

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 $\boldsymbol{b}_{f,0}, \boldsymbol{D}_{f,i,0}\boldsymbol{b}_{f,0}, \mathrm{diag}(\boldsymbol{A}^{*}_{f,i,0}\boldsymbol{b}_{f,0}) = O_p(n^{-1/2})$. 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, $\boldsymbol{R}_0(\hat{\boldsymbol{\beta}}_f)$ is expanded as follows:

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

By Lindberg central limit theorem, the first term of (A.3) is $O_p(n^{-1})$. Then, we get under conclusion:

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

Therefore, by calculating the inverse matrix, we calculate as follows:

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

Therefore. $\boldsymbol{R}_0^{-1}$ is expanded as follows:

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

Note that the second term of (A.5) is $O_p(n^{-1/2})$. Next, we calculate (A.1) and (A.2).

$$
\begin{aligned}
& \boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-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})\}\Big[\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 \\
& - \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 \\
& - \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\Big\}\boldsymbol{R}_0^{-1}\Big]\{\boldsymbol{A}_{i,0}^{-1/2} + \mathrm{diag}(\boldsymbol{A}_{f,i,0}^{*}\boldsymbol{b}_{f,0})\}\phi_o^{-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 \\
& - \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 \\
& - \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\Big\}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1} + O_p(n^{-1}).
\end{aligned}
$$

where $\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}(\hat{\boldsymbol{\beta}}_f) = O_p(n^{-1/2})$, and $\hat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_{i,0} = \boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0} = O_p(n^{-1/2})$, so (A.2) is calculated as follows:

$$
\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)'\left\{\boldsymbol{\Sigma}_{i,0}^{-1} - \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\}(\boldsymbol{\mu}_{i,0} - \hat{\boldsymbol{\mu}}_i)\right] = O(n^{-1}).
$$

Finally, we calculate (A.1).

$$
\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)\boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]\\
=&\mathrm{E}_y\left[2\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\left\{\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})\right\}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&-\mathrm{E}_y\left[\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^n\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\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] \qquad \text{(A.6)}\\
&-\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})'\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&-\mathrm{E}_y\left[\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^n\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\mathrm{diag}(\boldsymbol{A}_{f,j,0}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&+\mathrm{E}_y\left[2\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]+O(n^{-1}).
\end{aligned}
$$

For instance, we define under notation of summation.

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

Note that $\mathrm{E}[(\boldsymbol{y}_{i,}-\boldsymbol{\mu}_{i,0})\otimes(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'(\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0})]=\boldsymbol{0}_m$, (not $i=j=k$), so we can expand the first term of (A.6) as follows:

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

where $\boldsymbol{b}_{f,i,0}=\boldsymbol{H}_{f,n,0}^{-1}\boldsymbol{D}'_{f,i,0}\boldsymbol{A}_{i,0}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})$.

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 (A.6) is expanded as follows:

$$
\begin{aligned}
&-\mathrm{E}_y\left[\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^n\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
=&-\mathrm{E}_y\left[\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1,i\neq j}^n\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})'\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]\\
&+O(n^{-1})\\
=&-\mathrm{E}_y\left[2\sum_{i=1}^n(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})'\boldsymbol{\Sigma}_{i,0}^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]+O(n^{-1})\\
=&-2p+O(n^{-1}). \qquad \text{(A.8)}
\end{aligned}
$$

The fact that $\mathrm{E}_y[(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0}\otimes\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0}^{'})(\boldsymbol{y}_k-\boldsymbol{\mu}_{k,0}\otimes\boldsymbol{y}_l-\boldsymbol{\mu}_{l,0})]=0$ when 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 (A.6) 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}\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})^{'}\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})^{'}\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
=&-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\mathrm{diag}(\boldsymbol{A}_{f,i,0}\boldsymbol{b}_{f,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&-\mathrm{E}_y\left[\sum_{i\neq j}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\mathrm{diag}(\boldsymbol{A}_{f,j,0}\boldsymbol{b}_{f,i,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})^{'}\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1j,0}\right]\\
&-\mathrm{E}_y\left[\sum_{i\neq j}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\mathrm{diag}(\boldsymbol{A}_{f,j,0}\boldsymbol{b}_{f,j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})^{'}\boldsymbol{A}_{j,0}^{-1/2}\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1i,0}\right]\\
&+O(n^{-1})\\
=&O(n^{-1}).
\end{aligned} \tag{A.9}$$

Similarly, the forth term of (A.6) is expanded as follows:

$$\begin{aligned}
&-\mathrm{E}_y\left[\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{R}_0^{-1}\frac{2}{n}\sum_{j=1}^{n}\boldsymbol{A}_{j,0}^{-1/2}(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})(\boldsymbol{y}_j-\boldsymbol{\mu}_{j,0})^{'}\mathrm{diag}(\boldsymbol{A}_{f,j,0}\boldsymbol{b}_{f,0})\boldsymbol{R}_0^{-1}\phi_0^{-2}\boldsymbol{A}_{i,0}^{-1/2}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]\\
&=O(n^{-1}).
\end{aligned} \tag{A.10}$$

Furthermore, the fifth term of (A.6) is expanded as with (3.4).

$$\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_0^{-1}\boldsymbol{A}_{i,0}^{-1/2}\phi_0^{-1}\boldsymbol{D}_{i,0}\boldsymbol{b}_{1,0}\right]=2p. \tag{A.11}$$

By (A.7), (A.8), (A.9), (A.10) and (A.11), (A.1) is calculated as follows:

$$\mathrm{E}_y\left[2\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\mu}_{i,0})^{'}\left\{\boldsymbol{\Sigma}_{i,0}^{-1}-\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{R}_0^{-1}(\hat{\boldsymbol{\beta}}_f)\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_f)\hat{\phi}^{-1}(\hat{\boldsymbol{\beta}}_f)\right\}(\boldsymbol{\mu}_{i,0}-\hat{\boldsymbol{\mu}}_i)\right]=O(n^{-1}).$$

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

# References


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

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

[3] Inatsu, Y. (2014). Model selection criterion based on the prediction mean squared error in generalized estimating equations. *Master's Thesis, Department of Mathematics, Graduate School of Science, Hiroshima University, Hiroshima.*

[4] Inatsu, Y. & Imori, S. (2013). Model selection criterion based on the prediction mean squared error in generalized estimating equations. *TR13-10, Statistical Research Group, Hiroshima University, Hiroshima.*

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

[6] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

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

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

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

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

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

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

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

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

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