Consistency of AIC and its modification in the growth curve model under a large-(q, n) framework

Rie Enomoto*, Tetsuro Sakurai** and Yasunori Fujikoshi***,

*Department of Mathematical Information Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo, 162-8601 Japan
**Center of General Education, Tokyo University of Science, Suwa, 5000-1 Toyohira, Chino, Nagano 391-0292, Japan
***Department of Mathematics, Graduate School of Science, Hiroshima University, Higashi-Hiroshima, 739-8626, Japan

Abstract

The AIC and its modifications have been proposed for selecting the degree in a polynomial growth curve model under a large-sample framework and a high-dimensional framework by Satoh, Kobayashi and Fujikoshi [8] and Fujikoshi, Enomoto and Sakurai [6], respectively. They notes that the AIC and its modifications have no consistency property. In this paper we consider asymptotic properties of the AIC and its modification when the number q of groups or explanatory variables and the sample size n are large. First we show that the AIC has a consistency property under a large-(q, n) framework such that $q/n \to d \in [0, 1)$, under a condition on the noncentrality matrix, but the dimension p is fixed. Next we propose a modification of the AIC (denoted by MAIC) which is an asymptotic unbiased estimator of the risk under the asymptotic framework. It is shown that MAIC has a consistency property under a condition on the noncentrality matrix. Our results are checked numerically by conducting a Mote Carlo simulation.

AMS 2000 subject classification: primary 62H12; secondary 62H30
Key Words and Phrases: AIC, Consistency property, Growth curve model, Modified criterion, Large-(q, n) framework, Selection of models.

1 Introduction

The growth curve model introduced by Potthoff and Roy [7] is written as

$$\mathbf{Y} = \mathbf{A}\mathbf{\Theta}\mathbf{X} + \boldsymbol{\mathcal{E}}, \tag{1.1}$$

where $\mathbf{Y}; n \times p$ is an observation matrix, $\mathbf{A}; n \times q$ is a design matrix across individuals, $\mathbf{X}; k \times p$ is a design matrix within individuals, $\mathbf{\Theta}$ is an unknown matrix, and each row of $\boldsymbol{\mathcal{E}}$ is independent and identically distributed as a p-dimensional normal distribution with mean $\mathbf{0}$ and an unknown covariance matrix $\mathbf{\Sigma}$. We assume that that $n - p - k - 1 > 0$, and $\mathrm{rank}(\mathbf{X}) = k$. If we consider a polynomial regression of degree $k - 1$ on the time t with q groups, then

$$\mathbf{A} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_q} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_p \\ \vdots & \vdots & \vdots & \vdots \\ t_1^{k-1} & t_2^{k-1} & \cdots & t_p^{k-1} \end{pmatrix}.$$

Relating to the problem of deciding the degree in a polynomial growth curve model, consider a set of candidate models $M_1, \ldots, M_k$ where M_j is defined by

$$M_j;\ \mathbf{Y} = \mathbf{A}\mathbf{\Theta}_j\mathbf{X}_j + \boldsymbol{\mathcal{E}}, \quad j = 1, \ldots, k, \tag{1.2}$$

where $\mathbf{\Theta}_j$ is the $q \times j$ submatrix of $\mathbf{\Theta}$, and $\mathbf{X}_j$ is the $j \times p$ submatrix of $\mathbf{X}$ defined by

$$\mathbf{\Theta} = (\mathbf{\Theta}_j, \mathbf{\Theta}_{\bar{j}}), \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_j \\ \mathbf{X}_{\bar{j}} \end{pmatrix}.$$

Here we note that the design matrix $\mathbf{A}$ may be also an observation matrix of several explanatory variables. For such an application, see Satoh and Yanagihara [9]. There are several criteria for selecting models including the AIC (Akaike [1]). The AIC for M_j is given by

$$\mathrm{AIC} = n \log |\hat{\Sigma}_j| + np(\log 2\pi + 1) + 2\{qj + \frac{1}{2}p(p+1)\}, \tag{1.3}$$

where $\hat{\mathbf{\Sigma}}_j$ is the MLE of Σ under M_j, which is given by

$$\hat{\mathbf{\Sigma}}_j = \frac{1}{n}(\mathbf{Y} - \mathbf{A}\hat{\mathbf{\Theta}}_j\mathbf{X}_j)'(\mathbf{Y} - \mathbf{A}\hat{\mathbf{\Theta}}_j\mathbf{X}_j),$$

where $\hat{\boldsymbol{\Theta}}_j = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{S}^{-1}\mathbf{X}_j'(\mathbf{X}_j\mathbf{S}^{-1}\mathbf{X}_j')^{-1}$, $\mathbf{S} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\mathbf{A})\mathbf{Y}/(n-q)$, and $\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. The last term $\{qj + p(p+1)/2\}$ is the number of independent parameters under M_j. In addition to AIC, some bias-corrected criteria have been proposed. Satoh, Kobayashi and Fujikoshi [8] proposed a higher-order asymptotic unbiased estimator of the risk function under a large-sample framework,

$$p,\ q \text{ and } k \text{ are fixed},\ n \to \infty. \tag{1.4}$$

Recently, Fujikoshi, Enomoto and Sakurai [6] have proposed HAIC which is a higher-order asymptotic estimator of the risk under a high-dimensional asymptotic framework,

$$q \text{ and } k \text{ are fixed},\ p \to \infty,\ n \to \infty,\ p/n \to c \in [0,1). \tag{1.5}$$

For the two bias-corrected AICs, it was assumed that the true model is included in the full model M_k. The assumption is also assumed in this paper. So, without loss of generality, we may assume that the minimum model including the true model is M_{j_0}, and then the true model is expressed as

$$M_{j_0} : \mathbf{Y} \sim N_{n\times p}(\mathbf{A}\boldsymbol{\Theta}_0\mathbf{X}_{j_0}, \boldsymbol{\Sigma}_0 \otimes \mathbf{I}_n), \tag{1.6}$$

where $\boldsymbol{\Theta}_0$ is a given $q \times j_0$ matrix, and $\boldsymbol{\Sigma}_0$ is a given positive definite matrix. For simplicity, we write $\mathbf{X}_{j_0}$ as $\mathbf{X}_0$.

The purpose of this paper is to study asymptotic properties of the AIC when the number q of groups or explanatory variables and the sample size n are large, but the dimension p is fixed. In general, it may be happen when the number q of explanatory variables is large. Further, the number q of groups will increase in the following cases:

(1) The groups are based on many clusters.

(2) The groups are constructed by repeated measurements of each the subjects.

First we show that the AIC has a consistency property under a large-(q, n) framework such that

$$p \text{ and } k \text{ are fixed},\ q \to \infty,\ n \to \infty,\ q/n \to d \in [0,1), \tag{1.7}$$

under a condition on the noncentralty matrix defined in Lemma 2.1. The fact might be interesting, since the AIC has no consistency property under a large-sample framework (1.4) and a high-dimensional framework (1.5). Next we propose a modification of the AIC (denoted by MAIC) which is an asymptotic unbiased estimator of the risk under (1.7). Further, it is shown that MAIC has a consistency property under a condition on the noncentrality matrix. Our results are checked numerically by conducting a Mote Carlo simulation. Some future problems are discussed in the final section.

2 Preliminaries

In this section we prepare some distributional results on the AIC itself and its bias as an estimator of the risk. For a detail derivation, see Fujikoshi, Enomoto and Sakurai [6]. Let

$$\mathbf{H}_1^{(j)} = (\mathbf{X}_j \mathbf{\Sigma}_0^{-1/2})'(\mathbf{X}_j \mathbf{\Sigma}_0^{-1} \mathbf{X}_j')^{-1/2};\ p \times j, \quad j = 1, \ldots, k,$$

and consider a $p \times p$ orthogonal matrix

$$\mathbf{H} = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_k; *),$$

satisfying $\boldsymbol{h}_1 \in \mathcal{R}[\mathbf{H}_1^{(1)}], (\boldsymbol{h}_1, \boldsymbol{h}_2) \in \mathcal{R}[\mathbf{H}_1^{(2)}], \cdots, (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_k) \in \mathcal{R}[\mathbf{H}_1^{(k)}]$, and the remainder $p - k$ columns are any ones such that $\mathbf{H}$ is an orthogonal matrix. We partition $\mathbf{H}$ as

$$\mathbf{H} = (\mathbf{H}_1^{(j)},\ \mathbf{H}_2^{(j)}),\ \mathbf{H}_1^{(j)};\ p \times j, \quad j = 1, \ldots, k.$$

Using the orthogonal matrix $\mathbf{H}$, we define the random matrices W and B as follows;

$$\mathsf{W} = \mathbf{H}'\mathbf{\Sigma}_0^{-1/2}(n-q)\mathsf{S}\mathbf{\Sigma}_0^{-1/2}\mathbf{H}, \tag{2.8}$$

$$\mathsf{B} = \mathbf{H}'\{(\mathbf{A}'\mathbf{A})^{-1/2}\mathbf{A}'\mathsf{Y}\mathbf{\Sigma}_0^{-1/2}\}'(\mathbf{A}'\mathbf{A})^{-1/2}\mathbf{A}'\mathsf{Y}\mathbf{\Sigma}_0^{-1/2}\mathbf{H}. \tag{2.9}$$

Then W and B are independently distributed as $\mathrm{W}_p(n-q, \mathbf{I}_p)$ and $\mathrm{W}_p(q, \mathbf{I}_p; \mathbf{\Gamma}_j'\mathbf{\Gamma}_j)$, where $\mathbf{\Gamma} = (\mathbf{A}'\mathbf{A})^{1/2}\mathbf{\Theta}_0\mathbf{X}_0\mathbf{\Sigma}_0^{-1/2}\mathbf{H}$. We use the following result which is obtained from (2.7) and (2.8) in Fujikoshi, Enomoto and Sakurai [6].

Lemma 2.1. *Let* $\mathbf{W}$ *and* $\mathbf{B}$ *be the random matrices defined by* (2.8) *and* (2.9)*, respectively. Then*

$$\frac{|(n-q)\mathbf{S}|}{|n\hat{\boldsymbol{\Sigma}}_j|} = \frac{|\mathbf{W}_{(j)}|}{|\mathbf{W}_{(j)} + \mathbf{B}_{(j)}|}, \tag{2.10}$$

where $\mathbf{W}_{(j)}$ *and* $\mathbf{B}_{(j)}$ *are the last* $(p-j) \times (p-j)$ *submatrices of* $\mathbf{W}$ *and* $\mathbf{B}$ *by respectively, that is*

$$\mathbf{W} = \begin{pmatrix} * & * \\ * & \mathbf{W}_{(j)} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} * & * \\ * & \mathbf{B}_{(j)} \end{pmatrix}.$$

Further

$$\mathbf{W}_{(j)} \sim \mathrm{W}_{p-j}(n-q, \mathbf{I}_{p-j}), \quad \mathbf{B}_{(j)} \sim \mathrm{W}_{p-j}(q, \mathbf{I}_{p-j}; \boldsymbol{\Omega}_j),$$

where $\boldsymbol{\Omega}_j = \boldsymbol{\Gamma}_j'\boldsymbol{\Gamma}_j$*, and* $\boldsymbol{\Gamma}_j = (\mathbf{A}'\mathbf{A})^{1/2}\boldsymbol{\Theta}_0\mathbf{X}_0\boldsymbol{\Sigma}_0^{-1/2}\mathbf{H}_2^{(j)}$.

The matrix $\boldsymbol{\Omega}_j$ is simply called a noncentrality matrix. As is well known, the AIC was proposed as an approximately unbiased estimator of the risk defined by the expected -2log-predictive likelihood. Let $f(\mathbf{Y}; \boldsymbol{\Theta}_j, \boldsymbol{\Sigma}_j)$ be the density function of $\mathbf{Y}$ under M_j. Then the expected -2log-predictive likelihood of M_j is defined by

$$R_A = \mathrm{E}^*_{\mathbf{Y}}\mathrm{E}^*_{\mathbf{Y}_F}[-2\log f(\mathbf{Y}_F; \hat{\boldsymbol{\Theta}}_j, \hat{\Sigma}_j)], \tag{2.11}$$

where $\hat{\boldsymbol{\Sigma}}_j$ and $\hat{\boldsymbol{\Theta}}_j$ are the maximum likelihood estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Theta}$ under M_j, respectively. Here $\mathbf{Y}_F; n \times p$ may be regarded as a future random matrix that has the same distribution as $\mathbf{Y}$ and is independent of $\mathbf{Y}$, and E^* denotes the expectation with respect to the true model. The risk is expressed as

$$R_A = \mathrm{E}^*_{\mathbf{Y}}\mathrm{E}^*_{\mathbf{Y}_F}[-2\log f(\mathbf{Y}; \hat{\boldsymbol{\Theta}}_j, \hat{\boldsymbol{\Sigma}}_j)] + b_A, \tag{2.12}$$

where

$$b_A = \mathrm{E}^*_{\mathbf{Y}}\mathrm{E}^*_{\mathbf{Y}_F}[-2\log f(\mathbf{Y}_F; \hat{\boldsymbol{\Theta}}_j, \hat{\boldsymbol{\Sigma}}_j) + 2\log f(\mathbf{Y}; \hat{\boldsymbol{\Theta}}_j, \hat{\boldsymbol{\Sigma}}_j)]. \tag{2.13}$$

The AIC and its modifications have been proposed by regarding b_A as the bias term when we estimate R_A by

$$-2\log f(\mathbf{Y};\hat{\boldsymbol{\Theta}}_j,\hat{\boldsymbol{\Sigma}}_j) = n\log|\hat{\boldsymbol{\Sigma}}_j| + np(\log 2\pi + 1),$$

and by evaluating the bias term b_A. The bias for M_j is expressed as in the following Lemma 2.2.

Lemma 2.2. *Suppose that the true model is given by* (1.6)*. Then, the bias* b_A *for model* M_j *in* (2.12) *or* (2.13) *is expressed in terms of* $\mathbf{W}_{(j)}$ *and* $\mathbf{B}_{(j)}$ *in Lemma* 2.1 *as follows:*

$$b_A = b_{A1} + b_{A2}, \tag{2.14}$$

where

$$b_{A1} = -np + \frac{n(n+q)(n-q-1)j}{(n-q-p-1)(n-q-p+j-1)}, \tag{2.15}$$

$$b_{A2} = \mathrm{E}\left[n^2 \mathrm{tr}(\mathbf{W}_{(j)} + \mathbf{B}_{(j)})^{-1}\left(\mathbf{I} + \frac{1}{n}\boldsymbol{\Omega}_j\right)\right]. \tag{2.16}$$

3 Consistency of AIC

In this section we show that the asymptotic probability of selecting the true model by the AIC goes to 1 as the number q and the sample size n approaching to ∞ as in (1.7), under the several assumptions. Let $\mathcal{F}$ be the set of candidate models, which is denoted by $\mathcal{F} = \{1,\ldots,k\}$, where j means the model M_j. We denote the AIC for M_j by AIC_j. The best model chosen by minimizing the AIC is written as

$$\hat{j}_{\mathrm{A}} = \arg\min_{j\in\mathcal{F}} \mathrm{AIC}_j.$$

Our main assumptions are summarized as follows:

A1 (The true model M_0): $j_0 \in \mathcal{F}$.

A2 (The asymptotic framework): $q\to\infty,\ n\to\infty,\ q/n\to d\in[0,1)$.

A3 (The noncentrality matrix): For $j < j_0$,

$$\boldsymbol{\Omega}_j = n\boldsymbol{\Delta}_j = \mathrm{O}_g(n) \text{ and } \lim_{q/n\to d}\boldsymbol{\Delta}_j = \boldsymbol{\Delta}_j^*.$$

Here $\mathrm{O}_g(n^i)$ denotes the term of i-th order with respect to n under (1.7).

Theorem 3.1. *Suppose that the assumptions* A1, A2 *and* A3 *are satisfied. Let* d_{a} (≈ 0.797) *be the constant satisfying* $\log(1-d_{\mathrm{a}})+2d_{\mathrm{a}}=0$. *Further, assume that* $d \in [0, d_{\mathrm{a}})$, *and*

A4*: For any* $j < j_0$,

$$\log|\mathbf{I}_{p-j}+\mathbf{\Delta}_j^*| > (j_0-j)\{2d+\log(1-d)\}.$$

Then, the asymptotic probability of selecting the true model j_0 *by the* AIC *tends to* 1, i.e.

$$\lim_{q/n\to d} P(\hat{j}_{\mathrm{A}}=j_0)=1.$$

Proof. Using Lemma 2.1 we have

$$\begin{aligned}\mathrm{AIC}_j-\mathrm{AIC}_{j_0} &= -n\log\frac{|(n-q)\mathbf{S}|}{|n\hat{\mathbf{\Sigma}}_j|}-\left(-n\log\frac{|(n-q)\mathbf{S}|}{|n\hat{\mathbf{\Sigma}}_{j_0}|}\right)+2q(j-j_0)\\ &= -n\log\frac{|\mathbf{W}_{(j)}|}{|\mathbf{W}_{(j)}+\mathbf{B}_{(j)}|}-\left\{-n\log\frac{|\mathbf{W}_{(j_0)}|}{|\mathbf{W}_{(j_0)}+\mathbf{B}_{(j_0)}|}\right\} \qquad (3.17)\\ &\quad +2q(j-j_0).\end{aligned}$$

Let $\mathbf{V}_{(j)}$ and $\mathbf{U}_{(j)}$ be defined by

$$\begin{aligned}\mathbf{V}_{(j)} &= \sqrt{n-q}\left(\frac{1}{n-q}\mathbf{W}_{(j)}-\mathbf{I}_{p-j}\right), \text{ and}\\ \mathbf{U}_{(j)} &= \sqrt{q}\left(\frac{1}{q}\mathbf{B}_{(j)}-\mathbf{I}_{p-j}-\frac{n}{q}\mathbf{\Delta}_j\right),\end{aligned}$$

respectively. Then, $\mathbf{V}_{(j)}$ and $\mathbf{U}_{(j)}$ converge to normal distributions, and we have

$$\frac{1}{n}\mathbf{W}_{(j)}=\frac{n-q}{n}\cdot\frac{1}{n-q}\mathbf{W}_{(j)}\to(1-d)\mathbf{I}_{p-j}, \qquad (3.18)$$

$$\frac{1}{n}\mathbf{B}_{(j)}=\frac{q}{n}\frac{1}{q}\mathbf{U}_{(j)}\to d\left(\mathbf{I}_{p-j}+\frac{1}{d}\mathbf{\Delta}_j^*\right)=d\mathbf{I}_{p-j}+\mathbf{\Delta}_j^*. \qquad (3.19)$$

Therefore

$$
\begin{aligned}
-\log \frac{|\mathbf{W}_{(j)}|}{|\mathbf{W}_{(j)} + \mathbf{B}_{(j)}|} &\to -\log \frac{|(1-d)\mathbf{I}_{p-j}|}{|(1-d)\mathbf{I}_{p-j} + d\mathbf{I}_{p-j} + \boldsymbol{\Delta}_j^*|} \\
&= \log |\mathbf{I}_{p-j} + \boldsymbol{\Delta}_j^*| - (p-j)\log(1-d).
\end{aligned}
$$

Since $\boldsymbol{\Delta}_{j_0}^* = \mathbf{O}$, we have

$$
\frac{1}{n}(\mathrm{AIC}_j - \mathrm{AIC}_{j_0}) \to \log |\mathbf{I}_{p-j} + \boldsymbol{\Delta}_j^*| + (j - j_0)\{2d + \log(1-d)\}.
$$

By the way it is easily checked that if $0 < d < d_a$, $2d + \log(1-d) > 0$. Therefore, for $j = j_0 + 1, \ldots, k$, we have

$$
\frac{1}{n}(\mathrm{AIC}_j - \mathrm{AIC}_{j_0}) \to (j - j_0)\{2d + \log(1-d)\} > 0.
$$

Further, for $j = 1, \ldots, j_0 - 1$, from A4 we have

$$
\frac{1}{n}(\mathrm{AIC}_j - \mathrm{AIC}_{j_0}) \to \log |\mathbf{I}_{p-j} + \boldsymbol{\Delta}_j^*| - (j_0 - j)\{2d + \log(1-d)\} > 0.
$$

For the case $d = 0$, we can prove by considering the limt of $(1/q)(\mathrm{AIC}_j - \mathrm{AIC}_{j_0})$ in stead of $(1/n)(\mathrm{AIC}_j - \mathrm{AIC}_{j_0})$. These complete the proof. □

4 Modification of AIC

In this section we first obtain an asymptotic expansion of b_A, assuming A1, A2 and A3. Then, using the expansion we obtain an asymptotic unbiased estimator of b_A.

Note that $\mathbf{W}_{(j)} + \mathbf{B}_{(j)} \sim \mathrm{W}_{p-j}(n, \mathbf{I}_{p-j}; \boldsymbol{\Omega}_j)$. Therefore, from an asymptotic result (see, e.g., Fujikoshi [2]) we have

$$
b_{A2} = n(p-j) + 2(p-j+1)\xi_1 - \xi_2 + \mathrm{O}_g(n^{-1}),
$$

where

$$
\xi_1 = \mathrm{tr}\left(\mathbf{I}_{p-j} + \frac{1}{n}\boldsymbol{\Omega}_j\right)^{-1}, \quad \xi_2 = \xi_1^2 + \mathrm{tr}\left(\mathbf{I}_{p-j} + \frac{1}{n}\boldsymbol{\Omega}_j\right)^{-2}.
$$

In the special case $\mathbf{\Omega}_j = \mathbf{O}$,

$$b_{A2} = \frac{n^2(p-j)}{n-(p-j)-1} = n(p-j) + \frac{n(p-j)(p-j+1)}{n-p+j-1}. \tag{4.20}$$

These results are summarized as follows:

$$b_{A2} = \begin{cases} n(p-j) + \dfrac{n(p-j)(p-j+1)}{n-p+j-1}, & \mathbf{\Omega}_j = \mathbf{O}, \\ n(p-j) + 2(p-j+1)\xi_1 - \xi_2 + \mathrm{O}_g(n^{-1}), & \mathbf{\Omega}_j \neq \mathbf{O}. \end{cases}$$

Now we look for an estimator $\hat{b}_A$ in the following form:

$$\hat{b}_A = b_{A1} + n(p-j) + 2(p-j+1)\hat{\xi}_1 - \hat{\xi}_2. \tag{4.21}$$

We wish to determine $\hat{\xi}_1$ and $\hat{\xi}_2$ satisfying the following properties:

(1) When $\mathbf{\Omega} = \mathbf{O}$, $\mathrm{E}[\hat{b}_A] = b_A$.

(2) When $\mathbf{\Omega} \neq \mathbf{O}$, $\mathrm{E}[\hat{b}_A] = b_A + \mathrm{O}_g(n^{-1})$.

It is known (see, e.g., Fujikoshi, Enomoto and Sakurai [6]) that

$$\mathrm{tr}(n\hat{\mathbf{\Sigma}}_j)^{-1}(n-q)\mathbf{S} = j + \mathrm{tr}\mathbf{Q}_j,$$
$$\mathrm{tr}\left\{(n\hat{\mathbf{\Sigma}}_j)^{-1}(n-q)\mathbf{S}\right\}^2 = j + \mathrm{tr}\mathbf{Q}_j^2,$$

where $\mathbf{Q}_j = \mathbf{W}_{(j)}(\mathbf{W}_{(j)} + \mathbf{B}_{(j)})^{-1}$. Using (3.18) and (3.19) we have

$$\mathbf{Q}_j \to (1-d)\left(\mathbf{I}_{p-j} + \mathbf{\Delta}_j^*\right)^{-1}.$$

Based on these results, let us consider the naive estimators defined by

$$\tilde{\xi}_1 = \frac{n}{n-q}\left\{\mathrm{tr}(n\mathbf{\Sigma}_j)^{-1}(n-q)\mathbf{S} - j\right\},$$
$$\tilde{\xi}_2 = (\tilde{\xi}_1)^2 + \left(\frac{n}{n-q}\right)^2\left[\mathrm{tr}\{(n\hat{\mathbf{\Sigma}}_j)^{-1}(n-q)\mathbf{S}\}^2 - j\right].$$

Then we can see that

$$\tilde{\xi}_1 \to \xi_{10} = \mathrm{tr}(\mathbf{I} + \mathbf{\Delta}_j^*)^{-1},$$
$$\tilde{\xi}_2 \to \xi_{20} = \left\{\mathrm{tr}(\mathbf{I} + \mathbf{\Delta}_j^*)^{-1}\right\}^2 + \mathrm{tr}(\mathbf{I} + \mathbf{\Delta}_j^*)^{-2}.$$

When $\mathbf{\Omega}_j = \mathbf{O}$, $\mathbf{Q}_j$ is distributed (see, e.g., Muirhead [5], Fujikoshi, Ulyanov and Shimizu [4]) as a multivariate beta distribution $B_{p-j}(\frac{1}{2}(n-q), \frac{1}{2}q)$. Using the moment formulas (see, e.g., Fujikoshi and Satoh [3]) on $\mathbf{Q}_j$ we have

$$
\begin{aligned}
\mathrm{E}_0[\tilde{\xi}_1] &= \left(\frac{n}{n-q}\right)\mathrm{E}_0[\mathrm{tr}\mathbf{Q}_j] = p-j, \\
\mathrm{E}_0[\tilde{\xi}_2] &= \left(\frac{n}{n-q}\right)^2 \mathrm{E}_0[(\mathrm{tr}\mathbf{Q}_j)^2 + \mathrm{tr}\mathbf{Q}_j^2] \\
&= \frac{n(p-j)}{3(n-q)}\left\{\frac{2(n-q+2)(p-j+2)}{n+2} + \frac{(n-q-1)(p-j-1)}{n-1}\right\}.
\end{aligned}
$$

Here E_0 means the expectation when $\mathbf{\Omega} = \mathbf{O}$. Now we modify $\tilde{\xi}_1$ and $\tilde{\xi}_2$ as

$$
\hat{\xi}_1 = \tilde{\xi}_1, \text{ and } \hat{\xi}_2 = f\tilde{\xi}_2,
$$

where f is a constant satisfying that $f = 1 + \mathrm{O}_g(n^{-1})$. Our purpose is to determine f such that $\hat{b}_A$ is an exact biased estimator of b_A when $\mathbf{\Omega}_j = \mathbf{O}$. This is equivalent to determine f such that

$$
2(p-j+1)\mathrm{E}_0[\tilde{\xi}_1] - f\mathrm{E}_0[\tilde{\xi}_2] = \frac{n(p-j)(p-j+1)}{n-p+j-1}.
$$

Therefore, the constant f may be determined as

$$
\begin{aligned}
f &= \frac{1}{\mathrm{E}[\tilde{\xi}_2]}(p-j)(p-j+1)\left\{2 - \frac{n}{n-p+j-1}\right\} \\
&= \frac{3(n-q)(p-j+1)(n-2p+2j-2)}{n(n-p+j-1)} \\
&\quad \times \left\{\frac{2(n-q+2)(p-j+2)}{n+2} + \frac{(n-q-1)(p-j-1)}{n-1}\right\}^{-1},
\end{aligned} \tag{4.22}
$$

which is $1 + \mathrm{O}_g(n^{-1})$. Consequently, as a modification of AIC we propose

$$
\mathrm{MAIC} = n\log|\hat{\Sigma}_j| + np(\log 2\pi + 1) + \hat{b}_A, \tag{4.23}
$$

where

$$
\begin{aligned}
\hat{b}_A &= b_{A1} + \hat{b}_{A2} \\
&= b_{A1} + n(p-j) + 2(p-j+1)\hat{\xi}_1 - \hat{\xi}_2.
\end{aligned}
$$

Here b_{A1} is given by (2.15). The $\hat{\xi}_1$ and $\hat{\xi}_2$ are given by

$$\hat{\xi}_1 = \frac{n}{n-q}\{\mathrm{tr}(n\hat{\mathbf{\Sigma}}_j)^{-1}(n-q)\mathbf{S} - j\},$$
$$\hat{\xi}_2 = f\left[\hat{\xi}_1^2 + \left(\frac{n}{n-q}\right)^2\left[\mathrm{tr}\{(n\hat{\mathbf{\Sigma}}_j)^{-1}(n-q)\mathbf{S}\}^2 - j\right]\right],$$

where f is defined by (4.22).

5 Consistency of MAIC

In this section we examine a consistency property of MAIC proposed by (4.23). We denote the MAIC for M_j by MAIC_j. The best model chosen by minimizing the AIC is written as

$$\hat{j}_{\mathrm{MA}} = \arg\min_{j\in\mathcal{F}} \mathrm{MAIC}_j.$$

Further, we denote b_A and $\hat{b}_A$ for model M_j by $b_{A;j}$ and $\hat{b}_{A;j}$, respectively. Similar notations are used for $\hat{\xi}_1, \hat{\xi}_2, \xi_{10}, \xi_{20}$, etc. Then we have seen in Section 4 that

$$\hat{\xi}_{1;j} \to \xi_{10;j}, \quad \hat{\xi}_{2;j} \to \xi_{20;j}.$$

Therefore, it is easily seen that

$$\frac{1}{n}\hat{b}_{A;j} = \frac{2q}{n-q}j + \mathrm{O}_g(n^{-1}).$$

This implies that

$$\frac{1}{n}(\hat{b}_{A;j} - \hat{b}_{A;j_0}) \to \frac{2d}{1-d}(j - j_0).$$

Using asymptotic results on AIC in Section 3 we have

$$\frac{1}{n}(\mathrm{MAIC}_j - \mathrm{MAIC}_{j_0}) \to \log|\mathbf{I}_{p-j} + \mathbf{\Delta}_j^*| + (j - j_0)\left\{\frac{2d}{1-d} + \log(1-d)\right\}. \quad (5.24)$$

Note that $f(d) = 2d(1-d)^{-1} + \log(1-d)$ is positive for $0 < d < 1$. In fact, put $f(x) = 2x(1-x)^{-1} + \log(1-x)$ for $0 < x < 1$. Then $\lim_{x\to+0} f(x) = 0$, and $f'(x) = (1+x)(1-x)^{-2} > 0$. This implies $f(d) > 0$ for $0 < d < 1$. Using (5.24), we have a Theorem similar to Theorem 3.1.

Theorem 5.2. *Suppose that the assumptions* A1*,* A2 *and* A3 *in Theorem* 3.1 *are satisfied. Further, suppose that*

A5*: For any* $j < j_0$*,*

$$\log |\mathbf{I}_{p-j} + \mathbf{\Delta}_j^*| > (j_0 - j)\left\{\frac{2d}{1-d} + \log(1-d)\right\}.$$

Then, the asymptotic probability of selecting the true model j_0 *by the* MAIC *tends to* 1*,* i.e.

$$\lim_{q/n\to d} P(\hat{j}_{\mathrm{MA}} = j_0) = 1.$$

For Theorem 5.2, the assumption $d \in [0, d_a)$ in Theorem 3.1 is not necessary. However, Assumption A5 is required instead of Assumption A4.

6 Simulation study

In this section, we numerically examine the validity of our claims. The five candidate models $M_1, \ldots, M_5$, with several different values of n and $q = dn$, were considered for Monte Carlo simulations, where $p = 5$, $n = 50, 100, 200$, $n_1 = \cdots = n_q = n/q$ and $d = 0.1, 0.2$. We constructed a 5×5 matrix $\mathbf{X}$ of explanatory variables with $t_i = 1 + (i-1)(p-1)^{-1}$. The true model was determined by $\mathbf{\Theta}_0 = \mathbf{1}_2\mathbf{1}_2'$ and $\mathbf{\Sigma}_0$ whose (i, j)th element was defined by $\rho^{|i-j|}$, where $\rho = 0.2, 0.8$. Thus, M_2 was the true model, the true model were included in M_3, M_4, M_5, the true model was not included in M_1. Therefore, $\mathbf{\Omega} = \mathbf{O}$ when M_2, M_3, M_4, M_5 and $\mathbf{\Omega} \neq \mathbf{O}$ when M_1.

In the above simulation model, convergent values in the conditions A4 and A5 for consistency are calculated as follows:

ρ	d	$\log\|\mathbf{I}_{p-j}+\mathbf{\Delta}_j^*\|$	$2d+\log(1-d)$	$2d/(1-d)+\log(1-d)$
0.2	0.1	0.440	0.095	0.117
	0.2	0.440	0.177	0.277
0.8	0.1	0.614	0.095	0.117
	0.2	0.614	0.177	0.277

First, we studied performances of AIC and MAIC as estimators of the AIC-type risk R_A. For each of $M_1, \ldots, M_5$, we computed the averages of R_A, AIC and MAIC by Monte Carlo simulations with 10^4 replications. Table 1 shows the risk R_A and the biases of AIC and MAIC to R_A, defined by "$R_A -$ (the expectation of the information criterion)". In Table 1, j means the model M_j and the bold face denotes the true model. From Table 1, we can see that the biases of MAIC were smaller than the ones of AIC. In general, there is a tendency that the biases become large as q increases. But the tendency of MAIC is very small in the comparison with AIC. Further, AIC has a tendency of underestimating the risk.

Table 2 gives the selection probabilities of AIC and MAIC based on the simulation experiment. When q increases, the probabilities of selecting the true model by AIC and MAIC are near to 1. Further, we can see that when (n, q) is relatively small and $d = 0.2$, MAIC has a tendancy of selecting underspecified models, but such tenancy is not seen for AIC.

7 Concluding remarks

This paper discusses with the AIC and its modification for selecting the degrees in the growth curve model (1.1) under a large-(q, n) framework (1.7). It was shown that the AIC has a consistency property under the assumptions A1, A2, A3, A4 and $d \in [0, d_a)$, where d_a is the solution of $\log(1-d)+2d = 0$ and d_a is approximately 0.797. Next we proposed a modified AIC (denoted by MAIC), which is a higher-order asymptotic unbiased estimator of the risk of AIC. Further, it was shown that MAIC has a consistency property under A1, A2, A3 and A5 without the assumption of $d \in [0, d_a)$.

It is interesting to study similar properties of C_p and MC_p which were

Table 1. Risks and biases of AIC and MAIC

$\rho = 0.2$		R_A	AIC	MAIC	R_A	AIC	MAIC	R_A	AIC	MAIC
$d = 0.1$		$(n, q) = (50, 5)$			$(n, q) = (100, 10)$			$(n, q) = (200, 20)$		
	1	751.84	10.11	0.94	1477.63	7.38	-0.36	2935.55	8.96	0.14
	2	738.90	15.05	0.66	1448.24	13.03	-0.51	2873.64	16.54	-0.07
j	3	746.98	18.77	0.78	1461.70	17.35	-0.72	2898.93	23.26	-0.02
	4	754.20	21.54	0.84	1474.74	21.10	-0.70	2923.63	29.28	0.09
	5	760.72	23.48	0.91	1487.13	24.18	-0.58	2947.58	34.33	-0.04
$d = 0.2$		$(n, q) = (50, 10)$			$(n, q) = (100, 20)$			$(n, q) = (200, 40)$		
	1	765.45	20.26	0.32	1499.82	22.63	1.20	2971.86	29.87	-0.03
	2	764.77	33.97	-0.25	1490.22	40.45	0.82	2945.65	57.18	-0.43
j	3	784.01	45.16	-0.24	1523.03	56.44	1.13	3005.79	82.85	-0.22
	4	801.12	53.92	-0.40	1553.92	70.14	1.13	3064.21	106.39	-0.28
	5	816.56	60.78	-0.39	1583.22	81.95	1.14	3121.08	128.14	-0.30
$\rho = 0.8$		R_A	AIC	MAIC	R_A	AIC	MAIC	R_A	AIC	MAIC
$d = 0.1$		$(n, q) = (50, 5)$			$(n, q) = (100, 10)$			$(n, q) = (200, 20)$		
	1	563.77	9.18	0.25	1103.82	8.53	1.00	2185.42	9.25	0.61
	2	542.07	14.10	-0.28	1057.17	14.36	0.82	2088.67	16.80	0.19
j	3	549.96	17.59	-0.40	1070.82	18.87	0.79	2113.99	23.48	0.20
	4	557.31	20.50	-0.20	1083.64	22.41	0.61	2138.59	29.30	0.10
	5	563.88	22.51	-0.06	1095.99	25.36	0.60	2162.47	34.32	-0.04
$d = 0.2$		$(n, q) = (50, 10)$			$(n, q) = (100, 20)$			$(n, q) = (200, 40)$		
	1	578.14	20.37	0.68	1124.34	21.27	0.04	2221.89	29.57	-0.14
	2	569.38	34.81	0.58	1097.56	39.38	-0.25	2161.18	57.53	-0.08
j	3	588.43	45.75	0.35	1130.35	55.23	-0.07	2221.19	82.78	-0.29
	4	605.49	54.46	0.14	1161.15	68.91	-0.09	2279.43	106.20	-0.46
	5	621.25	61.68	0.51	1190.20	80.49	-0.32	2336.38	128.15	-0.29

proposed by Satoh, Kobayashi and Fujikoshi [8]. For the noncentrality matrix $\mathbf{\Omega}_j$, we assumed that $\mathbf{\Omega}_j = \mathrm{O}(n)$. It is also important to study asymptotic properties of AIC, MAIC, C_p and MC_p under $\mathbf{\Omega}_j = \mathrm{O}_g(nq)$. The works of these directions are ongoing.

In the traditional growth curve model it is assumed that the dimension p is small or moderate. However, it is also important to analysis the data such that p is large. This suggests to study asymptotic properties of AIC and C_p under a high-dimensional framework such that

$$p \to \infty,\ q \to \infty,\ n \to \infty,\ p/n \to c \in [0, 1),\ q/n \to d \in [0, 1). \qquad (7.25)$$

Table 2. Selection probabilities (%) of AIC and MAIC

$\rho = 0.2$		AIC	MAIC	AIC	MAIC	AIC	MAIC
$d = 0.1$		$(n, q) = (50, 5)$		$(n, q) = (100, 10)$		$(n, q) = (200, 20)$	
	1	0.6	5.3	0.1	0.3	0.0	0.0
	2	84.6	90.8	94.7	98.5	98.8	99.9
j	3	10.5	3.4	4.4	1.1	1.1	0.1
	4	3.3	0.5	0.7	0.1	0.0	0.0
	5	1.1	0.1	0.1	0.0	0.0	0.0
$d = 0.2$		$(n, q) = (50, 10)$		$(n, q) = (100, 20)$		$(n, q) = (200, 40)$	
	1	5.2	52.1	1.1	24.4	0.1	7.2
	2	86.0	47.7	96.5	75.6	99.7	92.8
j	3	7.0	0.2	2.2	0.0	0.2	0.0
	4	1.4	0.0	0.1	0.0	0.0	0.0
	5	0.4	0.0	0.0	0.0	0.0	0.0
$\rho = 0.8$		AIC	MAIC	AIC	MAIC	AIC	MAIC
$d = 0.1$		$(n, q) = (50, 5)$		$(n, q) = (100, 10)$		$(n, q) = (200, 20)$	
	1	0.0	0.4	0.0	0.0	0.0	0.0
	2	86.2	96.1	94.8	98.8	98.9	99.9
j	3	9.8	2.9	4.7	1.2	1.1	0.2
	4	2.9	0.5	0.5	0.0	0.0	0.0
	5	1.1	0.2	0.1	0.0	0.0	0.0
$d = 0.2$		$(n, q) = (50, 10)$		$(n, q) = (100, 20)$		$(n, q) = (200, 40)$	
	1	0.5	19.9	0.0	2.2	0.0	0.0
	2	90.4	79.9	97.5	97.8	99.7	100.0
j	3	7.4	0.2	2.3	0.0	0.3	0.0
	4	1.4	0.0	0.2	0.0	0.0	0.0
	5	0.3	0.0	0.0	0.0	0.0	0.0

Modifications of AIC and C_p and their properties should be also studied. These works are left as a future subject.

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 and F. Csáki), 267–281, Akadémiai Kiadó, Budapest, 1973.

[2] Y. Fujikoshi, Selection of variables in discriminant analysis and canonical correlation analysis, In *Multivariate Analysis-VI*, Ed. P.R. Krishnaian, 219–236, Elsevier Science Publishers B.V., 1985.

[3] Y. Fujikoshi and K. Satoh, *Modified AIC and C_p in multivariate linear regression*, Biometrika **84** (1997), 707–716.

[4] Y. Fujikoshi, V. V. Ulyanov, and R. Shimizu, Multivariate Statistics: High-Dimensional and Large-Sample Approximations, Wiley, Hobeken, N. J., 2010.

[5] R. J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley & Sons, N. Y., 1982.

[6] Y. Fujikoshi, R. Enomoto and T. Sakurai, *High-dimensional* AIC *in the growth curve model. Submitted for publication*, (2012).

[7] R. F. Potthoff and S. N. Roy, *A generalized multivariate analysis of variance model useful especially for growth curve problems*, Biometrika **51** (1964), 313–326.

[8] K. Satoh, M. Kobayashi and Y. Fujikoshi, *Variable selection for the growth curve model*, J. Multivariate Anal. **60** (1997), 277–292.

[9] K. Satoh and Y. Yanagihara, *Estimation of varying coefficients for a growth curve model*, Amer. J. Math. Management Sci. **30** (2010), 243–256.