AIC for Growth Curve Model with Monotone Missing Data

Ayaka Yagi*, Takashi Seo* and Yasunori Fujikoshi**

**Department of Applied Mathematics*
Tokyo University of Science
1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

***Department of Mathematics*
Hiroshima University
Higashi-Hiroshima, Hiroshima 739-0046, Japan

Abstract

In this article, we consider an AIC for a one-sample version of the growth curve model when the dataset has a monotone pattern of missing observations. It is well known that the AIC can be regarded as an approximately unbiased estimator of the AIC-type risk defined by the expected (-2) log-predictive likelihood. Here, the likelihood is based on the observed data. First, when the covariance matrix is known, we derive an AIC, which is an exact unbiased estimator of the AIC-type risk function. Next, when the covariance matrix is unknown, we derive a conventional AIC using the estimators based on the complete data set only. Finally, a numerical example is presented to illustrate our model selection procedure.

Key Words and Phrases: AIC-type risk; Maximum likelihood estimator, Missing data, Monotone missing data.

1 Introduction

The growth curve model in the case of one-sample problem introduced by Rao (1959) may be expressed as

$$\boldsymbol{Y} = \boldsymbol{1}_n \boldsymbol{\theta}' \boldsymbol{X}' + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{Y}$ is an $n \times p$ observation matrix, $\boldsymbol{1}_n$ is an $n \times 1$ vector with all elements as 1, $\boldsymbol{X}$ is a $p \times q$ within-individuals design matrix, with $\mathrm{rank}(\boldsymbol{X}) = q$, of explanatory variables $x_1, x_2, \ldots, x_q$, $\boldsymbol{\theta}$ is a $q \times 1$ unknown parameter vector, and

$$\underset{n\times p}{\boldsymbol{\varepsilon}} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_n)', \quad \boldsymbol{\varepsilon}_i \overset{i.i.d.}{\sim} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}),\ i = 1, 2, \ldots, n.$$

If the rows of $\boldsymbol{Y}$ are expressed as

$$\underset{n\times p}{\boldsymbol{Y}} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)',$$

then

$$\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \ \ i = 1, 2, \ldots, n,$$

where $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\theta}$. This is a special case of the growth curve model introduced by Potthoff and Roy (1964). If we consider a polynomial regression of degree $q-1$ on time t, then

$$\underset{p\times q}{\boldsymbol{X}} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{q-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{q-1} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & t_p & t_p^2 & \cdots & t_p^{q-1} \end{pmatrix}.$$

Specifically, matrix $\boldsymbol{X}$ in a first- or second-degree polynomial growth curve model is given as

$$\underset{p\times 2}{\boldsymbol{X}} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_p \end{pmatrix}, \text{ or } \underset{p\times 3}{\boldsymbol{X}} = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_p & t_p^2 \end{pmatrix},$$

respectively. It is important to determine the degree of polynomial growth curve model. In general, such a problem can be considered a one of selecting a best subset of $\{x_1, x_2, \ldots, x_q\}$. One of the approaches is to use a model selection criterion. In this article, we consider the AIC or its modifications. It is well known that the AIC was proposed as an approximately unbiased estimator of the AIC-type risk defined by the expected (-2) log-predictive likelihood (see Akaike, 1973). It is expressed as the sum of $(-2)\log(\text{likelihood})$ and the correction term. The latter term was proposed to be twice the number of independent parameters. Some refinements of the correction term in the growth curve model were studied by Satoh et al. (1997) and Fujikoshi, Enomoto, and Sakurai (2013).

In contrast, a variant of the AIC has been considered by Shimodaira (1994) and Cavanaugh and Shumway (1998) when the observed data are incomplete. However, such approaches have not been extended in the growth curve model with missing data. In this article we consider the construction of an AIC based on the observed data, especially with a monotone pattern of missing observations. We prove that the AIC is an exact unbiased estimator of the AIC-type risk function when the covariance matrix is known. When the covariance matrix is unknown, we propose a conventional AIC that may be applicable in the situation such that the number of subjects with missing observations is smaller than that of subjects with complete observations.

The remainder of this article is organized as follows. In Section 2, in the case of two-step monotone missing data, some preliminary notations, and the definition of AIC-type risk are presented. It is shown that the correction term in AIC-type risk is exactly $2q$, where q is the dimension of $\boldsymbol{\theta}$, when the covariance matrix is known. These results are extended for general monotone missing data. In Section 3, we derive an AIC when we estimate the covariance matrix by using the set of complete data. Finally, we present the numerical results in Section 4 and conclusions in Section 5. It is noteworthy that the AIC's modifications proposed herein are based on an exact theory. Therefore, the proposed criteria are expected to work even in a high-dimensional case.

2 When Covariance Matrix is Known

2.1 Two-step Monotone Missing Data

Suppose that the observation matrix $\boldsymbol{Y}$ in (1) consists of two-step monotone missing data. Then, without loss of generality, the first n_1 samples are complete, and the remainder n_2 samples have been observed for the first p_1 components only. Specifically, let

$$(\boldsymbol{Y}_{11}, \boldsymbol{Y}_{12}) = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n_1})',$$

where $\boldsymbol{Y}_{11} : n_1 \times p_1, \boldsymbol{Y}_{12} : n_1 \times p_2$ and $p = p_1 + p_2$. Here, $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n_1}$ have a p-dimensional normal distribution with the mean vector $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{X}$ is a $p \times q$ $(p \leq q)$ known design matrix, and $\boldsymbol{\theta}$ is a $q \times 1$ unknown parameter vector. Further, let $\boldsymbol{y}_{21}, \boldsymbol{y}_{22}, \ldots, \boldsymbol{y}_{2n_2}$ have a p_1-dimensional normal distribution with mean vector $\boldsymbol{\mu}_1 = \boldsymbol{X}_1\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}_{11}$, where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ \boldsymbol{\mu}_2')'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$, and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Then, we write $\boldsymbol{Y}_{21} = (\boldsymbol{y}_{21}, \boldsymbol{y}_{22}, \ldots, \boldsymbol{y}_{2n_2})'$ as an $n_2 \times p_1$ matrix, and

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_{11} & \boldsymbol{Y}_{12} \\ \boldsymbol{Y}_{21} & * \end{pmatrix}, \tag{2}$$

where "$*$" indicates a missing part, and we assume that each row of $(\boldsymbol{Y}_{11} \ \boldsymbol{Y}_{12})$ and that of $\boldsymbol{Y}_{21}$ are mutually independent. For observation matrix $\boldsymbol{Y}$ in (2), our model is summarized as follows:

M1$_2$: Each row of $(\boldsymbol{Y}_{11}\ \boldsymbol{Y}_{12})$ and that of $\boldsymbol{Y}_{21}$ are mutually independent and normal with the same mean vectors $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_1$ and covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{11}$, respectively.

M2$_2$:

$$\mathrm{E}[\underset{n_1\times p}{(\boldsymbol{Y}_{11}\ \boldsymbol{Y}_{12})}] = \underset{n_1\times 1}{\mathbf{1}_{n_1}}\underset{1\times q}{\boldsymbol{\theta}'}\underset{q\times p}{\boldsymbol{X}'},\quad \mathrm{E}[\underset{n_2\times p_1}{\boldsymbol{Y}_{21}}] = \underset{n_2\times 1}{\mathbf{1}_{n_2}}\underset{1\times q}{\boldsymbol{\theta}'}\underset{q\times p_1}{\boldsymbol{X}_1'}.$$

Now, we consider the AIC for growth curve model satisfies M1$_2$ and M2$_2$. Following Akaike (1973), the AIC-type-risk may be defined (e.g., Sugiura (1978), Fujikoshi and Satoh (1997)) as

$$\mathrm{RI}_g = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\left[-2\log f(\boldsymbol{Y}^\star;\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Sigma}})\right],$$

where $\boldsymbol{Y}^\star$ is an independent copy of $\boldsymbol{Y}$ and f is a normal density function. We note that the distribution of $\boldsymbol{Y}^\star$ is the same as that of $\boldsymbol{Y}$. Then, the risk can be written as

$$\mathrm{RI}_g = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}[-2\log f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Sigma}})] + \mathrm{CT}_g$$

where

$$\mathrm{CT}_g = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\left[-2\log\frac{f(\boldsymbol{Y}^\star;\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Sigma}})}{f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Sigma}})}\right], \tag{3}$$

and $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Sigma}}$ are maximum likelihood estimators (MLEs) of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, respectively. CT_g is called the correction term when we estimate RI_g by a naive estimator $-2\log f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Sigma}})$. The expectation in (3) should be evaluated under the true density of $\boldsymbol{Y}$. Next, we assume the following:

A1$_2$: The true density of $\boldsymbol{Y}$ is given by $f(\boldsymbol{Y};\boldsymbol{\theta}_0,\boldsymbol{\Sigma}_0)$, where $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_0$ are given values of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, respectively. The true density is included in the model with M1$_2$ and M2$_2$.

A2$_2$: The covariance matrix $\boldsymbol{\Sigma}$ is known.

Under A1$_2$ and A2$_2$, the risk and correction term are written as

$$\mathrm{RI} = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\left[-2\log f(\boldsymbol{Y}^\star;\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})\right],\quad \text{and}\quad \mathrm{CT} = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\left[-2\log\frac{f(\boldsymbol{Y}^\star;\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})}{f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})}\right], \tag{4}$$

respectively. In the following, we write $f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})$ as $\widehat{f}(\boldsymbol{Y})$ and true density $f(\boldsymbol{Y};\boldsymbol{\theta}_0,\boldsymbol{\Sigma}_0)$ as $f(\boldsymbol{Y})$.

Note that $\widehat{f}(\boldsymbol{Y}^\star)$ can be written as

$$\widehat{f}(\boldsymbol{Y}^\star) =\widehat{f}(\boldsymbol{Y}^\star_{11})\widehat{f}(\boldsymbol{Y}^\star_{12}|\boldsymbol{Y}^\star_{11})\widehat{f}(\boldsymbol{Y}^\star_{21}),$$

where

$$\begin{aligned}
\widehat{f}(\boldsymbol{Y}^\star_{11}) =&(2\pi)^{-\frac{1}{2}n_1p_1}|\boldsymbol{\Sigma}_{11}|^{-\frac{1}{2}n_1}\mathrm{etr}\Big\{-\frac{1}{2}(\boldsymbol{Y}^\star_{11}-\boldsymbol{1}_{n_1}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}^\star_{11}-\boldsymbol{1}_{n_1}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')'\Big\},\\
\widehat{f}(\boldsymbol{Y}^\star_{12}|\boldsymbol{Y}^\star_{11}) =&(2\pi)^{-\frac{1}{2}n_1p_2}|\boldsymbol{\Sigma}_{22\cdot 1}|^{-\frac{1}{2}n_1}\\
&\times \mathrm{etr}\Big\{-\frac{1}{2}(\boldsymbol{Y}^\star_{12}-\boldsymbol{1}_{n_1}\widehat{\boldsymbol{\theta}}'(\boldsymbol{X}_2'-\boldsymbol{X}_1'\mathcal{B})-\boldsymbol{Y}^\star_{11}\mathcal{B}_{12})\\
&\times \boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}^\star_{12}-\boldsymbol{1}_{n_1}\widehat{\boldsymbol{\theta}}'(\boldsymbol{X}_2'-\boldsymbol{X}_1'\mathcal{B})-\boldsymbol{Y}^\star_{11}\mathcal{B})'\Big\},\\
\widehat{f}(\boldsymbol{Y}^\star_{21}) =&(2\pi)^{-\frac{1}{2}n_2p_1}|\boldsymbol{\Sigma}_{11}|^{-\frac{1}{2}n_2}\mathrm{etr}\Big\{-\frac{1}{2}(\boldsymbol{Y}^\star_{21}-\boldsymbol{1}_{n_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}^\star_{21}-\boldsymbol{1}_{n_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')'\Big\},
\end{aligned}$$

Here, $\mathcal{B}_{12}=\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. Similarly, $f(\boldsymbol{Y})=f(\boldsymbol{Y}_{11})f(\boldsymbol{Y}_{12}|\boldsymbol{Y}_{11})f(\boldsymbol{Y}_{21})$. Therefore, the correction term in (4) is given by

$$\mathrm{CT}=(\mathrm{A})_1-(\mathrm{B})_1+(\mathrm{A})_2-(\mathrm{B})_2,$$

where

$$\begin{aligned}
(\mathrm{A})_1&=\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}^\star_{(12)1}-\boldsymbol{1}_{N_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')'(\boldsymbol{Y}^\star_{(12)1}-\boldsymbol{1}_{N_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')\Big],\\
(\mathrm{B})_1&=\mathrm{E}_{\boldsymbol{Y}}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_{(12)1}-\boldsymbol{1}_{N_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')'(\boldsymbol{Y}_{(12)1}-\boldsymbol{1}_{N_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')\Big],\\
(\mathrm{A})_2&=\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}^\star_{12}-\boldsymbol{1}_{n_1}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_2'-\boldsymbol{Y}^\star_{11}\mathcal{B}_{12})'(\boldsymbol{Y}^\star_{12}-\boldsymbol{1}_{n_1}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_2'-\boldsymbol{Y}^\star_{11}\mathcal{B}_{12})\Big],\\
(\mathrm{B})_2&=\mathrm{E}_{\boldsymbol{Y}}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{12}-\boldsymbol{1}_{n_1}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_2'-\boldsymbol{Y}_{11}\mathcal{B}_{12})'(\boldsymbol{Y}_{12}-\boldsymbol{1}_{n_1}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_2'-\boldsymbol{Y}_{11}\mathcal{B}_{12})\Big],
\end{aligned}$$

and $N_2=n_1+n_2$, $\underset{q\times p_2}{\widetilde{\boldsymbol{X}}_2'}=\boldsymbol{X}_2'-\boldsymbol{X}_1'\mathcal{B}_{12}$. Then, it is shown that

$$(\mathrm{A})_1-(\mathrm{B})_1=2N_2\,\mathrm{tr}\,\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1',\quad (\mathrm{A})_2-(\mathrm{B})_2=2n_1\,\mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\widetilde{\boldsymbol{X}}_2\boldsymbol{M}^{-1}\widetilde{\boldsymbol{X}}_2',$$

and $\mathrm{CT}=2q$. For the details, see Appendix. Thus, we can obtain the following result:

Theorem 1 *Suppose that a two-step monotone data matrix* $\boldsymbol{Y}$ *in* (2) *satisfies* $\mathrm{M1}_2$ *and* $\mathrm{M2}_2$. *Then, under assumptions* $\mathrm{A1}_2$ *and* $\mathrm{A2}_2$*, the correction term in* (4) *is given by*

$$\mathrm{CT}\equiv\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\left[-2\log\frac{f(\boldsymbol{Y}^\star;\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})}{f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})}\right]=2q.$$

The number $2q$ equals to $2\times$ (the number of unknown parameters) under $\mathrm{A1}_2$ and $\mathrm{A2}_2$. We can express a usual AIC as

$$\mathrm{AIC} = -2\log f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma}) + 2q,$$

and from Theorem 1, the AIC is an exact unbiased estimator of RI.

2.2 k-step Monotone Missing Data

In this section, we consider the case of k-step monotone missing data whose observation matrix $\boldsymbol{Y}$ is of the form

$$\boldsymbol{Y} = \begin{pmatrix} \overbrace{\boldsymbol{Y}_{11}}^{p_1} & \overbrace{\boldsymbol{Y}_{12}}^{p_2} & \cdots & \overbrace{\boldsymbol{Y}_{1,k-i+1}}^{p_{k-i+1}} & \cdots & \overbrace{\boldsymbol{Y}_{1,k-1}}^{p_{k-1}} & \overbrace{\boldsymbol{Y}_{1k}}^{p_k} \\ \boldsymbol{Y}_{21} & \boldsymbol{Y}_{22} & \cdots & \boldsymbol{Y}_{2k-i+1} & \cdots & \boldsymbol{Y}_{2,k-1} & * \\ \vdots & \vdots & & \vdots & & & \vdots \\ \boldsymbol{Y}_{i1} & \boldsymbol{Y}_{i,2} & \cdots & \boldsymbol{Y}_{i,k-i+1} & * & \cdots & * \\ \vdots & \vdots & & & & & \vdots \\ \boldsymbol{Y}_{k-1,1} & \boldsymbol{Y}_{k-1,2} & * & \cdots & \cdots & \cdots & * \\ \boldsymbol{Y}_{k1} & * & \cdots & \cdots & \cdots & \cdots & * \end{pmatrix} \begin{matrix} \}n_1 \\ \}n_2 \\ \\ \}n_i \\ \\ \}n_{k-1} \\ \}n_k \end{matrix}, \qquad (5)$$

where "$*$" indicates a missing part. Further, let

$$\underset{n_i\times p_{(12\ldots,k-i+1)}}{\boldsymbol{Y}_{i(12\ldots,k-i+1)}} = (\boldsymbol{Y}_{i1}\ \boldsymbol{Y}_{i2}\cdots\boldsymbol{Y}_{i,k-i+1}), \quad \underset{p_{(12\ldots,k-i+1)}\times 1}{\boldsymbol{\mu}_{(12\ldots,k-i+1)}} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_{k-i+1} \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \\ \\ \}p_{k-i+1} \end{matrix},$$

$$\underset{p_{(12,\ldots,k-i+1)}\times p_{(12,\ldots,k-i+1)}}{\boldsymbol{\Sigma}_{(12,\ldots,k-i+1)(12,\ldots,k-i+1)}} = \begin{pmatrix} \overbrace{\boldsymbol{\Sigma}_{11}}^{p_1} & \overbrace{\boldsymbol{\Sigma}_{12}}^{p_2} & \cdots & \overbrace{\boldsymbol{\Sigma}_{1,k-i+1}}^{p_{k-i+1}} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2,k-i+1} \\ \vdots & & & \vdots \\ \boldsymbol{\Sigma}_{k-i+1,1} & \boldsymbol{\Sigma}_{k-i+1,2} & \cdots & \boldsymbol{\Sigma}_{k-i+1,k-i+1} \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \\ \\ \}p_{k-i+1} \end{matrix},$$

$$\underset{p_{(12\ldots,k-i+1)}\times q}{\boldsymbol{X}_{(12\ldots,k-i+1)}} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_{k-i+1} \end{pmatrix} \begin{matrix} \}p_1 \\ \}p_2 \\ \\ \}p_{k-i+1} \end{matrix}, \quad p_{(12\ldots,k-i+1)} = \sum_{j=1}^{k-i+1} p_j, \quad i = 1, 2, \ldots, k,$$

$$\boldsymbol{\mu}_{(12\ldots k)} = \boldsymbol{\mu}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{(12\ldots k)(12\ldots k)}, \quad \underset{p\times q}{\boldsymbol{X}} = \boldsymbol{X}_{(12\ldots k)}, \quad p = p_{(12\ldots k)},$$

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_q)'.$$

For observation matrix $\boldsymbol{Y}$ in (5), our model is summarized as follows:

M1$_k$: Each row of $\boldsymbol{Y}_{i(12\ldots,k-i+1)}$ is mutually independent and normal with the same mean vectors $\boldsymbol{\mu}_{(12\ldots,k-i+1)}$ and covariance matrices $\boldsymbol{\Sigma}_{(12,\ldots,k-i+1)(12,\ldots,k-i+1)}$.

M2$_k$:

$$\mathrm{E}[\boldsymbol{Y}_{i(12\ldots,k-i+1)}] = \mathbf{1}_{n_i}\boldsymbol{\mu}'_{(12\ldots,k-i+1)},\ i = 1, 2, \ldots, k, \tag{6}$$

where $\boldsymbol{\mu}_{(12\ldots,k-i+1)} = \boldsymbol{X}_{(12\ldots,k-i+1)}\boldsymbol{\theta}$.

Similar to Section 2.1, we assume the following:

A1$_k$: The true density of $\boldsymbol{Y}$ is given by $f(\boldsymbol{Y}; \boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_0$ are given values of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, respectively. The true density is included in the model with M1$_k$ and M2$_k$.

A2$_k$: The covariance matrix $\boldsymbol{\Sigma}$ is known.

Under A1$_k$ and A2$_k$, the risk and correction terms for the k-step monotone missing data are written as

$$\mathrm{RI} = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\left[-2\log f(\boldsymbol{Y}^\star; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})\right] \quad \text{and} \quad \mathrm{CT} = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\left[-2\log \frac{f(\boldsymbol{Y}^\star; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})}{f(\boldsymbol{Y}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})}\right], \tag{7}$$

respectively. We note that

$$\begin{aligned}
f(\boldsymbol{Y}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{k} f(\boldsymbol{Y}_{i(12\ldots,k-i+1)}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \\
&= \prod_{i=1}^{k} f(\boldsymbol{Y}_{i1}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \prod_{i=1}^{k} \prod_{j=2}^{k-i+1} f(\boldsymbol{Y}_{ij} | \boldsymbol{Y}_{i(12\ldots,j-1)}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \\
&= f(\boldsymbol{Y}_{(12\ldots k)1}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \prod_{i=1}^{k-1} f(\boldsymbol{Y}_{(12\ldots i)k-i+1} | \boldsymbol{Y}_{(12\ldots i)(12\ldots,k-i)}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \\
&= f(\boldsymbol{Y}_{(12\ldots k)1}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \prod_{j=2}^{k} f(\boldsymbol{Y}_{(12\ldots,k-j+1)j} | \boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}),
\end{aligned}$$

where

$$
\begin{aligned}
&f(\boldsymbol{Y}_{(12\ldots k)1};\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})\\
&=(2\pi)^{-\frac{1}{2}N_k p_1}|\boldsymbol{\Sigma}_{11}|^{-\frac{1}{2}N_k}\,\mathrm{etr}\Big\{-\frac{1}{2}(\boldsymbol{Y}_{(12\ldots k)1}-\mathbf{1}_{N_k}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_{(12\ldots k)1}-\mathbf{1}_{N_k}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')'\Big\},\\
&f(\boldsymbol{Y}_{(12\ldots i)j}|\boldsymbol{Y}_{(12\ldots i)(12\ldots,j-1)};\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})\\
&=(2\pi)^{-\frac{1}{2}N_i p_j}|\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}|^{-\frac{1}{2}N_i}\,\mathrm{etr}\Big[-\frac{1}{2}\Big\{\boldsymbol{Y}_{(12\ldots i)j}-\mathbf{1}_{N_i}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_j'-\boldsymbol{Y}_{(12\ldots i)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j}\Big\}\\
&\quad\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1}\Big\{\boldsymbol{Y}_{(12\ldots i)j}-\mathbf{1}_{N_i}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_j'-\boldsymbol{Y}_{(12\ldots i)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j}\Big\}'\Big],\\
&\hspace{20em} i=k-j+1,\ j=2,3,\ldots,k,
\end{aligned}
$$

and

$$
\begin{aligned}
&N_i=\sum_{j=1}^{i}n_j,\ i=k-j+1,\quad \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}=\boldsymbol{\Sigma}_{jj}-\boldsymbol{\Sigma}_{(12\ldots,j-1)j}\boldsymbol{\Sigma}_{(12\ldots,j-1)(12\ldots,j-1)}^{-1}\boldsymbol{\Sigma}_{(12\ldots,j-1)j}\\
&\widetilde{\boldsymbol{X}}_j'=\boldsymbol{X}_j'-\boldsymbol{X}_{(12\ldots,j-1)}'\mathcal{B}_{(12\ldots,j-1)j},\quad \mathcal{B}_{(12\ldots,j-1)j}=\boldsymbol{\Sigma}_{(12\ldots,j-1)(12\ldots,j-1)}^{-1}\boldsymbol{\Sigma}_{(12\ldots,j-1)j},\\
&\underset{p_{(12\ldots,j-1)}\times p_j}{\boldsymbol{\Sigma}_{(12\ldots,j-1)j}}=\begin{pmatrix}\boldsymbol{\Sigma}_{1j}\\ \boldsymbol{\Sigma}_{2j}\\ \vdots\\ \boldsymbol{\Sigma}_{j-1,j}\end{pmatrix},\ j=2,3,\ldots,k.
\end{aligned}
$$

Similarly, $f(\boldsymbol{Y}^{\star};\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})=\prod_{i=1}^{k}f(\boldsymbol{Y}^{\star}_{i(12\ldots,k-i+1)};\widehat{\boldsymbol{\theta}},\boldsymbol{\Sigma})$. Further, we note that $\widehat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is known (see Yagi, Seo and Fujikoshi (2021)).

Therefore, the correction term in (7) is given by

$$
\mathrm{CT}=\sum_{j=1}^{k}\{(\mathrm{A})_j-(\mathrm{B})_j\},
$$

where

$$
\begin{aligned}
(\mathrm{A})_1=&\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^{\star}}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}^{\star}_{(12\ldots k)1}-\mathbf{1}_{N_k}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')'(\boldsymbol{Y}^{\star}_{(12\ldots k)1}-\mathbf{1}_{N_k}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')\Big],\\
(\mathrm{B})_1=&\mathrm{E}_{\boldsymbol{Y}}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_{(12\ldots k)1}-\mathbf{1}_{N_k}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')'(\boldsymbol{Y}_{(12\ldots k)1}-\mathbf{1}_{N_k}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')\Big],\\
(\mathrm{A})_j=&\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^{\star}}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1}(\boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)j}-\mathbf{1}_{N_{k-j+1}}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_j'-\boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j})'\\
&\quad(\boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)j}-\mathbf{1}_{N_{k-j+1}}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_j'-\boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j})\Big],\\
(\mathrm{B})_j=&\mathrm{E}_{\boldsymbol{Y}}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1}(\boldsymbol{Y}_{(12\ldots,k-j+1)j}-\mathbf{1}_{N_{k-j+1}}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_j'-\boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j})'\\
&\quad(\boldsymbol{Y}_{(12\ldots,k-j+1)j}-\mathbf{1}_{N_{k-j+1}}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_j'-\boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j})\Big],\ j=2,3,\ldots,k.
\end{aligned}
$$

Then, it is shown that

$$(A)_1 - (B)_1 = 2N_k \operatorname{tr} \boldsymbol{A}_1 \boldsymbol{M}^{-1},$$
$$(A)_j - (B)_j = 2N_{k-j+1} \operatorname{tr} \boldsymbol{A}_j \boldsymbol{M}^{-1}, \ j = 2, 3, \ldots, k,$$

where

$$\boldsymbol{A}_1 = \boldsymbol{X}_1' \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{X}_1, \quad \boldsymbol{A}_j = \widetilde{\boldsymbol{X}}_j' \boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1} \widetilde{\boldsymbol{X}}_j, \quad j = 2, 3, \ldots, k, \quad \boldsymbol{M} = \sum_{j=1}^{k} N_{k-j+1} \boldsymbol{A}_j.$$

Hence, we can obtain the following result:

Theorem 2 *Suppose that a k-step monotone data matrix $\boldsymbol{Y}$ in (5) satisfies* $\mathrm{M1}_k$ *and* $\mathrm{M2}_k$*. Then, under assumptions* $\mathrm{A1}_k$ *and* $\mathrm{A2}_k$*, the correction term in* (7) *is given by*

$$\mathrm{CT} \equiv \mathrm{E}_{\boldsymbol{Y}} \mathrm{E}_{\boldsymbol{Y}^\star} \left[-2 \log \frac{f(\boldsymbol{Y}^\star; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})}{f(\boldsymbol{Y}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})} \right] = 2q.$$

For the proof of Theorem 2, refer to Appendix. From Theorem 2, we can express a usual AIC as

$$\mathrm{AIC} = -2 \log f(\boldsymbol{Y}; \widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) + 2q, \tag{8}$$

which is an exact unbiased estimator of RI, where $\boldsymbol{Y}$ is given by (5).

3 When all Parameters are Unknown

In this section, we consider the construction of a naive AIC when all the parameters are unknown. Suppose that a k-step monotone data $\boldsymbol{Y}$ is given as in (5), and let

$$\begin{aligned} \boldsymbol{Y}_i &= (\boldsymbol{Y}_{i1}, \boldsymbol{Y}_{i2}, \ldots, \boldsymbol{Y}_{i,k-i+1}) \\ &= (\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, \ldots, \boldsymbol{y}_{in_i})', \quad i = 1, 2, \ldots, k. \end{aligned} \tag{9}$$

Then,

$$\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, \ldots, \boldsymbol{y}_{in_i} \overset{i.i.d.}{\sim} N_{r_i}(\boldsymbol{X}^{(i)} \boldsymbol{\theta}, \boldsymbol{\Sigma}^{(i)}), \quad i = 1, 2, \ldots, k, \tag{10}$$

where $r_i = p_1+p_2+\cdots+p_{k-i+1}$, $i = 1, 2, \ldots, k$, $\boldsymbol{X}^{(i)}$ is the upper $r_i \times q$ matrix of $\boldsymbol{X}$, and $\boldsymbol{\Sigma}^{(i)}$ is the upper left $r_i \times r_i$ matrix of $\boldsymbol{\Sigma}$. The (-2) loglikelihood of observed data $\boldsymbol{Y} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_k\}$ can be expressed as

$$\begin{aligned}\ell(\boldsymbol{\theta}, \boldsymbol{\Sigma}; \boldsymbol{Y}) &= \sum_{i=1}^{k} \left\{ n_i \log |\boldsymbol{\Sigma}^{(i)}| + n_i r_i \log(2\pi) \right\} + \sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathrm{tr}\, {\boldsymbol{\Sigma}^{(i)}}^{-1} (\boldsymbol{y}_{ij} - \boldsymbol{X}^{(i)}\boldsymbol{\theta})(\boldsymbol{y}_{ij} - \boldsymbol{X}^{(i)}\boldsymbol{\theta})', \\ &= \ell_1(\boldsymbol{\theta}, \boldsymbol{\Sigma}; \boldsymbol{Y}_1) + \ell_2(\boldsymbol{\theta}, \boldsymbol{\Sigma}; \boldsymbol{Y}_d).\end{aligned}$$

Here,

$$\ell_1(\boldsymbol{\theta}, \boldsymbol{\Sigma}; \boldsymbol{Y}_1) = n_1 \log |\boldsymbol{\Sigma}| + n_1 p \log(2\pi) + \sum_{j=1}^{n_1} \mathrm{tr}\, \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_{1j} - \boldsymbol{X}\boldsymbol{\theta})(\boldsymbol{y}_{1j} - \boldsymbol{X}\boldsymbol{\theta})',$$

$$\ell_2(\boldsymbol{\theta}, \boldsymbol{\Sigma}; \boldsymbol{Y}_d) = \sum_{i=2}^{k} \left\{ n_i \log |\boldsymbol{\Sigma}^{(i)}| + n_i r_i \log(2\pi) \right\} + \sum_{i=2}^{k}\sum_{j=1}^{n_i} \mathrm{tr}\, {\boldsymbol{\Sigma}^{(i)}}^{-1} (\boldsymbol{y}_{ij} - \boldsymbol{X}^{(i)}\boldsymbol{\theta})(\boldsymbol{y}_{ij} - \boldsymbol{X}^{(i)}\boldsymbol{\theta})',$$

where $\boldsymbol{Y}_d = \{\boldsymbol{Y}_2, \boldsymbol{Y}_3, \ldots, \boldsymbol{Y}_k\}$.

When n_1 is large and $(n_2 + n_3 + \cdots + n_k)/n_1$ is small, we shall be able to obtain reasonable estimators for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ by considering the MLEs based on the likelihood of $\boldsymbol{Y}_1$ only. From the result on the MLEs for complete data, such quasi MLEs are expressed (see, e.g. Fujikoshi, Ulyanov, and Shimizu (2010)) as

$$\widetilde{\boldsymbol{\theta}} = (\boldsymbol{X}'\widetilde{\boldsymbol{S}}_1^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widetilde{\boldsymbol{S}}_1^{-1}\bar{\boldsymbol{y}}_1, \quad \widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{S}}_1 + (\bar{\boldsymbol{y}}_1 - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})(\bar{\boldsymbol{y}}_1 - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})',$$

where

$$\overline{\boldsymbol{y}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} \boldsymbol{y}_{1j}, \quad \widetilde{\boldsymbol{S}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} (\boldsymbol{y}_{1j} - \bar{\boldsymbol{y}}_1)(\boldsymbol{y}_{1j} - \bar{\boldsymbol{y}}_1)'.$$

Let $\boldsymbol{Y}^\star$ be an independent copy of $\boldsymbol{Y}$, and decompose $\boldsymbol{Y}^\star$ as in (9). Then,

$$\begin{aligned}\boldsymbol{Y}_i^\star &= (\boldsymbol{Y}_{i1}^\star, \boldsymbol{Y}_{i2}^\star, \ldots, \boldsymbol{Y}_{i,k-i+1}^\star) \\ &= (\boldsymbol{y}_{i1}^\star, \boldsymbol{y}_{i2}^\star, \ldots, \boldsymbol{y}_{in_i}^\star)', \quad i = 1, 2, \ldots, k,\end{aligned}$$

and

$$\boldsymbol{y}_{i1}^\star, \boldsymbol{y}_{i2}^\star, \ldots, \boldsymbol{y}_{in_i}^\star \overset{i.i.d.}{\sim} N_{r_i}(\boldsymbol{X}^{(i)}\boldsymbol{\theta}, \boldsymbol{\Sigma}^{(i)}), \quad i = 1, 2, \ldots, k,$$

Using observed data set $\boldsymbol{Y}$ and quasi MLEs $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\Sigma}}$, we consider the following AIC-type risk:

$$\begin{aligned}\widetilde{\mathrm{RI}} &= \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}[\ell(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Sigma}}; \boldsymbol{Y}^\star)] \\ &= \mathrm{E}_{\boldsymbol{Y}}[\ell(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Sigma}}; \boldsymbol{Y})] + \widetilde{\mathrm{CT}},\end{aligned} \tag{11}$$

where

$$\widetilde{\mathrm{CT}} = \mathrm{E}_{\boldsymbol{Y}} \mathrm{E}_{\boldsymbol{Y}^\star}[\ell(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Sigma}}; \boldsymbol{Y}^\star) - \ell(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Sigma}}; \boldsymbol{Y})].$$

Let correction term $\widetilde{\mathrm{CT}}$ decompose as

$$\widetilde{\mathrm{CT}} = \widetilde{\mathrm{CT}}_1 + \widetilde{\mathrm{CT}}_2,$$

where $\widetilde{\mathrm{CT}}_i = \mathrm{E}_{\boldsymbol{Y}} \mathrm{E}_{\boldsymbol{Y}^\star}[\ell_i(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Sigma}}; \boldsymbol{Y}^\star) - \ell_i(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Sigma}}; \boldsymbol{Y})], \quad i = 1, 2$. It is clear that

$$\widetilde{\mathrm{CT}}_1 = \sum_{j=1}^{n_1} \mathrm{E}_{\boldsymbol{Y}} \mathrm{E}_{\boldsymbol{Y}^\star} \left[\mathrm{tr}\, \widetilde{\boldsymbol{\Sigma}}^{-1} \left\{ (\boldsymbol{y}_{1j}^\star - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})(\boldsymbol{y}_{1j}^\star - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})' - (\boldsymbol{y}_{1j} - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})(\boldsymbol{y}_{1j} - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})' \right\} \right].$$

Considering expectation with respect to $\boldsymbol{y}_{1j}^\star$ and using (17), we note that

$$\widetilde{\mathrm{CT}}_1 = \mathrm{E}_{\boldsymbol{Y}} \left[\mathrm{tr}\, \widetilde{\boldsymbol{\Sigma}}^{-1} \left\{ n_1 \boldsymbol{\Sigma} + n_1 \left\{ \boldsymbol{X}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\} \left\{ \boldsymbol{X}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\}' - n_1 \widetilde{\boldsymbol{\Sigma}} \right\} \right]$$

Further, the above $\widetilde{\mathrm{CT}}_1$ can be reduced to the following (see, e.g., Fujikoshi et al. (2010, pp.371-372)):

$$\begin{aligned} \widetilde{\mathrm{CT}}_1 = & -n_1 p + \frac{(p-q)n_1^2}{n_1 - p + q - 1} + \frac{q n_1 (n_1 - 2)(n_1 + 1)}{(n_1 - p - 2)(n_1 - p + q - 2)} \\ \equiv & \widetilde{C}_0. \end{aligned} \tag{12}$$

Second correction term $\widetilde{\mathrm{CT}}_2$ is expressed as

$$\widetilde{\mathrm{CT}}_2 = \sum_{i=2}^{k} \sum_{j=1}^{n_i} \mathrm{E}_{\boldsymbol{Y}} \mathrm{E}_{\boldsymbol{Y}^\star} \left[\mathrm{tr}\, \widetilde{\boldsymbol{\Sigma}}^{(i)^{-1}} \left\{ (\boldsymbol{y}_{ij}^\star - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})(\boldsymbol{y}_{ij}^\star - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})' - (\boldsymbol{y}_{ij} - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})(\boldsymbol{y}_{ij} - \boldsymbol{X}\widetilde{\boldsymbol{\theta}})' \right\} \right].$$

Here, note that if $i \geq 2$, $\{\boldsymbol{y}_{ij}^\star\}$ and $\{\boldsymbol{y}_{ij}\}$ are independent of $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\Sigma}}$, implying that $\widetilde{\mathrm{CT}}_2 = 0$. Thus, we can obtain the following result:

Theorem 3 *For a general k-step monotone data* $\boldsymbol{Y}$ *in* (5)*, assume a growth curve model* (10)*. Let* $\widetilde{\mathrm{RI}}$ *and* $\widetilde{\mathrm{CT}}$ *in* (11) *be the AIC-type risk and correction terms based on the quasi MLEs in* (17)*. Let*

$$\widetilde{\mathrm{AIC}} = \ell(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Sigma}}; \boldsymbol{Y}) + \widetilde{\mathrm{C}}_0, \tag{13}$$

where $\widetilde{\mathrm{C}}_0$ *is given by* (12)*. Assume that the true model is included in a growth curve model* (10)*. Then,* $\widetilde{\mathrm{AIC}}$ *is an exact unbiased estimator for* $\widetilde{\mathrm{RI}}$.

When n_1 is large, correction term $\widetilde{\mathrm{C}}_0$ can be expanded as

$$\widetilde{\mathrm{C}}_0 = 2\left\{q + \frac{1}{2}p(p+1)\right\} + O(n_1^{-1}),$$

whose first term is twice the number of unknown parameters.

4 Numerical Example

In this section, we consider the data on the ramus heights of 20 boys, presented in Table 2 of Elston and Grizzle (1962), to illustrate the results of this article. The ramus heights have been measured in mm for each boy at 8, $8\frac{1}{2}$, 9, and $9\frac{1}{2}$ years of age. The data are expressed as $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ with $n = 20$. Sample mean vector $\overline{\boldsymbol{y}}$ and sample covariance matrix $\boldsymbol{S}$ are given as follows:

$$\overline{\boldsymbol{y}} = \begin{pmatrix} 48.66 \\ 49.63 \\ 50.57 \\ 51.45 \end{pmatrix}, \quad \boldsymbol{S} = \begin{pmatrix} 6.330 & 6.189 & 5.777 & 5.548 \\ 6.189 & 6.449 & 6.153 & 5.923 \\ 5.777 & 6.153 & 6.918 & 6.946 \\ 5.548 & 5.923 & 6.946 & 7.465 \end{pmatrix}.$$

A growth curve model is expressed as

$$\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \; i = 1, 2, \ldots, n, \tag{14}$$

where $\boldsymbol{\mu} = \boldsymbol{X\theta}$, $\boldsymbol{X}$ is a given $p \times q$ $(\leq p)$ matrix, and $\boldsymbol{\theta}$ is an unknown $q \times 1$ vector. The case of ramus height data is a special case with $p = 4$, $n = 20$, and

$$\boldsymbol{X} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_4 \end{pmatrix}, \; \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \quad \text{or} \quad \boldsymbol{X} = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_4 & t_4^2 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix},$$

with $t_1 = 8$, $t_2 = 8\frac{1}{2}$, $t_3 = 9$, $t_4 = 9\frac{1}{2}$. It is important to determine the degree of polynomial growth curve model (precisely $q = 2$ or 3).

Now, to study the case of monotone missing data, let us assume that by discarding some data, the data will be missing completely at random and consist of a three-step monotone pattern, as follows:

$$\boldsymbol{Y}_{1(123)} = \begin{pmatrix} 47.8 & 48.8 & 49.0 & 49.7 \\ 46.4 & 47.3 & 47.7 & 48.4 \\ 46.3 & 46.8 & 47.8 & 48.5 \\ 45.1 & 45.3 & 46.1 & 47.2 \\ 47.6 & 48.5 & 48.9 & 49.3 \\ 52.5 & 53.2 & 53.3 & 53.7 \\ 51.2 & 53.0 & 54.3 & 54.5 \\ 49.8 & 50.0 & 50.3 & 52.7 \\ 48.1 & 50.8 & 52.3 & 54.4 \\ 45.0 & 47.0 & 47.3 & 48.3 \\ 51.2 & 51.4 & 51.6 & 51.9 \\ 48.5 & 49.2 & 53.0 & 55.5 \\ 52.1 & 52.8 & 53.7 & 55.0 \\ 48.2 & 48.9 & 49.3 & 49.8 \end{pmatrix}, \quad \boldsymbol{Y}_{2(12)} = \begin{pmatrix} 49.6 & 50.4 & 51.2 \\ 50.7 & 51.7 & 52.7 \\ 47.2 & 47.7 & 48.4 \end{pmatrix}, \quad \boldsymbol{Y}_{31} = \begin{pmatrix} 53.3 & 54.6 \\ 46.2 & 47.5 \\ 46.3 & 47.6 \end{pmatrix}.$$

For this three-step monotone missing data, we obtain the MLEs of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ for $q = 2$ and $q = 3$ using the methods described by Yagi et al. (2021), as follows:

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} 33.20 \\ 1.937 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 6.015 & 5.881 & 5.889 & 5.732 \\ 5.881 & 6.128 & 6.218 & 6.097 \\ 5.889 & 6.218 & 7.032 & 7.242 \\ 5.732 & 6.097 & 7.242 & 7.985 \end{pmatrix} \quad (q = 2), \tag{15}$$

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} 39.75 \\ 0.4005 \\ 0.08967 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 6.015 & 5.880 & 5.887 & 5.739 \\ 5.880 & 6.127 & 6.214 & 6.102 \\ 5.887 & 6.214 & 7.030 & 7.257 \\ 5.739 & 6.102 & 7.257 & 8.014 \end{pmatrix} \quad (q = 3). \tag{16}$$

We are interested in examining whether the growth is linear ($q = 2$) or quadratic ($q = 3$). Related to this problem, consider two models $\mathcal{M}_1 : q = 2$ and $\mathcal{M}_2 : q = 3$.

First consider the case of $\boldsymbol{\Sigma}$ being known. Assume that $\boldsymbol{\Sigma}$ is equal to $\widehat{\boldsymbol{\Sigma}}$ in (15) or (16). Then, for two models $\mathcal{M}_1$ and $\mathcal{M}_2$, the AIC in (8) is presented in Table 1. Table 1 shows that $\mathcal{M}_1$ is better than $\mathcal{M}_2$.

Next, consider the case of $\boldsymbol{\Sigma}$ being unknown. For this case, we proposed a conventional AIC, $\widetilde{\mathrm{AIC}}$, which is presented in (13). This criterion is given as follows. Table 1 suggests that $\mathcal{M}_1$ is also better than $\mathcal{M}_2$.

Table 1. AIC and $\widetilde{\mathrm{AIC}}$ for the three-step monotone missing data.

	AIC	$\widetilde{\mathrm{AIC}}$
$\mathcal{M}_1$	202.4	242.1
$\mathcal{M}_2$	204.2	245.6

To examine an effect of missing data on the model selection criterion, we consider a formal AIC and a corrected AIC, which are denoted by AIC_0 and CAIC_0, respectively, based on the density $f(\boldsymbol{Y};\boldsymbol{\theta},\boldsymbol{\Sigma})$ of $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)'$ in (14). Then, we have

$$\mathrm{AIC}_0 = -2\log f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Sigma}}) + 2\left\{q + \frac{1}{2}p(p+1)\right\},$$

where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Sigma}}$ are the MLEs of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, satisfying

$$\max_{\boldsymbol{\theta},\boldsymbol{\Sigma}} f(\boldsymbol{Y};\boldsymbol{\theta},\boldsymbol{\Sigma}) = f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Sigma}}).$$

The MLEs are given (e.g., Fujikoshi et al. (2010)) as follows:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{S}^{-1}\bar{\boldsymbol{y}}, \quad \widehat{\boldsymbol{\Sigma}} = \boldsymbol{S} + (\bar{\boldsymbol{y}} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})(\bar{\boldsymbol{y}} - \boldsymbol{X}\widehat{\boldsymbol{\theta}})'. \tag{17}$$

Its corrected AIC is given (e.g., Fujikoshi et al. (2010)) as follows:

$$\mathrm{CAIC}_0 = -2\log f(\boldsymbol{Y};\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Sigma}}) + C_0,$$

where

$$C_0 = -np + \frac{(p-q)n^2}{n-p+q-1} + \frac{qn(n-2)(n+1)}{(n-p-2)(n-p+q-2)}.$$

These criteria are presented in Table 2. Both criteria show that $\mathcal{M}_1$ is better than $\mathcal{M}_2$, similar to the case of missing data. Although not very significant, the difference between CAIC_0s is larger than that between AIC_0s.

Table 2. AIC and corrected AIC for the complete data.

	AIC_0	CAIC_0
$\mathcal{M}_1$	248.6	259.1
$\mathcal{M}_2$	250.4	261.9

5 Conclusion

In this article, we considered an AIC for selecting a set of explanatory variables or a degree of polynomial regression in a growth curve model when the dataset has a monotone pattern of missing observations. We noted that a formal AIC is an exact unbiased estimator of the

AIC-type risk function defined by the expected log-predictive likelihood when the covariance matrix is known. When the covariance matrix is unknown, we considered a quasi AIC-type risk based on the estimators obtained from a set of complete data only and not all of the dataset. We proposed a quasi AIC, which is an exact unbiased estimator for the quasi AIC-type risk. Through a numerical experiment, it was shown that the proposed AICs work well for monotone missing data.

When all the parameters are unknown, we proposed a conventional $\widetilde{\mathrm{AIC}}$ in (13). It will be interesting to consider a model selection criterion as an estimator of the AIC-type risk based on the exact MLEs. We considered the case of one group. It is also important to extend to the case of several groups of growth data. Further, it is interesting to examine similar modifications for other model selection criteria such as $\mathrm{C_p}$.

Acknowledgment

The first and second authors' research is partly supported by a Grant-in-Aid for Young Scientists (JSPS KAKENHI Grant Number JP19K20225) and a Grant-in-Aid for Scientific Research (C) (JSPS KAKENHI Grant Number JP17K00058), respectively. The third author's research is partially supported by the Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (C) (JSPS KAKENHI Grant Number JP16K00047), 2016-2018.

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

[2] Cavanaugh, J. E. and Shumway, R. H. (1998). An Akaike information criterion for model selection in the presence of incomplete data. *Journal of Statistical Planning and Inference*, **67**, 45–65.

[3] Elston, R. C. and Grizzle, J. E. (1962). Estimation of time-response curves and their confidence bands. *Biometrics*, **18**, 148-159.

[4] Fujikoshi, Y., Enomoto, R., and Sakurai, T. (2013). High-dimensional AIC in the growth curve model. *Journal of Multivariate Analysis*, **122**, 239–250.

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

[6] Fujikoshi, Y., Ulyanov, V. V., and Shimizu, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, N.J.

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

[8] Rao, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika*, **46**, 49–58.

[9] Satoh, K., Kobayashi, M., and Fujikoshi, Y. (1997). Variable selection for the growth curve model. *Journal of Multivariate Analysis*, **60**, 277–292.

[10] Shimodaira, H. (1994). A new criterion for selecting models from partially observed data. *In: Cheeseman, P., Oldford, R.W. (Eds.), Selecting Models from Data: Artificial Intelligence and Statistics IV, Lecture Notes in Statistics, Springer, New York*, **89**, 21–29.

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

[12] Yagi, A., Seo, T., and Fujikoshi, Y. (2021). Maximum likelihood estimators in growth curve model with monotone missing data. to appear in *American Journal of Mathematical and Management Sciences*.

Appendix

Proof of Theorem 1

First, we calculate $(\mathrm{A})_1$. Because $(\boldsymbol{Y}^\star_{(12)1}-\mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1')$ and $(\mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1'-\mathbf{1}_{N_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')$ are independent, $(\mathrm{A})_1$ can be written as

$$\begin{aligned}(\mathrm{A})_1 =&N_2p_1+\operatorname{tr}\boldsymbol{\Sigma}_{11}^{-1}\mathrm{E}_{\boldsymbol{Y}}\left[(\mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1'-\mathbf{1}_{N_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')'(\mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1'-\mathbf{1}_{N_2}\widehat{\boldsymbol{\theta}}'\boldsymbol{X}_1')\right]\\ =&N_2p_1+N_2\operatorname{tr}\boldsymbol{A}_1\boldsymbol{M}^{-1},\end{aligned}$$

where

$$\boldsymbol{M}=N_2A_1+N_1A_2,\quad \boldsymbol{A}_1=\boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1,\quad \boldsymbol{A}_2=\widetilde{\boldsymbol{X}}_2'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\widetilde{\boldsymbol{X}}_2.$$

We note that $\mathrm{E}\big[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\big]=\mathrm{Var}(\widehat{\boldsymbol{\theta}})=\boldsymbol{M}^{-1}$, $\mathrm{E}_{\boldsymbol{Y}^\star}\Big[(\boldsymbol{Y}^\star_{(12)1}-\mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1')'(\boldsymbol{Y}^\star_{(12)1}-\mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1')\Big]=N_2\boldsymbol{\Sigma}_{11}$, and $\mathrm{E}_{\boldsymbol{Y}^\star}\Big[\boldsymbol{Y}^\star_{(12)1}-\mathbf{1}_{N_2}\boldsymbol{\theta}'\boldsymbol{X}_1'\Big]=\mathbf{0}$. Similar to the calculation of $(\mathrm{A})_1$, we have

$$(\mathrm{B})_1=N_2p_1+N_2\operatorname{tr}\boldsymbol{A}_1\boldsymbol{M}^{-1}-2\mathrm{E}_{\boldsymbol{Y}}\left[\operatorname{tr}\boldsymbol{a}_1(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\right],$$

where

$$\begin{aligned}&\widehat{\boldsymbol{\theta}}=\boldsymbol{M}^{-1}(\boldsymbol{a}_1+\boldsymbol{a}_2),\quad \boldsymbol{a}_1=\boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}'_{(12)1}\mathbf{1}_{N_2}\sim N_q(N_2\boldsymbol{A}_1\boldsymbol{\theta},N_2\boldsymbol{A}_1),\\ &\boldsymbol{a}_2=\widetilde{\boldsymbol{X}}_2'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{12}-\boldsymbol{Y}_{11}\mathcal{B}_{12})'\mathbf{1}_{N_1}\sim N_q(N_1\boldsymbol{A}_2\boldsymbol{\theta},N_1\boldsymbol{A}_2),\end{aligned}$$

and $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are independent (for the details, refer to Yagi et al. (2021)). Because $\mathrm{E}[\boldsymbol{a}_1\boldsymbol{a}_1']=N_2\boldsymbol{A}_1+N_2^2\boldsymbol{A}_1\boldsymbol{\theta}\boldsymbol{\theta}'\boldsymbol{A}_1$ and $\mathrm{E}[\boldsymbol{a}_1\boldsymbol{a}_2']=N_1N_2\boldsymbol{A}_1\boldsymbol{\theta}\boldsymbol{\theta}'\boldsymbol{A}_2$, it can be inferred that

$$\mathrm{E}_{\boldsymbol{Y}}\left[\operatorname{tr}\boldsymbol{a}_1(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\right]=N_2\operatorname{tr}\boldsymbol{A}_1\boldsymbol{M}^{-1}.$$

Therefore, we can obtain

$$(\mathrm{A})_1-(\mathrm{B})_1=2N_2\operatorname{tr}\boldsymbol{A}_1\boldsymbol{M}^{-1}.$$

Next, for the calculation of $(\mathrm{A})_2-(\mathrm{B})_2$, because $\mathrm{E}[\boldsymbol{Y}^\star_{12}|\boldsymbol{Y}^\star_{11}]=\mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2'+\boldsymbol{Y}^\star_{11}\mathcal{B}_{12}$,

$$\boldsymbol{Y}^\star_{12}-\mathbf{1}_{N_1}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_2'-\boldsymbol{Y}^\star_{11}\mathcal{B}_{12}=(\boldsymbol{Y}^\star_{12}-\mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2'-\boldsymbol{Y}^\star_{11}\mathcal{B}_{12})-\mathbf{1}_{N_1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\widetilde{\boldsymbol{X}}_2'. \tag{18}$$

Further, because the first and second terms on the right side of equation (18) are independent,

$$\begin{aligned}(\mathrm{A})_2 =&\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{Y}^\star}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{12}^\star - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}^\star\mathcal{B}_{12})'(\boldsymbol{Y}_{12}^\star - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}^\star\mathcal{B}_{12})\\ &+ N_1\,\mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\widetilde{\boldsymbol{X}}_2(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\widetilde{\boldsymbol{X}}_2'\Big]\\ =&N_1p_2 + N_1\,\mathrm{tr}\,\boldsymbol{A}_2\boldsymbol{M}^{-1}.\end{aligned}$$

We note that $\mathrm{E}_{\boldsymbol{Y}^\star}\Big[(\boldsymbol{Y}_{12}^\star - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}^\star\mathcal{B}_{12})'(\boldsymbol{Y}_{12}^\star - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}^\star\mathcal{B}_{12})\Big] = N_1\boldsymbol{\Sigma}_{22\cdot 1}$. Similar to the calculation of $(\mathrm{A})_2$, we have

$$\begin{aligned}(\mathrm{B})_2 =&\mathrm{E}_{\boldsymbol{Y}}\Big[\mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{12} - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}\mathcal{B}_{12})'(\boldsymbol{Y}_{12} - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}\mathcal{B}_{12})\\ &+ N_1\,\mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\widetilde{\boldsymbol{X}}_2(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\widetilde{\boldsymbol{X}}_2' - 2\,\mathrm{tr}\,\boldsymbol{\Sigma}_{22\cdot 1}^{-1}(\boldsymbol{Y}_{12} - \mathbf{1}_{N_1}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_2' - \boldsymbol{Y}_{11}\mathcal{B}_{12})'\mathbf{1}_{N_1}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\widetilde{\boldsymbol{X}}_2'\Big]\\ =&N_1p_2 + N_1\,\mathrm{tr}\,\boldsymbol{A}_2\boldsymbol{M}^{-1} - 2\mathrm{E}[\mathrm{tr}\,\boldsymbol{a}_2(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'].\end{aligned}$$

Then, after some calculations, we obtain $\mathrm{E}[\mathrm{tr}\,\boldsymbol{a}_2(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'] = N_1\,\mathrm{tr}\,\boldsymbol{A}_2\boldsymbol{M}^{-1}$. Thus,

$$(\mathrm{A})_2 - (\mathrm{B})_2 = 2N_1\,\mathrm{tr}\,\boldsymbol{A}_2\boldsymbol{M}^{-1}.$$

Finally, we can obtain the following result:

$$\begin{aligned}\mathrm{CT} &= (\mathrm{A})_1 - (\mathrm{B})_1 + (\mathrm{A})_2 - (\mathrm{B})_2\\ &= 2N_2\,\mathrm{tr}\,\boldsymbol{A}_1\boldsymbol{M}^{-1} + 2N_1\,\mathrm{tr}\,\boldsymbol{A}_2\boldsymbol{M}^{-1}\\ &= 2\,\mathrm{tr}(N_2\boldsymbol{A}_1 + N_1\boldsymbol{A}_2)\boldsymbol{M}^{-1}\\ &= 2q.\end{aligned}$$

□

Proof of Theorem 2

In the derivation, we use the MLE of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is known with k-step monotone missing data given by Yagi et al., (2021), that is,

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{M}^{-1}\sum_{j=1}^{k}\boldsymbol{a}_j,$$

where

$$
\begin{aligned}
\boldsymbol{M} &= \sum_{j=1}^{k} N_{k-j+1}\boldsymbol{A}_j, \quad \boldsymbol{a}_1 = \boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_{(12\ldots k)1}'\boldsymbol{1}_{N_k}, \quad \boldsymbol{A}_1 = \boldsymbol{X}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1,\\
\boldsymbol{a}_j &= \widetilde{\boldsymbol{X}}_j'\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1}\Big(\boldsymbol{Y}_{(12\ldots,k-j+1)j} - \boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j}\Big)'\boldsymbol{1}_{N_{k-j+1}},\\
\boldsymbol{A}_j &= \widetilde{\boldsymbol{X}}_j'\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1}\widetilde{\boldsymbol{X}}_j, \quad j = 2, 3, \ldots, k.
\end{aligned}
$$

For the calculation of $(\mathrm{A})_j$ $(j = 2, 3, \ldots, k)$, because

$$
\begin{aligned}
&\mathrm{E}[\boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)j}|\boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)(12\ldots,j-1)}] = \boldsymbol{1}_{N_{k-j+1}}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_j' + \boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j},\\
&\boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)j} - \boldsymbol{1}_{N_{k-j+1}}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_j' - \boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j}\\
&= (\boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)j} - \boldsymbol{1}_{N_{k-j+1}}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_j' - \boldsymbol{Y}^{\star}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j}) - \boldsymbol{1}_{N_{k-j+1}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\widetilde{\boldsymbol{X}}_j',
\end{aligned}
\tag{19}
$$

the first and second terms on the right side of equation (19) are independent, and it holds that $\mathrm{E}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$, $\mathrm{Var}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{M}^{-1}$,

$$
\begin{aligned}
(A)_j &= N_{k-j+1}p_j + \mathrm{E}\Big[N_{k-j+1}\,\mathrm{tr}\,\boldsymbol{\Sigma}_{jj\cdot 12\ldots,j-1}^{-1}\widetilde{\boldsymbol{X}}_j(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\widetilde{\boldsymbol{X}}_j'\Big]\\
&= N_{k-j+1}p_j + N_{k-j+1}\,\mathrm{tr}\,\boldsymbol{A}_j\boldsymbol{M}^{-1}.
\end{aligned}
$$

As for $(\mathrm{B})_j$ $(j = 2, 3, \ldots, k)$, because

$$
\begin{aligned}
&\mathrm{E}[\boldsymbol{Y}_{(12\ldots,k-j+1)j}|\boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)}] = \boldsymbol{1}_{N_{k-j+1}}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_j' + \boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j},\\
&\boldsymbol{Y}_{(12\ldots,k-j+1)j} - \boldsymbol{1}_{N_{k-j+1}}\widehat{\boldsymbol{\theta}}'\widetilde{\boldsymbol{X}}_j' - \boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j}\\
&= \boldsymbol{Y}_{(12\ldots,k-j+1)j} - \boldsymbol{1}_{N_{k-j+1}}\boldsymbol{\theta}'\widetilde{\boldsymbol{X}}_j' - \boldsymbol{Y}_{(12\ldots,k-j+1)(12\ldots,j-1)}\mathcal{B}_{(12\ldots,j-1)j} - \boldsymbol{1}_{N_{k-j+1}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\widetilde{\boldsymbol{X}}_j'
\end{aligned}
\tag{20}
$$

but the first and second terms on the right side of equation (20) are not independent. Therefore

$$
(B)_j = N_{k-j+1}p_j + N_{k-j+1}\,\mathrm{tr}\,\boldsymbol{A}_j\boldsymbol{M}^{-1} - 2\mathrm{E}[\mathrm{tr}\,\boldsymbol{a}_j(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})']
$$

As for the calculation of $\mathrm{E}\Big[\mathrm{tr}\,\boldsymbol{a}_j(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\Big]$,

$$
\begin{aligned}
\mathrm{E}[\mathrm{tr}\,\boldsymbol{a}_j(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'] &= \mathrm{E}[\mathrm{tr}\{\boldsymbol{a}_j(\boldsymbol{a}_1 + \boldsymbol{a}_2 + \cdots + \boldsymbol{a}_k)'\boldsymbol{M}^{-1}\} - \mathrm{tr}\,\boldsymbol{a}_j\boldsymbol{\theta}']\\
&= \mathrm{E}\Big[\sum_{\substack{i=1\\ i\neq j}}^{k}\mathrm{tr}\,\boldsymbol{a}_j\boldsymbol{a}_i'\boldsymbol{M}^{-1} + \mathrm{tr}\,\boldsymbol{a}_j\boldsymbol{a}_j'\boldsymbol{M}^{-1} - \mathrm{tr}\,\boldsymbol{a}_j\boldsymbol{\theta}'\Big].
\end{aligned}
$$

Then, using Lemma 1 of Yagi et al. (2021),

$$
\begin{aligned}
\sum_{\substack{i=1\\ i\neq j}}^{k} \operatorname{tr} \mathrm{E}[\boldsymbol{a}_j \boldsymbol{a}_i'] \boldsymbol{M}^{-1} &= \sum_{\substack{i=1\\ i\neq j}}^{k} N_{k-j+1} N_{k-i+1} \operatorname{tr} \boldsymbol{A}_j \boldsymbol{\theta}\boldsymbol{\theta}' \boldsymbol{A}_i \boldsymbol{M}^{-1}, \\
\operatorname{tr} \mathrm{E}[\boldsymbol{a}_j \boldsymbol{a}_j'] \boldsymbol{M}^{-1} &= N_{k-j+1} \operatorname{tr} \boldsymbol{A}_j \boldsymbol{M}^{-1} + N_{k-j+1}^2 \operatorname{tr} \boldsymbol{A}_j \boldsymbol{\theta}\boldsymbol{\theta}' \boldsymbol{A}_j \boldsymbol{M}^{-1}, \\
\operatorname{tr} \mathrm{E}[\boldsymbol{a}_j] \boldsymbol{\theta}' &= N_{k-j+1} \operatorname{tr} \boldsymbol{A}_j \boldsymbol{\theta}\boldsymbol{\theta}'.
\end{aligned}
$$

Therefore, $\mathrm{E}[\operatorname{tr} \boldsymbol{a}_j (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'] = N_{k-j+1} \operatorname{tr} \boldsymbol{A}_j \boldsymbol{M}^{-1}$, and we can obtain

$$
(A)_j - (B)_j = 2N_{k-j+1} \operatorname{tr} \boldsymbol{A}_j \boldsymbol{M}^{-1},\ j = 2, 3, \ldots, k.
$$

Similarly, for the case of $j = 1$, it can also be confirmed that

$$
(A)_1 - (B)_1 = 2N_k \operatorname{tr} \boldsymbol{A}_1 \boldsymbol{M}^{-1}.
$$

Thus,

$$
\begin{aligned}
\mathrm{CT}_k &= \sum_{j=1}^{k} \{(A)_j - (B)_j\} \\
&= 2 \operatorname{tr} \sum_{j=1}^{k} N_{k-j+1} \boldsymbol{A}_j \boldsymbol{M}^{-1} \\
&= 2q.
\end{aligned}
$$

□