Growth Curve Model with Bilinear Random Coefficients

(Last Modified: November 22, 2019)

Shinpei Imori[1*], Dietrich von Rosen[2,3], and Ryoya Oda[1].

[1] *Graduate School of Science, Hiroshima University, Japan.*

[2] *Department of Energy and Technology, Swedish University of Agricultural Sciences, Sweden.*

[3] *Department of Mathematics, Linköping University, Sweden.*

Abstract

In the present paper, we derive a new multivariate model to fit correlated data, representing a general model class. Our model is an extension of the Growth Curve model (also called generalized multivariate analysis of variance model) by additionally assuming randomness of regression coefficients like in linear mixed models. Each random coefficient has a linear or a bilinear form with respect to explanatory variables. In our model, the covariance matrices of the random coefficients is allowed to be singular. This yields flexible covariance structures of response data but the parameter space includes a boundary, and thus maximum likelihood estimators (MLEs) of the unknown parameters have more complicated forms than the crude Growth Curve models. We derive the MLEs in the proposed model by solving an optimization problem, and derive sufficient conditions for consistency property of the MLEs. Through simulation studies, we confirmed performance of the MLEs when the sample size and the size of the response variable are large.

Key words: Consistency; Growth Curve model; Maximum likelihood estimators; Random coefficients.

[*] Corresponding author, *E-mail address*: imori@hiroshima-u.ac.jp

1 Introduction

Today we are able to simultaneously measure many responses repeatedly over time intervals of various lengths and as a consequence a huge amount of data is obtained (sometimes termed big data). Often this leads researchers to formulate complicated models to capture the information provided by the collected data. The statistical paradigm consists of proposing suitable models and then validate these models against data. If the model does not seem to fit a new model is formulated and these steps are iterated until one is satisfied with the agreement. Moreover, it is important that the model which fits data is interpretable, meaning that parameters and estimators should be interpretable.

The statistical paradigm implies that models cannot be too complicated and data cannot be too complex if there should be a possibility to validate the models appropriately. In this article we extend the linear multivariate random coefficient regression models to bilinear random coefficient regression models which leads to a model class which is relatively flexible and where it seems possible to perform model validation.

The stem in our model, is usually called Growth Curve model or GMANOVA (generalized multivariate analysis of variance model). The model was formulated by [1] although other authors had earlier studied similar models. A general review of the model can, for example, be found in [2]. Moreover, there are a number of books which discuss the model, for example [3] focuses on testing, [4] presents a classical approach of how to work with the model and for a recent contribution see [5], where up-to-date knowledge about the model and some extensions are presented.

Linear mixed models can be defined via

$$x = A\beta + Z\theta + e, \tag{1}$$

where A and Z are known design matrices, β is a fixed parameter and, θ is a vector of random variables, which is normally distributed with mean 0_p, and which is independent of $e \sim N_p(0_p, \Sigma_x)$, where Σ_x is an unknown covariance matrix. This is a very general model class and in particular it is referred to the book by [6], where many results can be found for the model. If $A = Z$ we have a so called random coefficient regression model. References related to random coefficient regression models can, for example, be found in an article by [7] and in more recent work by [8] who considered the testing of fixed effects of the model. A multivariate version of (1) equals

$$X = ABC + Z\Theta + E_x, \tag{2}$$

where A and Z are as in (1) and C is a new design matrix, B is an unknown regression coefficients, $\Theta \sim N_{q,n}(O_{q,n}, \Sigma_\Theta, I_n)$ is independent of $E_x \sim N_{p,n}(0_p, \Sigma_x, I_n)$ and Σ_Θ is an unknown covariance matrix. For any matrices $A_1, A_2 \in \mathbb{R}^{a\times b}$, $A_3 \in \mathbb{R}^{a\times a}$ and $A_4 \in \mathbb{R}^{b\times b}$, $A_1 \sim N_{a,b}(A_2, A_3, A_4)$ is identical to $\mathrm{vec}(A_1) \sim N_{ab}(\mathrm{vec}(A_2), A_4 \otimes A_3)$, where $\mathrm{vec}(\cdot)$ is the vec operator and the symbol $\otimes$ denotes the Kronecker product. Note that $\mathrm{Var}[\mathrm{vec}(X)] = I_n \otimes (Z\Sigma_\Theta Z^\top + \Sigma_x)$. If $\Sigma_\Theta = O_{q,q}$, i.e., there is no random coefficient, the model in (2) is the above mentioned classical Growth Curve model of [1]. The model in (2) is also called Growth Curve model with random effects, [e.g. see 9, 10, 11, 12, 13]. In these articles many more references to earlier published achievements related to (2) can be found. If in (2) $A = Z$ we have a multivariate random coefficient regression model.

Our work concerns a generalization of the multivariate random coefficient regression model which is based on

$$X = ABC + A\mathcal{E}_\beta + \mathcal{E}_\gamma C + A\mathcal{E}_\theta C + \mathcal{E}_x, \tag{3}$$

where A and C are known matrices and $\mathcal{E}_\beta$, $\mathcal{E}_\gamma$, $\mathcal{E}_\theta$ and $\mathcal{E}_x$ are independently and normally distributed. More details for the model are provided in Section 2. We can consider the model in (3) as a model which reflects two attributes such as spatial and temporal variation as well as their interaction effect. Later we show that the model in fact can be written $X = ABC + DEF$ for some new matrices D and F which are functions of the dispersion matrices of the random effects, given in (3).

In this article maximum likelihood estimators (MLEs) are derived where the technical problem is that estimators may lay on the boundary of the parameter space which is a well known phenomenon and has beautifully been treated by [14] when considering a univariate random coefficient model. Furthermore, these results where extended by [15] who considered different tests in an extended Growth Curve model with random effects.

To conclude the introduction it is noted that the proposed model which will be studied is relatively general, has the possibility to handle complicated observational studies and for the model it is possible to carry out model validation, although the model validation part is postponed to another publication.

The model with all its details is presented in Section 2, and Appendix A shows how to obtain a canonical form of the model in Section 2. Thereafter in Section 3 MLEs are stated where the proof of the results are given in Appendix B and Appendix C. Consistency of MLEs is considered in Section 4, where the proofs except for $\hat{B}$ which is an estimator of B are shown in Appendix D because to show the consistency of $\hat{B}$ is trivial. In Section 5 the estimators are studied via simulations and in Section 6 a few concluding remarks are presented.

2 Growth Curve model with bilinear random coefficients

In this section, we propose a Growth Curve model with bilinear random coefficients. For all $i = 1, \ldots, p$ and $j = 1, \ldots, n$, an observed dataset is denoted by $\{(x_{ij}, a_{is}, c_{tj}); s = 1, \ldots, q, t = 1, \ldots, k\}$, where x_{ij} is a response variable, a_{is} and c_{tj} are non-stochastic explanatory variables. Then, a relationship between response variable and explanatory variables is obtained as follows:

$$x_{ij} = \sum_{s=1}^{q} a_{is}\beta_{sj} + \sum_{t=1}^{k} \gamma_{it}c_{tj} + \sum_{s=1}^{q}\sum_{t=1}^{k} a_{is}\theta_{st}c_{tj} + \varepsilon_{ij}, \tag{4}$$

where β_{sj}, γ_{it} and θ_{st} are regression coefficients, $\varepsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$ is a measurement error and $\sigma^2 > 0$ is an unknown parameter. A matrix form of (4) is given by

$$X = A\mathcal{B} + \Gamma C + A\Theta C + \mathcal{E}_x, \tag{5}$$

where $(X)_{ij} = x_{ij}$, $(A)_{is} = a_{is}$, $(\mathcal{B})_{sj} = \beta_{sj}$, $(\Gamma)_{it} = \gamma_{it}$, $(C)_{tj} = c_{tj}$, $(\Theta)_{st} = \theta_{st}$ and $(\mathcal{E}_x)_{ij} = \varepsilon_{ij}$ for all $i = 1, \ldots, p$, $j = 1, \ldots, n$, $s = 1, \ldots, q$ and $t = 1, \ldots, k$. Note that $\mathcal{E}_x \sim N_{p,n}(O_{p,n}, \sigma^2 I_p, I_n)$. We can regard $A\mathcal{B}$ and ΓC as single effects with respect to the design matrices A and C, respectively, and $A\Theta C$ expresses an interaction effect between A and C. For simplicity, we assume A and C to be of full rank. In this paper, we also assume that q and k are fixed constants and $np - nq - kp > 0$. We consider an asymptotic framework where p and/or n are allowed to diverge.

The regression coefficients $\mathcal{B}$, Γ and Θ are considered as random coefficients. The random structures are given as follows:

$$\begin{aligned} \mathcal{B} &= B_\beta C + \mathcal{E}_\beta, \\ \Gamma &= A B_\gamma + \mathcal{E}_\gamma, \\ \Theta &= B_\theta + \mathcal{E}_\theta, \end{aligned} \tag{6}$$

where B_β, B_γ and B_θ are $q \times k$ unknown regression coefficients, and $\mathcal{E}_\beta$, $\mathcal{E}_\gamma$ and $\mathcal{E}_\theta$ are error terms. The error distributions are assumed to be

$$\mathcal{E}_\beta \sim N_{q,n}(O_{q,n}, \Sigma, I_n),$$
$$\mathcal{E}_\gamma \sim N_{p,k}(O_{p,k}, \delta^2 I_p, \Psi),$$
$$\mathcal{E}_\theta \sim N_{q,k}(O_{q,k}, \Sigma, \Psi),$$

where Σ and Ψ are unknown covariance matrices. It is worth noting that Σ and Ψ are allowed to be singular. This implies that the parameter space of Σ and Ψ include its boundary, which violates usual regularity assumptions. Moreover, we assume all error matrices $\mathcal{E}_x$, $\mathcal{E}_\beta$, $\mathcal{E}_\gamma$ and $\mathcal{E}_\theta$ to be independent. Because $A\Theta C$ is a bilinear function of A and C although $A\mathcal{B}$ and ΓC are linear functions of A and C, respectively, we call (5) the Growth Curve model with bilinear random coefficients.

Combining (5) and (6), we obtain (3). Because $\mathrm{Var}[\mathrm{vec}(X)] = (C^\top \Psi C + I_n) \otimes (A\Sigma A^\top + \delta^2 I_p)$, we can define the Growth Curve model with bilinear random coefficients as follows:

$$X = ABC + DEF, \tag{7}$$

where $B = B_\beta + B_\gamma + B_\theta$, $D^2 = A\Sigma A^\top + \delta^2 I_p$, $F^2 = C^\top \Psi C + I_n$ and $E \sim N_{p,n}(O_{p,n}, I_p, I_n)$. This model is flexible because this can express various covariance structures by varying Σ and Ψ.

3 Maximum likelihood estimators

In this section, we derive the MLEs of the unknown parameters in (7). Without loss of generality, we assume a canonical form for (7), i.e., $A = (I_q, O_{q,p-q})^\top$ and $C = (I_k, O_{k,n-k})$. A transformation of the model from the general form to the canonical form is shown in Appendix A. In order to derive the MLEs easily, we use a bijective parameter transformation

$\Sigma_* = \Sigma + \delta^2 I_q$, $\Psi_* = \delta^2(\Psi + I_k)$. Thus, we firstly estimate Σ_* and Ψ_*, afterwords, we can obtain the estimators of Σ and Ψ by using the inverse transformation.

Since X is distributed as $N_{p,n}(ABC, D^2, F^2)$, the log-likelihood function $\ell(B, \delta^2, \Sigma_*, \Psi_*|X)$ is given by

$$\begin{aligned}\ell(B, \delta^2, \Sigma_*, \Psi_*|X) = &-\frac{np}{2}\log 2\pi - \frac{n}{2}\log|D^2| - \frac{p}{2}\log|F^2| \\ &- \frac{1}{2}\mathrm{tr}\{D^{-2}(X - ABC)F^{-2}(X - ABC)^\top\}.\end{aligned}$$

Dividing X into four parts as follows:

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} : \quad \begin{matrix} q \times k & q \times (n-k) \\ (p-q) \times k & (p-q) \times (n-k) \end{matrix}$$

we can see that under the canonical form $\log|D^2| = \log|\Sigma_*| + (p - q)\log\delta^2$, $\log|F^2| = \log|\Psi_*| - k\log\delta^2$ and

$$\begin{aligned}&\mathrm{tr}\{D^{-2}(X - ABC)F^{-2}(X - ABC)^\top\} \\ &= \delta^2\mathrm{tr}\{\Psi_*^{-1}(X_{11} - B)^\top\Sigma_*^{-1}(X_{11} - B)\} \\ &\quad + \mathrm{tr}(\Sigma_*^{-1}X_{12}X_{12}^\top) + \mathrm{tr}(\Psi_*^{-1}X_{21}^\top X_{21}) + \delta^{-2}\mathrm{tr}(X_{22}X_{22}^\top).\end{aligned}$$

Hence, -2 log-likelihood function is simplified as follows:

$$\begin{aligned}-2\ell(B, \delta^2, \Sigma_*, \Psi_*|X) \propto\ &(np - nq - kp)\log\delta^2 + n\log|\Sigma_*| + p\log|\Psi_*| \\ &+ \delta^2\mathrm{tr}\{\Psi_*^{-1}(X_{11} - B)\Sigma_*^{-1}(X_{11} - B)^\top\} \\ &+ \mathrm{tr}(\Sigma_*^{-1}X_{12}X_{12}^\top) + \mathrm{tr}(\Psi_*^{-1}X_{21}^\top X_{21}) + \delta^{-2}\mathrm{tr}(X_{22}X_{22}^\top).\end{aligned} \tag{8}$$

As a preparation of deriving the MLEs, applying spectral decomposition to $X_{12}X_{12}^\top$ and

$X_{21}^\top X_{21}$, we obtain

$$X_{12}X_{12}^\top = Q_{12}^\top \mathrm{diag}\{\lambda_1^{(12)}, \ldots, \lambda_q^{(12)}\}Q_{12},$$
$$X_{21}^\top X_{21} = Q_{21}^\top \mathrm{diag}\{\lambda_1^{(21)}, \ldots, \lambda_k^{(21)}\}Q_{21},$$

where Q_{12} and Q_{21} are orthogonal matrices, $\lambda_s^{(12)} = \lambda_s(X_{12}X_{12}^\top)$, $\lambda_t^{(21)} = \lambda_t(X_{21}^\top X_{21})$ and $\lambda_\ell(\cdot)$ denotes the ℓth largest eigenvalue. Specially, we denote $\lambda_0(\cdot) = \infty$. Using these symbols, we define the following estimators of B, δ^2, Σ_* and Ψ_* by

$$\begin{aligned}
\hat{B} &= X_{11} = A^\top X C^\top, \\
\hat{\delta}^2(q_\sigma, k_\psi) &= \frac{\mathrm{tr}(X_{22}X_{22}^\top) + \sum_{s>q_\sigma} \lambda_s^{(12)} + \sum_{t>k_\psi} \lambda_t^{(21)}}{np - nq_\sigma - k_\psi p}, \\
\hat{\Sigma}_*(q_\sigma, k_\psi) &= Q_{12}^\top \mathrm{diag}\{\lambda_1^{(12)}/n, \ldots, \lambda_{q_\sigma}^{(12)}/n, \hat{\delta}^2(q_\sigma, k_\psi), \ldots, \hat{\delta}^2(q_\sigma, k_\psi)\}Q_{12}, \\
\hat{\Psi}_*(q_\sigma, k_\psi) &= Q_{21}^\top \mathrm{diag}\{\lambda_1^{(21)}/p, \ldots, \lambda_{k_\psi}^{(21)}/p, \hat{\delta}^2(q_\sigma, k_\psi), \ldots, \hat{\delta}^2(q_\sigma, k_\psi)\}Q_{21},
\end{aligned} \tag{9}$$

and a subset of $\mathcal{L}_F = \{(q_\sigma, k_\psi) | 0 \le q_\sigma \le q, 0 \le k_\psi \le k\}$ by

$$\mathcal{L} = \{(q_\sigma, k_\psi) \in \mathcal{L}_F | \lambda_{q_\sigma}^{(12)}/n \ge \hat{\delta}^2(q_\sigma, k_\psi), \lambda_{k_\psi}^{(21)}/p \ge \hat{\delta}^2(q_\sigma, k_\psi)\},$$

where $\hat{\Sigma}_*(0, k_\psi) = \hat{\delta}(0, k_\psi)I_q$ and $\hat{\Psi}_*(q_\sigma, 0) = \hat{\delta}(q_\sigma, 0)I_k$.

Then, the MLEs can be calculated according to the next theorem:

Theorem 1. Suppose that the model in (7) has the canonical form. Then, it follows that

$$\max_{B,\delta^2,\Sigma_*,\Psi_*} \ell(B, \delta^2, \Sigma, \Psi | X) = \max_{(q_\sigma,k_\psi)\in\mathcal{L}} \ell(\hat{B}, \hat{\delta}^2(q_\sigma, k_\psi), \hat{\Sigma}_*(q_\sigma, k_\psi), \hat{\Psi}_*(q_\sigma, k_\psi)|X).$$

A proof of Theorem 1 is given in Appendix B. Theorem 1 implies that the optimization problem of the log-likelihood function with respect to the unknown parameters is reduced to a discrete optimization problem of (q_σ, k_ψ). Although the computational burden on the optimization of (q_σ, k_ψ) is not heavy because the number of possible combinations of (q_σ, k_ψ) is at most $(q+1)(k+1)$, we can in the next theorem give a smaller range

$$\mathcal{L}_R = \{(q_\sigma, k_\psi) \in \mathcal{L} | \forall (q'_\sigma, k'_\psi) \in \mathcal{L} \setminus \{(q_\sigma, k_\psi)\}, q_\sigma > q'_\sigma \text{ or } k_\psi > k'_\psi\} \tag{10}$$

for searching for the optimal (q_σ, k_ψ). The proof of the next theorem is given in Appendix C.

Theorem 2. Suppose that the model (7) is written in a canonical form. The optimum value of $(q_\sigma, k_\psi) \in \mathcal{L}$ in Theorem 1 belongs to $\mathcal{L}_R$, that is,

$$(\hat{q}_\sigma, \hat{k}_\psi) = \arg \max_{(q_\sigma, k_\psi) \in \mathcal{L}} \ell(q_\sigma, k_\psi | X) \in \mathcal{L}_R,$$

where $\ell(q_\sigma, k_\psi | X) = \ell(\hat{B}, \hat{\delta}^2(q_\sigma, k_\psi), \hat{\Sigma}_*(q_\sigma, k_\psi), \hat{\Psi}_*(q_\sigma, k_\psi) | X)$.

For example, for each $k_\psi = 1, \ldots, k$, let $q_\sigma(k_\psi)$ be the maximum q_σ such that $(q_\sigma, k_\psi) \in \mathcal{L}$. Then, it suffices to search for the optimal $(\hat{q}_\sigma, \hat{k}_\psi)$ among $\{(q_\sigma(k_\psi), k_\psi) | k_\psi = 1, \ldots, k\}$.

4 Consistency of MLEs

In this section, we show the consistency of the MLEs. The MLEs in the previous section are derived under the canonical form. Thus, applying the inverse transformation, we obtain estimators of the unknown parameters in the general form.

Let us present the estimators given (q_σ, k_ψ), in the general form

$$\begin{aligned}
\hat{B} &= (A^\top A)^{-1} A^\top X C^\top (C C^\top)^{-1}, \\
\hat{\Sigma}(q_\sigma, k_\psi) &= (A^\top A)^{-1/2} \{\hat{\Sigma}_*(q_\sigma, k_\psi) - \hat{\delta}^2(q_\sigma, k_\psi) I_q\} (A^\top A)^{-1/2}, \\
\hat{\Psi}(q_\sigma, k_\psi) &= (C C^\top)^{-1/2} \{\hat{\Psi}_*(q_\sigma, k_\psi) / \hat{\delta}^2(q_\sigma, k_\psi) - I_k\} (C C^\top)^{-1/2},
\end{aligned}$$

where $\hat{\delta}^2(q_\sigma, k_\psi)$, $\hat{\Sigma}_*(q_\sigma, k_\psi)$ and $\hat{\Psi}_*(q_\sigma, k_\psi)$ defined in (9) are derived in the canonical form.

Let $\hat{\delta}^2 = \hat{\delta}^2(\hat{q}_\sigma, \hat{k}_\psi)$, $\hat{\Sigma} = \hat{\Sigma}(\hat{q}_\sigma, \hat{k}_\psi)$ and $\hat{\Psi} = \hat{\Psi}(\hat{q}_\sigma, \hat{k}_\psi)$ as the optimal estimators of δ^2, Σ and Ψ, respectively, where $(\hat{q}_\sigma, \hat{k}_\psi)$ is defined in Theorem 2.

Hereafter, we assume eigenvalues of the design matrices to satisfy that there exist ζ_a and $\zeta_c \in (0, 1]$ such that for all n and p, it holds that

$$0 < \zeta_a \leq \lambda_q(A^\top A)/p, \quad 0 < \zeta_c \leq \lambda_k(CC^\top)/n. \tag{11}$$

Considering the following three asymptotic frameworks: (i) $n \to \infty$, $p \to \infty$, (ii) $n \to \infty$, p is fixed and (iii) n is fixed, $p \to \infty$, we show the consistency of $\hat{\delta}^2$, $\hat{\Sigma}$ and $\hat{\Psi}$ under these frameworks and the condition (11). The next two lemmas are needed and shown in Appendix D. Define $q_\sigma^* = \mathrm{rank}(\Sigma)$ and $k_\psi^* = \mathrm{rank}(\Psi)$.

Lemma 1. Suppose that $\lambda_s(\Sigma)$ $(s = 1, \ldots, q)$ and $\lambda_t(\Psi)$ $(t = 1, \ldots, k)$ are constants. Then, for each framework (i)–(iii),

$$Pr((q_\sigma, k_\psi) \in \mathcal{L}_R \Rightarrow q_\sigma \geq q_\sigma^*, k_\psi \geq k_\psi^*) \to 1,$$

where $\mathcal{L}_R$ is defined in (10).

This lemma indicates that for arbitrary large probability $\hat{q}_\sigma$ and $\hat{k}_\psi$ are not lower than the true ranks of Σ and Ψ, respectively, when n and/or p are sufficiently large. On the other hand, it follows from the following lemma that for fixed (q_σ, k_ψ), which satisfies $q_\sigma \geq q_\sigma^*$ and $k_\psi \geq k_\psi^*$, then $\hat{\delta}^2$, $\hat{\Sigma}$ and $\hat{\Psi}$ are consistent.

Lemma 2. Suppose that $\lambda_s(\Sigma)$ $(s = 1, \ldots, q)$ and $\lambda_t(\Psi)$ $(t = 1, \ldots, k)$ are constants. Then, under the condition (11), for all $q_\sigma \geq q_\sigma^*$ and $k_\psi \geq k_\psi^*$,

$$
\begin{aligned}
|\hat{\delta}^2(q_\sigma, k_\psi) - \delta^2| &\xrightarrow{p} 0, \quad \text{for (i) -- (iii);} \\
\|\hat{\Sigma}(q_\sigma, k_\psi) - \Sigma\|_2 &\xrightarrow{p} 0, \quad \text{for (i) and (ii);} \\
\|\hat{\Psi}(q_\sigma, k_\psi) - \Psi\|_2 &\xrightarrow{p} 0, \quad \text{for (i) and (iii),}
\end{aligned}
$$

where $\|G\|_2 = \lambda_1(G^\top G)^{1/2}$ for a matrix G and the frameworks (i) – (iii) are mentioned in Lemma 1 and defined before the lemma.

Because q and k are fixed constants, by combining Lemma 1 and Lemma 2, the consistency of the MLEs can be established immediately.

Theorem 3. Let $\hat{\delta}^2 = \hat{\delta}^2(\hat{q}_\sigma, \hat{k}_\psi)$, $\hat{\Sigma} = \hat{\Sigma}(\hat{q}_\sigma, \hat{k}_\psi)$ and $\hat{\Psi} = \hat{\Psi}(\hat{q}_\sigma, \hat{k}_\psi)$. Suppose that $\lambda_s(\Sigma)$ $(s = 1, \ldots, q)$ and $\lambda_t(\Psi)$ $(t = 1, \ldots, k)$ are constants. Then, under the condition (11),

$$
\begin{aligned}
|\hat{\delta}^2 - \delta^2| &\xrightarrow{p} 0, \quad \text{for (i) -- (iii);} \\
\|\hat{\Sigma} - \Sigma\|_2 &\xrightarrow{p} 0, \quad \text{for (i) and (ii);} \\
\|\hat{\Psi} - \Psi\|_2 &\xrightarrow{p} 0, \quad \text{for (i) and (iii).}
\end{aligned}
$$

Next we show the consistency of $\hat{B}$.

Theorem 4. Under the condition (11),

(a) if $\lambda_1(\Sigma) \to 0$ as $p \to \infty$, then

$$\|\hat{B} - B\|_2 \xrightarrow{p} 0, \quad p \to \infty.$$

(b) if $\lambda_1(\Psi) \to 0$ as $n \to \infty$, then

$$\|\hat{B} - B\|_2 \xrightarrow{p} 0, \quad n \to \infty.$$

A proof of this theorem is omitted because that is straightforward from the variance of $\hat{B}$. This theorem indicates that when the true variance Σ or Ψ is small, the MLE of B is consistent. Note that even when $\lambda_1(\Sigma) \to 0$ or $\lambda_1(\Psi) \to 0$, it is possible that the corresponding covariance matrices of the response variable, D^2 and F^2, do not converge to $\delta^2 I_p$ and I_n, respectively.

5 Simulations

In this section, through 1000 Monte Carlo simulations, we calculate mean squared errors (MSEs) of $\hat{B}$, $\hat{\delta}^2$, $\hat{\Sigma}$ and $\hat{\Psi}$ in order to show the performance of the MLEs derived in Section 3. For any matrix $\hat{\Theta}$, which is an estimator of Θ, the MSE of $\hat{\Theta}$ is defined by

$$\mathrm{MSE}(\hat{\Theta}) = E[\|\hat{\Theta} - \Theta\|_F^2],$$

where the expectation $E[\cdot]$ is approximated by an average value of 1000 simulation results and $\|\cdot\|_F$ denotes the Frobenius norm. Note that if $\mathrm{MSE}(\hat{\Theta})$ goes to zero, then $\hat{\Theta} \xrightarrow{p} \Theta$ [see e.g., 16].

For each iteration, the response matrix X is generated according to

$$X \sim N_{p,n}(ABC, D^2, F^2),$$

where $D^2 = A\Sigma A^\top + \delta^2 I_p$ and $F^2 = C^\top \Psi C + I_n$. The design matrices A and C are given by

$$A = (1_p, A'), \quad A' \sim N_{p,q-1}(O_{p,q-1}, I_p, I_{q-1}),$$
$$C = (1_n, C')^\top, \quad C' \sim N_{n,k-1}(O_{n,k-1}, I_n, I_{k-1}).$$

The entries of B are given by i.i.d. copies from $N(0,1)$, and we set $q = 5$, $k = 3$, $\delta^2 = 2$,

$\Sigma = \text{diag}(1, 1, 0, 0, 0)$ and $\Psi = \alpha_n \text{diag}(1, 0, 0)$ with a positive sequence α_n. Three choices of α_n have been studied:

Case 1: $\alpha_n = 1$;

Case 2: $\alpha_n = n^{-1/2}$;

Case 3: $\alpha_n = n^{-1}$.

The covariance matrices, Σ and Ψ, are singular. For each case, n and p, we prepare three scenarios, (i) $n \to \infty$ and $p \to \infty$; (ii) $n \to \infty$ and p is fixed; (iii) n is fixed and $p \to \infty$. In the simulations, n and p vary from 20 to 3000.

It follows from Theorem 3 that when n tends to infinity $\hat{\Sigma}$ is consistent and when p tends to infinity $\hat{\Psi}$ is consistent, whereas if either n or p is large $\hat{\delta}^2$ is consistent. However, the condition for consistency of $\hat{B}$, given in Theorem 4, is only satisfied in the Cases 2 and 3 when n goes to infinity.

Tables 1, 2 and 3 present the MSEs of $\hat{B}$, $\hat{\delta}^2$, $\hat{\Sigma}$ and $\hat{\Psi}$, for Cases 1, 2 and 3, respectively. It can be seen from Tables 2 and 3 that the MSE of $\hat{B}$ is small when n is large while it does not go to zero even when n and p are large in Table 1. This may be due to the condition for the consistency of $\hat{B}$ given in Theorem 4. In order to check the condition, it needs to estimate Σ and Ψ, precisely. As mentioned previously, it follows from Theorem 3 that $\hat{\Sigma} \xrightarrow{p} \Sigma$ and $\hat{\Psi} \xrightarrow{p} \Psi$ with $n \to \infty$ and $p \to \infty$, respectively. Therefore, we next check the performance of $\hat{\Sigma}$ and $\hat{\Psi}$ in finite sample situations. The MSE of $\hat{\Sigma}$ can be small when n is large regardless of p in all tables although it does not approach zero when n is small. An opposite result can be seen for the MSE of $\hat{\Psi}$ in Table 1. On the other hand, the MSE of $\hat{\Psi}$ is small even when p is finite in Tables 2 and 3. This is because the entries of $\hat{\Psi}$ approach zero in Cases 2 and 3 when n tends to infinity. Moreover, in all tables, the MSE of $\hat{\delta}^2$ become small when n and/or p are small. These results agree with the consistency property for the estimators shown in

Theorem 3. Thus, the adequacy of the MLEs has been confirmed through this simulation study.

Table 1: (Case 1: $\alpha_n = 1$) Mean squared errors (MSE) and their standard errors (SE) for $\hat{B}$, $\hat{\delta}^2$, $\hat{\Sigma}$ and $\hat{\Psi}$ are presented when n and p vary between 20 and 3000. All entries of MSE and SE are rounded to 3 digits.

		$\hat{B}$		$\hat{\delta}^2$		$\hat{\Sigma}$		$\hat{\Psi}$	
n	p	MSE	SE	MSE	SE	MSE	SE	MSE	SE
20	20	3.127	0.079	0.038	0.001	0.456	0.008	0.138	0.005
50	50	2.263	0.064	0.004	0.000	0.136	0.003	0.046	0.002
100	100	2.100	0.062	0.001	0.000	0.064	0.002	0.023	0.001
200	200	2.048	0.065	0.000	0.000	0.030	0.001	0.011	0.000
500	500	1.928	0.059	0.000	0.000	0.012	0.000	0.004	0.000
1000	1000	2.169	0.071	0.000	0.000	0.006	0.000	0.002	0.000
2000	2000	2.038	0.065	0.000	0.000	0.003	0.000	0.001	0.000
3000	3000	2.048	0.060	0.000	0.000	0.002	0.000	0.001	0.000
50	20	2.877	0.071	0.013	0.001	0.194	0.004	0.133	0.004
100	20	2.800	0.075	0.006	0.000	0.097	0.002	0.145	0.005
200	20	2.774	0.068	0.003	0.000	0.048	0.001	0.135	0.004
500	20	2.789	0.079	0.001	0.000	0.020	0.000	0.138	0.004
1000	20	2.660	0.071	0.001	0.000	0.010	0.000	0.136	0.004
2000	20	2.733	0.074	0.000	0.000	0.005	0.000	0.138	0.004
3000	20	2.662	0.070	0.000	0.000	0.003	0.000	0.140	0.005
20	50	2.474	0.064	0.011	0.000	0.351	0.008	0.052	0.002
20	100	2.341	0.068	0.005	0.000	0.322	0.007	0.027	0.001
20	200	2.321	0.066	0.002	0.000	0.320	0.008	0.013	0.001
20	500	2.317	0.068	0.001	0.000	0.303	0.007	0.005	0.000
20	1000	2.244	0.064	0.000	0.000	0.304	0.007	0.003	0.000
20	2000	2.236	0.066	0.000	0.000	0.299	0.007	0.001	0.000
20	3000	2.339	0.065	0.000	0.000	0.289	0.006	0.001	0.000

6 Concluding remarks

The present paper proposes a new multivariate model, which is an extension of the Growth Curve model and the linear mixed model. This model has a bilinear random coefficients structure and expresses a flexible model class for correlated data. Moreover, we derive the

Table 2: (Case 2: $\alpha_n = n^{-1/2}$) Mean squared errors (MSE) and their standard errors (SE) for $\hat{B}$, $\hat{\delta}^2$, $\hat{\Sigma}$ and $\hat{\Psi}$ are presented when n and p vary between 20 and 3000. All entries of MSE and SE are rounded to 3 digits.

		$\hat{B}$		$\hat{\delta}^2$		$\hat{\Sigma}$		$\hat{\Psi}$	
n	p	MSE	SE	MSE	SE	MSE	SE	MSE	SE
20	20	1.009	0.021	0.038	0.002	0.456	0.008	0.011	0.000
50	50	0.434	0.010	0.004	0.000	0.136	0.003	0.001	0.000
100	100	0.269	0.007	0.001	0.000	0.064	0.002	0.000	0.000
200	200	0.175	0.005	0.000	0.000	0.030	0.001	0.000	0.000
500	500	0.098	0.003	0.000	0.000	0.012	0.000	0.000	0.000
1000	1000	0.075	0.002	0.000	0.000	0.006	0.000	0.000	0.000
2000	2000	0.049	0.001	0.000	0.000	0.003	0.000	0.000	0.000
3000	3000	0.039	0.001	0.000	0.000	0.002	0.000	0.000	0.000
50	20	0.553	0.012	0.013	0.001	0.194	0.004	0.004	0.000
100	20	0.358	0.008	0.006	0.000	0.097	0.002	0.002	0.000
200	20	0.235	0.005	0.003	0.000	0.048	0.001	0.001	0.000
500	20	0.141	0.004	0.001	0.000	0.020	0.000	0.000	0.000
1000	20	0.092	0.002	0.001	0.000	0.010	0.000	0.000	0.000
2000	20	0.065	0.002	0.000	0.000	0.005	0.000	0.000	0.000
3000	20	0.051	0.001	0.000	0.000	0.003	0.000	0.000	0.000
20	50	0.811	0.018	0.011	0.000	0.351	0.008	0.004	0.000
20	100	0.759	0.019	0.005	0.000	0.322	0.007	0.002	0.000
20	200	0.753	0.018	0.002	0.000	0.320	0.008	0.001	0.000
20	500	0.751	0.019	0.001	0.000	0.303	0.007	0.000	0.000
20	1000	0.731	0.018	0.000	0.000	0.304	0.007	0.000	0.000
20	2000	0.724	0.018	0.000	0.000	0.299	0.007	0.000	0.000
20	3000	0.749	0.018	0.000	0.000	0.289	0.006	0.000	0.000

Table 3: (Case 3: $\alpha_n = n^{-1}$) Mean squared errors (MSE) and their standard errors (SE) for $\hat{B}$, $\hat{\delta}^2$, $\hat{\Sigma}$ and $\hat{\Psi}$ are presented when n and p vary between 20 and 3000. All entries of MSE and SE are rounded to 3 digits.

		$\hat{B}$		$\hat{\delta}^2$		$\hat{\Sigma}$		$\hat{\Psi}$	
n	p	MSE	SE	MSE	SE	MSE	SE	MSE	SE
20	20	0.534	0.009	0.040	0.002	0.456	0.008	0.002	0.000
50	50	0.176	0.003	0.004	0.000	0.136	0.003	0.000	0.000
100	100	0.086	0.002	0.001	0.000	0.064	0.002	0.000	0.000
200	200	0.042	0.001	0.000	0.000	0.030	0.001	0.000	0.000
500	500	0.016	0.000	0.000	0.000	0.012	0.000	0.000	0.000
1000	1000	0.009	0.000	0.000	0.000	0.006	0.000	0.000	0.000
2000	2000	0.004	0.000	0.000	0.000	0.003	0.000	0.000	0.000
3000	3000	0.003	0.000	0.000	0.000	0.002	0.000	0.000	0.000
50	20	0.224	0.004	0.013	0.001	0.194	0.004	0.000	0.000
100	20	0.114	0.002	0.006	0.000	0.097	0.002	0.000	0.000
200	20	0.055	0.001	0.003	0.000	0.048	0.001	0.000	0.000
500	20	0.022	0.000	0.001	0.000	0.020	0.000	0.000	0.000
1000	20	0.011	0.000	0.001	0.000	0.010	0.000	0.000	0.000
2000	20	0.005	0.000	0.000	0.000	0.005	0.000	0.000	0.000
3000	20	0.004	0.000	0.000	0.000	0.003	0.000	0.000	0.000
20	50	0.438	0.008	0.011	0.000	0.351	0.008	0.001	0.000
20	100	0.406	0.008	0.005	0.000	0.322	0.007	0.000	0.000
20	200	0.402	0.008	0.002	0.000	0.320	0.008	0.000	0.000
20	500	0.400	0.009	0.001	0.000	0.303	0.007	0.000	0.000
20	1000	0.393	0.008	0.000	0.000	0.304	0.007	0.000	0.000
20	2000	0.386	0.008	0.000	0.000	0.299	0.007	0.000	0.000
20	3000	0.394	0.008	0.000	0.000	0.289	0.006	0.000	0.000

MLEs of the unknown parameters as well as their consistency property. Because the parameter space includes the boundary and the maximum can take place on the boundary of the parameter space, the structure as well as distributions of the MLEs except for $\hat{B}$ is complicated so that their properties such as finding of an asymptotic distribution looks hard to show theoretically. From the simulation studies, we demonstrate the finite sample performance of the MLEs that can be equipped with a consistency property even when both n and p are large.

It is possible to extend our result to the case when we have N independent observations of X in (3), i.e. $X_1, \ldots, X_N$. Thus, we have independent observations of a bilinear random regression model which indeed means that we have a trilinear scenario. Nowadays, some results exist for trilinear models and a recent reference is [17], where also other works are mentioned. However, our extended model differs significantly from the model considered in [17], i.e. they do not include random coefficient regression effects in their model. It will be possible to produce estimators of the unknown parameters using N samples, $X_1, \ldots, X_N$, corresponding to Theorems 1 and 2 in this paper. It will also be possible to select the true rank of the covariance matrices of the random coefficients by using an information criteria as considered in [18]. We can expect that this approach enables us to select the true rank with a probability tending to one, under appropriate conditions. However, we leave this for future works.

Acknowledgements

Shinpei Imori is supported in part by JSPS KAKENHI Grant Number 17K12650 and “Funds for the Development of Human Resources in Science and Technology” under MEXT, through the “Home for Innovative Researchers and Academic Knowledge Users (HIRAKU)” consortium. Dietrich von Rosen is supported by the Swedish Research Council (2017-03003). Ryoya Oda is supported by Research Fellowship for Young Scientists from the Japan Society for the

Promotion of Science.

References

[1] R. F. Potthoff, S. Roy, A generalized multivariate analysis of variance model useful especially for growth curve problems, Biometrika 51 (3-4) (1964) 313–326.

[2] D. von Rosen, The growth curve model: a review, Communications in Statistics-Theory and Methods 20 (9) (1991) 2791–2822.

[3] T. Kariya, Testing in the multivariate general linear model, Kinokuniya, 1985.

[4] A. Kshirsagar, W. Smith, Growth Curves, Vol. 145, CRC Press, 1995.

[5] D. von Rosen, Bilinear regression analysis: An Introduction, Vol. 220, Springer, 2018.

[6] E. Demidenko, Mixed models: Theory and applications, Wiley Series in Probability and Statistics, Wiley-Interscience, 2004.

[7] C. M. Rutter, R. M. Elashoff, Analysis of longitudinal data: random coefficient regression modelling, Statistics in Medicine 13 (12) (1994) 1211–1231.

[8] J. Volaufova, L. R. LaMotte, A simulation comparison of approximate tests for fixed effects in random coefficients growth curve models, Communications in Statistics-Simulation and Computation 42 (2) (2013) 344–359.

[9] G. Reinsel, Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure, Journal of the American Statistical Association 77 (377) (1982) 190–195.

[10] N. Lange, N. M. Laird, The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters, Journal of the American Statistical Association 84 (405) (1989) 241–247.

[11] T. Yokoyama, Efficiency of the MLE in a multivariate parallel profile model with random effects, Hiroshima Mathematical Journal 35 (2) (2005) 197–203.

[12] W.-C. Ip, M.-X. Wu, S.-G. Wang, H. Wong, Estimation for parameters of interest in random effects growth curve models, Journal of Multivariate Analysis 98 (2) (2007) 317–327.

[13] Z. Li, Testing for random effects in growth curve models, Communications in Statistics-Theory and Methods 44 (3) (2015) 564–572.

[14] J. R. Schott, Multivariate maximum likelihood estimators for the mixed linear model, Sankhyā: The Indian Journal of Statistics, Series B (1985) 179–185.

[15] Y. Fujikoshi, D. von Rosen, LR tests for random-coefficient covariance structures in an extended growth curve model, Journal of Multivariate Analysis 75 (2) (2000) 245–268.

[16] E. L. Lehmann, Elements of large-sample theory, Springer Science & Business Media, 2004.

[17] K. Filipiak, D. Klein, Estimation of parameters under a generalized growth curve model, Journal of Multivariate Analysis 158 (2017) 73–86.

[18] S. Imori, D. von Rosen, Covariance components selection in high-dimensional growth curve model with random coefficients, Journal of Multivariate Analysis 136 (2015) 86–94.

[19] T. Siotani, Y. Fujikoshi, T. Hayakawa, Modern multivariate statistical analysis, a graduate course and handbook, American Sciences Press, 1985.

Appendix A Transformation from general form to canonical form

We present one way of transforming the Growth Curve model to a canonical form. Let P_A and P_C be defined by

$$P_A = (A(A^\top A)^{-1/2}, Q_A), \quad P_C = (C^\top (CC^\top)^{-1/2}, Q_C),$$

where $Q_A^\top A = O_{p-q,q}$, $Q_A^\top Q_A = I_{p-q}$, $CQ_C = O_{k,n-k}$ and $Q_C^\top Q_C = I_{n-k}$ are fulfilled. Thus,

$$P_A^\top A = ((A^\top A)^{1/2}, O_{q,p-q})^\top, \quad CP_C = ((CC^\top)^{1/2}, O_{k,n-k}).$$

Moreover, it follows that $P_A^\top P_A = I_p$, $P_C^\top P_C = I_n$ and

$$P_A^\top D^2 P_A = (I_q, O_{q,p-q})^\top (A^\top A)^{1/2} \Sigma (A^\top A)^{1/2} (I_q, O_{q,p-q}) + \delta^2 I_p,$$
$$P_C^\top F^2 P_C = (I_k, O_{k,n-k})^\top (CC^\top)^{1/2} \Psi (CC^\top)^{1/2} (I_k, O_{k,n-k}) + I_n.$$

Using P_A and P_C indicates that

$$\tilde{X} := P_A^\top X P_C = P_A^\top (ABC + DEF) P_C = \tilde{A}\tilde{B}\tilde{C} + \tilde{D}E\tilde{F},$$

where $\tilde{A} = (I_q, O_{q,p-q})^\top$, $\tilde{B} = (A^\top A)^{1/2} B (CC^\top)^{1/2}$, $\tilde{C} = (I_k, O_{k,n-k})$, $\tilde{D}^2 = \tilde{A}\tilde{\Sigma}\tilde{A}^\top + \delta^2 I_p$, $\tilde{F}^2 = \tilde{C}^\top \tilde{\Psi} \tilde{C} + I_n$, $\tilde{\Sigma} = (A^\top A)^{1/2} \Sigma (A^\top A)^{1/2}$ and $\tilde{\Psi} = (CC^\top)^{1/2} \Psi (CC^\top)^{1/2}$. Hence, by replacing X, A, B, C, Σ and Ψ by $\tilde{X}$, $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, $\tilde{\Sigma}$ and $\tilde{\Psi}$, respectively, we obtain the canonical form in (7).

Appendix B Proof of Theorem 1

Proof. Since the fourth term on the right hand side in (8), $\delta^2 \mathrm{tr}\{\Psi_*^{-1}(X_{11}-B)\Sigma_*^{-1}(X_{11}-B)^\top\}$ is non-negative, the optimum value of B is given by $\hat{B} = X_{11}$. On the other hand, it follows from [14] that the maximum point of $\ell(\hat{B}, \Sigma_*, \Psi_*, \delta^2|X)$ attains when Σ_* and Ψ_* satisfy that $Q_{12}\Sigma_* Q_{12}^\top$ and $Q_{21}\Psi_* Q_{21}^\top$ are diagonal matrices, respectively. Hence, we re-parametrize $Q_{12}\Sigma_* Q_{12}^\top = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_q^2)$ and $Q_{21}\Psi_* Q_{21}^\top = \mathrm{diag}(\psi_1^2, \ldots, \psi_k^2)$. Moreover, the assumptions $\Sigma_* - \delta^2 I_q \geq 0$ and $\Psi_* - \delta^2 I_k \geq 0$ imply $\sigma_s^2 \geq \delta^2$ and $\psi_t^2 \geq \delta^2$, respectively. Therefore, instead of minimizing (8),

$$
\begin{aligned}
\min_{\tilde{\sigma}_s^2, \tilde{\psi}_t^2, \tilde{\delta}^2} \; & -(np - nq - kp) \log \tilde{\delta}^2 + \tilde{\delta}^2 \mathrm{tr}(X_{22} X_{22}^\top) \\
& + \sum_{s=1}^{q} \{-n \log \tilde{\sigma}_s^2 + \lambda_s^{(12)} \tilde{\sigma}_s^2\} \\
& + \sum_{t=1}^{k} \{-p \log \tilde{\psi}_t^2 + \lambda_t^{(21)} \tilde{\psi}_t^2\}.
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
\text{subject to} \;\; & \tilde{\sigma}_s^2 - \tilde{\delta}^2 \leq 0, \quad s = 1, \ldots, q, \\
& \tilde{\psi}_t^2 - \tilde{\delta}^2 \leq 0, \quad t = 1, \ldots, k,
\end{aligned} \tag{13}
$$

where $\tilde{\delta}^2 = \delta^{-2}$, $\tilde{\sigma}_s^2 = \sigma_s^{-2}$ and $\tilde{\psi}_t^2 = \psi_t^{-2}$. We note that this optimization problem consisting of (12) and (13) is a convex problem because $np - nq - kp$ is assumed to be positive. Hence, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient conditions for the optimization of (12) given (13). By using KKT multipliers $\lambda_{\sigma,s} \geq 0$ and $\lambda_{\psi,t} \geq 0$, along with

(13), we obtain the following KKT conditions:

$$
\begin{aligned}
&-(np - nq - kp)\tilde{\delta}^{-2} + \mathrm{tr}(X_{22}X_{22}^{\top}) - \sum_{s=1}^{q} \lambda_{\sigma,s} - \sum_{t=1}^{k} \lambda_{\psi,t} = 0, \\
&-n\tilde{\sigma}_s^{-2} + \lambda_s^{(12)} + \lambda_{\sigma,s} = 0, \quad s = 1, \ldots, q, \\
&-p\tilde{\psi}_t^{-2} + \lambda_t^{(21)} + \lambda_{\psi,t} = 0, \quad t = 1, \ldots, k, \\
&\lambda_{\sigma,s}(\tilde{\sigma}_s^2 - \tilde{\delta}^2) = 0, \quad s = 1, \ldots, q, \\
&\lambda_{\psi,t}(\tilde{\psi}_t^2 - \tilde{\delta}^2) = 0, \quad t = 1, \ldots, k.
\end{aligned}
$$

Let two index sets I_σ and I_ψ be defined by

$$
I_\sigma = \{1 \le s \le q | \lambda_{\sigma,s} = 0\}, \quad I_\psi = \{1 \le t \le k | \lambda_{\psi,t} = 0\}.
$$

The 2nd $-$ 5th KKT conditions show that

$$
\begin{aligned}
\tilde{\sigma}_s^{-2} &= \begin{cases} \tilde{\delta}^{-2}, & \lambda_{\sigma,s} \neq 0 \Leftrightarrow s \notin I_\sigma, \\ \lambda_s^{(12)}/n, & \lambda_{\sigma,s} = 0 \Leftrightarrow s \in I_\sigma, \end{cases} \\
\tilde{\psi}_t^{-2} &= \begin{cases} \tilde{\delta}^{-2}, & \lambda_{\psi,t} \neq 0 \Leftrightarrow t \notin I_\psi, \\ \lambda_t^{(21)}/p, & \lambda_{\psi,t} = 0 \Leftrightarrow t \in I_\psi. \end{cases}
\end{aligned}
\tag{14}
$$

Moreover, it follows from (14) and the 1st KKT condition that

$$
\begin{aligned}
0 &= -(np - nq - kp)\tilde{\delta}^{-2} + \mathrm{tr}(X_{22}X_{22}^\top) - \sum_{s=1}^{q} \lambda_{\sigma,s} - \sum_{t=1}^{k} \lambda_{\psi,t} \\
&= -(np - nq - kp)\tilde{\delta}^{-2} + \mathrm{tr}(X_{22}X_{22}^\top) - \sum_{s \notin I_\sigma} \lambda_{\sigma,s} - \sum_{t \notin I_\psi} \lambda_{\psi,t} \\
&= -(np - nq - kp)\tilde{\delta}^{-2} + \mathrm{tr}(X_{22}X_{22}^\top) - \sum_{s \notin I_\sigma} \{n\tilde{\delta}^{-2} - \lambda_s^{(12)}\} - \sum_{t \notin I_\psi} \{p\tilde{\delta}^{-2} - \lambda_t^{(21)}\} \\
&= -(pn - n\#I_\sigma - p\#I_\psi)\tilde{\delta}^{-2} + \mathrm{tr}(X_{22}X_{22}^\top) + \sum_{s \notin I_\sigma} \lambda_s^{(12)} + \sum_{t \notin I_\psi} \lambda_t^{(21)},
\end{aligned}
$$

where $\#I_\sigma$ and $\#I_\psi$ are the number of elements in the sets I_σ and I_ψ, respectively. Hence, given I_σ and I_ψ, estimators of σ_s^2, ψ_t^2 and δ^2 equal

$$
\begin{aligned}
\hat{\delta}^2(I_\sigma, I_\psi) &= \frac{\mathrm{tr}(X_{22}X_{22}^\top) + \sum_{s \notin I_\sigma} \lambda_s^{(12)} + \sum_{t \notin I_\psi} \lambda_t^{(21)}}{np - n\#I_\sigma - p\#I_\psi}, \\
\hat{\sigma}_s^2(I_\sigma, I_\psi) &= \begin{cases} \hat{\delta}^2(I_\sigma, I_\psi), & s \notin I_\sigma, \\ \lambda_s^{(12)}/n, & s \in I_\sigma, \end{cases} \\
\hat{\psi}_t^2(I_\sigma, I_\psi) &= \begin{cases} \hat{\delta}^2(I_\sigma, I_\psi), & t \notin I_\psi, \\ \lambda_t^{(21)}/p, & t \in I_\psi. \end{cases}
\end{aligned}
\tag{15}
$$

Because of (13), they have to satisfy that $\hat{\delta}^2(I_\sigma, I_\psi) \leq \hat{\sigma}_s^2(I_\sigma, I_\psi)$ and $\hat{\delta}^2(I_\sigma, I_\psi) \leq \hat{\psi}_t^2(I_\sigma, I_\psi)$ for all $s \in I_\sigma$ and $t \in I_\psi$. Note that if I_σ and I_ψ are full sets, then the MLEs of δ^2, Σ_* and Ψ_* are given by $\hat{\delta}^2(I_\sigma, I_\psi) = \mathrm{tr}(X_{22}X_{22}^\top)/(np - nq - kp)$, $\hat{\Sigma}_*(I_\sigma, I_\psi) = X_{12}X_{12}^\top/n$ and $\hat{\Psi}_*(I_\sigma, I_\psi) = X_{21}^\top X_{21}/p$, respectively.

Next, we want to optimize the index sets I_σ and I_ψ. Let $\ell(I_\sigma, I_\psi)$ be the maximum likelihood function given I_σ and I_ψ, where

$$
-2\ell(I_\sigma, I_\psi) \propto (pn - n\#I_\sigma - p\#I_\psi) \log \hat{\delta}^2(I_\sigma, I_\psi) + n \sum_{s \in I_\sigma} \log \frac{\lambda_s^{(12)}}{n} + p \sum_{t \in I_\psi} \log \frac{\lambda_t^{(21)}}{p}. \tag{16}
$$

At first, we fix I_ψ, and compare the index sets $I_\sigma^{(1)} = I_\sigma^{(0)} \cup \{s_1\}$ and $I_\sigma^{(2)} = I_\sigma^{(0)} \cup \{s_2\}$, where $I_\sigma^{(0)} \subsetneq \{1, \dots, q\}$ and $s_1, s_2 \notin I_\sigma^{(0)}$ satisfying $s_1 < s_2$. When $s_1 < s_2$, it holds that $\lambda_{s_1}^{(12)} \geq \lambda_{s_2}^{(12)}$. Thus, it also holds from (15) that $\hat{\delta}^2(I_\sigma^{(1)}, I_\psi) \leq \hat{\delta}^2(I_\sigma^{(2)}, I_\psi)$ and $\hat{\sigma}_{s_2}^2(I_\sigma^{(2)}, I_\psi) \leq \hat{\sigma}_{s_1}^2(I_\sigma^{(1)}, I_\psi)$. Suppose that the estimators based on $I_\sigma^{(2)}$ satisfies (13), i.e., $\hat{\delta}^2(I_\sigma^{(2)}, I_\psi) \leq \hat{\sigma}_s^2(I_\sigma^{(2)}, I_\psi)$ for all $s \in I_\sigma^{(2)}$. Then, it follows that

$$\hat{\delta}^2(I_\sigma^{(1)}, I_\psi) \leq \hat{\delta}^2(I_\sigma^{(2)}, I_\psi) \leq \hat{\sigma}_{s_2}^2(I_\sigma^{(2)}, I_\psi) \leq \hat{\sigma}_{s_1}^2(I_\sigma^{(1)}, I_\psi). \tag{17}$$

Hence, $\hat{\delta}^2(I_\sigma^{(1)}, I_\psi) \leq \hat{\sigma}_s^2(I_\sigma^{(1)}, I_\psi)$ is established for all $s \in I_\sigma^{(1)}$, that is, the estimators based on $I_\sigma^{(1)}$ satisfies (13).

Now we want to show that $-2\ell(I_\sigma^{(1)}, I_\psi) \leq -2\ell(I_\sigma^{(2)}, I_\psi)$, which indicates that $I_\sigma^{(1)}$ is better than $I_\sigma^{(2)}$ in terms of the likelihood function. To simplify notation, we denote

$$\begin{aligned}
K &= \operatorname{tr}(X_{22}X_{22}^\top) + \sum_{s \notin I_\sigma^{(0)}} \lambda_s^{(12)} + \sum_{t \notin I_\psi} \lambda_t^{(21)}, \\
L &= n \sum_{s \in I_\sigma^{(0)}} \log \frac{\lambda_s^{(12)}}{n} + p \sum_{t \in I_\psi} \log \frac{\lambda_t^{(21)}}{p}, \\
M &= np - n\#I_\sigma^{(0)} - p\#I_\psi.
\end{aligned}$$

Here, we use the inequality that for $a, b, c, d > 0$,

$$\frac{b}{a} \leq \frac{d}{c} \Rightarrow \frac{b}{a} \leq \frac{b+d}{a+c} \leq \frac{d}{c}. \tag{18}$$

Let $a = M - n$, $b = K - \lambda_{s_2}^{(12)}$, $c = n$ and $d = \lambda_{s_2}^{(12)}$. Because $b/a = \hat{\delta}^2(I_\sigma^{(2)}, I_\psi) \leq \hat{\sigma}_{s_2}^2(I_\sigma^{(2)}, I_\psi) = \lambda_{s_2}^{(12)}/n = d/c$ is established from (17), it follows from (18) that

$$\frac{K}{M} = \frac{b+d}{a+c} \leq \frac{d}{c} = \frac{\lambda_{s_2}^{(12)}}{n} \leq \frac{\lambda_{s_1}^{(12)}}{n}, \tag{19}$$

where the last inequality follows from the assumption that $s_1 < s_2$. For $x \in (0, K/n)$, denote

$$f(x) = (M-n)\log\frac{K-nx}{M-n} + n\log x + L,$$

which satisfies $f(\lambda^{(12)}_{s_m}/n) = -2\ell(I^{(m)}_\sigma, I_\psi)$ for $m = 1, 2$. It is easy to show that f is concave and its critical point is given by $x = K/M$, which implies that f is monotonically decreasing when $x \geq K/M$. This and (19) clarify that $-2\ell(I^{(1)}_\sigma, I_\psi) \leq -2\ell(I^{(2)}_\sigma, I_\psi)$ when the estimators based on $I^{(2)}_\sigma$ satisfies (13).

Hence, there exists a non-negative integer q_σ such that the optimum set for I_σ is obtained by $\hat{I}_\sigma \equiv \{1, \ldots, q_\sigma\}$. In a similar way, we can show that an optimal index set of I_ψ can be expressed by $\hat{I}_\psi \equiv \{1, \ldots, k_\psi\}$, where k_ψ is a non-negative integer. Note that $q_\sigma = 0$ and $k_\psi = 0$ imply $\hat{I}_\sigma$ and $\hat{I}_\psi$ are empty, respectively. Thus, the proof is completed. □

Appendix C Proof of Theorem 2

Proof. Let $I^{(0)}_\sigma = \{1, \ldots, q_\sigma\}$, $I^{(1)}_\sigma = \{1, \ldots, q_\sigma+1\}$, $I^{(0)}_\psi = \{1, \ldots, k_\psi\}$ and $I^{(1)}_\psi = \{1, \ldots, k_\psi+1\}$. At first, we show that $-2\ell(I^{(1)}_\sigma, I^{(0)}_\psi) \leq -2\ell(I^{(0)}_\sigma, I^{(0)}_\psi)$ when $(q_\sigma+1, k_\psi) \in \mathcal{L}$. Denoting K, L and M as in Appendix B, we can see that

$$-2\ell(I^{(0)}_\sigma, I^{(0)}_\psi) + 2\ell(I^{(1)}_\sigma, I^{(0)}_\psi) = M\log\frac{K}{M} - (M-n)\log\frac{K-\lambda^{(12)}_{q_\sigma+1}}{M-n} - n\log\frac{\lambda^{(12)}_{q_\sigma+1}}{n}.$$

Note that since $(q_\sigma+1, k_\psi)$ belongs to $\mathcal{L}$, it holds that $\{K-\lambda^{(12)}_{q_\sigma+1}\}/(M-n) = \hat{\delta}^2(I^{(1)}_\sigma, I^{(0)}_\psi) \leq \hat{\sigma}^2_{q_\sigma+1}(I^{(1)}_\sigma, I^{(0)}_\psi) = \lambda^{(12)}_{q_\sigma+1}/n$. From (18) with $a = M-n$, $b = K-\lambda^{(12)}_{q_\sigma+1}$, $c = n$ and $d = \lambda^{(12)}_{q_\sigma+1}$, it follows that $K/M = (b+d)/(a+c) \leq d/c = \lambda^{(12)}_{q_\sigma+1}/n < K/n$. Here, for $x \in (0, K/n)$, denote

$$g(x) = M\log\frac{K}{M} - (M-n)\log\frac{K-nx}{M-n} - n\log x.$$

Because g is convex and its critical point is given by K/M, we can see that

$$-2\ell(I_\sigma^{(0)}, I_\psi^{(0)}) + 2\ell(I_\sigma^{(1)}, I_\psi^{(0)}) = g(\lambda_{q_\sigma+1}^{(12)}/n) \geq g(K/M) = 0.$$

Likewise, it can be seen that $-2\ell(I_\sigma^{(0)}, I_\psi^{(1)}) \leq -2\ell(I_\sigma^{(0)}, I_\psi^{(0)})$ when $(q_\sigma, k_\psi+1) \in \mathcal{L}$. Hence, the proof is completed. □

Appendix D Proofs of Lemma 1 and Lemma 2

Proof. Firstly, we prepare the following lemma to show Lemma 1.

Lemma 3. For any $a \geq 0$ and symmetric matrices $A_1, A_2 \in \mathbb{R}^{m\times m}$ such that A_1 and A_2 are positive and non-negative definite, respectively, it holds that for all $\ell \in 1, \ldots, m$,

$$\{a + \lambda_\ell(A_2)\}\lambda_m(A_1) \leq \lambda_\ell(A_1(A_2 + aI_m)) \leq \{a + \lambda_\ell(A_2)\}\lambda_1(A_1).$$

It follows from some linear algebra [see e.g., 19] that for all $\ell \in 0, \ldots, m-1$,

$$\begin{aligned}\lambda_{\ell+1}(A_1(A_2 + aI_m)) &= \lambda_{\ell+1}(A_1^{1/2}(A_2 + aI_m)A_1^{1/2}) \\ &= \inf_{F_\ell} \sup_{F_\ell^\top x = 0_\ell} \frac{x^\top A_1^{1/2}(A_2 + aI_m)A_1^{1/2}x}{x^\top x} \\ &= \inf_{F_\ell} \sup_{F_\ell^\top y = 0_\ell} \frac{y^\top (A_2 + aI_m)y}{y^\top A_1^{-1} y},\end{aligned}$$

where $F_\ell \in \mathbb{R}^{m\times\ell}$. Because eigenvalues of A_1^{-1} are inverse of eigenvalues of A_1, which is positive definite, for all $y \in \mathbb{R}^m$,

$$\begin{aligned}\frac{y^\top (A_2 + aI_m)y}{y^\top A_1^{-1} y} &\leq \lambda_m(A_1^{-1})^{-1}\frac{y^\top (A_2 + aI_m)y}{y^\top y} = \lambda_1(A_1)\frac{y^\top (A_2 + aI_m)y}{y^\top y}, \\ \frac{y^\top (A_2 + aI_m)y}{y^\top A_1^{-1} y} &\geq \lambda_1(A_1^{-1})^{-1}\frac{y^\top (A_2 + aI_m)y}{y^\top y} = \lambda_m(A_1)\frac{y^\top (A_2 + aI_m)y}{y^\top y}.\end{aligned}$$

Note that

$$\inf_{F_\ell} \sup_{F_\ell^\top y = 0_\ell} \frac{y^\top (A_2 + aI_m) y}{y^\top y} = \lambda_{\ell+1}(A_2 + aI_m) = a + \lambda_{\ell+1}(A_2).$$

Hence, the proof is completed. □

Proof of Lemma 1

Proof. At first, we show $Pr((q_\sigma^*, k_\sigma^*) \in \mathcal{L}) \to 1$ for all asymptotic frameworks (i)–(iii) of the lemma, that is,

$$Pr(\lambda_{q_\sigma^*}^{(12)}/n \geq \hat{\delta}^2(q_\sigma^*, k_\sigma^*), \lambda_{k_\psi^*}^{(21)}/p \geq \hat{\delta}^2(q_\sigma^*, k_\sigma^*)) \to 1.$$

We firstly show that for all $q_\sigma \geq q_\sigma^*$ and $k_\psi \geq k_\psi^*$, $\hat{\delta}^2(q_\sigma, k_\psi)$ converges to δ^2 in probability. By considering the transformation in Appendix A, we can see that

$$\tilde{A}\tilde{B}\tilde{C} = \begin{pmatrix} \tilde{B} & O \\ O & O \end{pmatrix}, \quad \tilde{D}^2 = \begin{pmatrix} \tilde{\Sigma} + \delta^2 I_q & O_{q,p-q} \\ O_{p-q,q} & \delta^2 I_{p-q} \end{pmatrix}, \quad \tilde{F}^2 = \begin{pmatrix} \tilde{\Psi} + I_k & O_{k,n-k} \\ O_{n-k,k} & I_{n-k} \end{pmatrix}.$$

Splitting the error matrix E into four parts like X, i.e.,

$$E = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} : \quad \begin{matrix} q \times k & q \times (n-k) \\ (p-q) \times k & (p-q) \times (n-k) \end{matrix}$$

we can see that

$$\tilde{X} = \begin{pmatrix} \tilde{B} + (\tilde{\Sigma} + \delta^2 I_q)^{1/2} E_{11} (\tilde{\Psi} + I_k)^{1/2} & (\tilde{\Sigma} + \delta^2 I_q)^{1/2} E_{12} \\ \delta E_{21} (\tilde{\Psi} + I_k)^{1/2} & \delta E_{22} \end{pmatrix},$$

where $\tilde{X} = \tilde{A}\tilde{B}\tilde{C} + \tilde{D}E\tilde{F}$. Hence, X_{12} and X_{21} can be expressed as

$$X_{12} = (\tilde{\Sigma} + \delta^2 I_q)^{1/2} E_{12}, \quad X_{21} = \delta E_{21}(\tilde{\Psi} + I_k)^{1/2},$$

where $E_{12} \sim N_{q,n-k}(O_{q,n-k}, I_q, I_{n-k})$ and $E_{21} \sim N_{p-q,k}(O_{p-q,k}, I_{p-q}, I_k)$. Note that $\tilde{\Sigma} = (A^\top A)^{1/2}\Sigma(A^\top A)^{1/2}$ and $\tilde{\Psi} = (CC^\top)^{1/2}\Psi(CC^\top)^{1/2}$. Define

$$W_{12} = E_{12}E_{12}^\top \sim W_q(I_q, n-k), \quad W_{21} = E_{21}^\top E_{21} \sim W_k(I_k, p-q). \tag{20}$$

Then, $\lambda_{q_\sigma}^{(12)}$ and $\lambda_{k_\psi}^{(21)}$ can be expressed as follows:

$$\begin{aligned} \lambda_{q_\sigma}^{(12)} &= \lambda_{q_\sigma}(X_{12}X_{12}^\top) = \lambda_{q_\sigma}((\tilde{\Sigma} + \delta^2 I_q)W_{12}), \\ \lambda_{k_\psi}^{(21)} &= \lambda_{k_\psi}(X_{21}X_{21}^\top) = \delta^2 \lambda_{k_\psi}((\tilde{\Psi} + I_k)W_{21}). \end{aligned} \tag{21}$$

Here, let us fix $q_\sigma > q_\sigma^*$ and $k_\psi > k_\psi^*$. Because $\lambda_{q_\sigma}(\Sigma) = \lambda_{k_\psi}(\Psi) = 0$, Lemma 3 yields that $\lambda_{q_\sigma}(\tilde{\Sigma}) = \lambda_{k_\psi}(\tilde{\Psi}) = 0$. Combining this result, (21) and Lemma 3, we have

$$\begin{aligned} \delta^2 \lambda_q(W_{12}) \le \lambda_{q_\sigma}^{(12)} \le \delta^2 \lambda_1(W_{12}), \\ \delta^2 \lambda_k(W_{21}) \le \lambda_{k_\psi}^{(21)} \le \delta^2 \lambda_1(W_{21}). \end{aligned} \tag{22}$$

Because $W_{12}/n \xrightarrow{p} I_q$ as $n \to \infty$ and $W_{21}/p \xrightarrow{p} I_k$ as $p \to \infty$, we can see that $\lambda_1(W_{12})/n$, $\lambda_q(W_{12})/n \xrightarrow{p} 1$ and $\lambda_1(W_{21})/p$, $\lambda_k(W_{21})/p \xrightarrow{p} 1$ when n and p go to infinity, respectively. Therefore, (22) indicates that

$$\begin{aligned} &\lambda_{q_\sigma}^{(12)}/n \xrightarrow{p} \delta^2 \;\text{ for (i), (ii)},\; \lambda_{q_\sigma}^{(12)}/n = O_p(1) \;\text{ for (iii)}, \\ &\lambda_{k_\psi}^{(21)}/p \xrightarrow{p} \delta^2 \;\text{ for (i), (iii)},\; \lambda_{k_\psi}^{(21)}/p = O_p(1) \;\text{ for (ii)}. \end{aligned}$$

On the other hand, $\mathrm{tr}(X_{22}X_{22}^\top)/\delta^2 \sim \chi^2_{(n-k)(p-q)}$ and $\chi^2_{(n-k)(p-q)}/(n-k)(p-q) \xrightarrow{p} 1$ for (i)–(iii). Hence, it holds that

$$\hat{\delta}^2(q_\sigma, k_\psi) = \frac{\mathrm{tr}(X_{22}X_{22}^\top) + \sum_{s>q_\sigma} \lambda_s^{(12)} + \sum_{t>k_\psi} \lambda_t^{(21)}}{np - nq_\sigma - k_\psi p} \xrightarrow{p} \delta^2 \tag{23}$$

for (i)–(iii) when $q_\sigma \geq q_\sigma^*$ and $k_\psi \geq k_\psi^*$.

Next, we obtain the lower bounds of $\lambda^{(12)}_{q_\sigma^*}$ and $\lambda^{(21)}_{k_\psi^*}$. Using Lemma 3 and the condition (11), we have

$$\lambda_{q_\sigma^*}(\tilde{\Sigma} + \delta^2 I_q) \geq \lambda_q(A^\top A)\lambda_{q_\sigma^*}(\Sigma) + \delta^2 \geq p\zeta_a \lambda_{q_\sigma^*}(\Sigma) + \delta^2,$$
$$\lambda_{k_\psi^*}(\tilde{\Psi} + I_k) \geq \lambda_k(CC^\top)\lambda_{k_\psi^*}(\Psi) + 1 \geq n\zeta_c \lambda_{k_\psi^*}(\Psi) + 1.$$

By applying this evaluation and Lemma 3 into (21), it follows that

$$\begin{aligned} \lambda^{(12)}_{q_\sigma^*} &\geq \lambda_{q_\sigma^*}(\tilde{\Sigma} + \delta^2 I_q)\lambda_q(W_{12}) \geq \{p\zeta_a\lambda_{q_\sigma^*}(\Sigma) + \delta^2\}\lambda_q(W_{12}), \\ \lambda^{(21)}_{k_\psi^*} &\geq \delta^2 \lambda_{k_\psi^*}(\tilde{\Psi} + I_k)\lambda_k(W_{21}) \geq \delta^2\{n\zeta_c\lambda_{k_\psi^*}(\Psi) + 1\}\lambda_k(W_{21}). \end{aligned} \tag{24}$$

If $\lambda_q(W_{12})/n \geq \{\zeta_a\lambda_{q_\sigma^*}(\Sigma)/2 + \delta^2\}/\{p\zeta_a\lambda_{q_\sigma^*}(\Sigma) + \delta^2\}$, then it follows from (24) that $\lambda^{(12)}_{q_\sigma^*}/n \geq \zeta_a\lambda_{q_\sigma^*}(\Sigma)/2 + \delta^2$. Here, we consider the case when $n \to \infty$, i.e., (i) and (ii). Because $\{\zeta_a\lambda_{q_\sigma^*}(\Sigma)/2 + \delta^2\}/\{p\zeta_a\lambda_{q_\sigma^*}(\Sigma) + \delta^2\} \leq 1 + \zeta_a\lambda_{q_\sigma^*}(\Sigma)/(2\delta^2) < 1$ and $\lambda_q(W_{12})/n \xrightarrow{p} 1$, it holds that for (i) and (ii),

$$Pr(\lambda_q(W_{12})/n \geq \{\zeta_a\lambda_{q_\sigma^*}(\Sigma)/2 + \delta^2\}/\{p\zeta_a\lambda_{q_\sigma^*}(\Sigma) + \delta^2\}) \to 1.$$

Hence, we can see that

$$Pr(\lambda^{(12)}_{q_\sigma^*}/n \geq \zeta_a\lambda_{q_\sigma^*}(\Sigma)/2 + \delta^2) \to 1, \tag{25}$$

for (i) and (ii). In the case (iii), because $\lambda_{q_\sigma^*}(\Sigma) > 0$,

$$Pr(\lambda_q(W_{12})/n \geq \{\zeta_a \lambda_{q_\sigma^*}(\Sigma)/2 + \delta^2\}/\{p\zeta_a \lambda_{q_\sigma^*}(\Sigma) + \delta^2\}) \to 1.$$

Hence (25) holds for (iii). In a similar way, for (i)–(iii), it follows that

$$Pr(\lambda_{k_\psi^*}^{(21)}/p \geq \delta^2\{\zeta_c \lambda_{k_\psi^*}(\Psi)/2 + 1\}) \to 1. \tag{26}$$

Combining (23), (25) and (26), it can be seen that $Pr((q_\sigma^*, k_\psi^*) \in \mathcal{L}) \to 1$ for (i)–(iii). Thus, for all $q_\sigma < q_\sigma^*$ and $k_\psi < k_\psi^*$, $Pr((q_\sigma, k_\psi) \notin \mathcal{L}_R) \to 1$ for all (i)–(iii).

For the proof of the lemma, it suffices to show that (q_σ, k_ψ), which satisfies $q_\sigma \geq q_\sigma^*, k_\psi < k_\psi^*$ or $q_\sigma < q_\sigma^*, k_\psi \geq k_\psi^*$, does not belong to $\mathcal{L}_R$ with a probability tending to 1. Because of symmetry, we only consider the case $q_\sigma \geq q_\sigma^*$ and $k_\psi < k_\psi^*$. Suppose that there exists $(q_\sigma, k_\psi) \in \mathcal{L}$ such that $q_\sigma \geq q_\sigma^*$ and $k_\psi < k_\psi^*$. Then, from the definition of $\mathcal{L}$, it follows that

$$\hat{\delta}^2(q_\sigma, k_\psi) \leq \frac{\lambda_{q_\sigma}^{(12)}}{n}, \quad \hat{\delta}^2(q_\sigma, k_\psi) \leq \frac{\lambda_{k_\psi}^{(21)}}{p}.$$

On the other hand, because $k_\psi < k_\psi^*$ is assumed, if $(q_\sigma, k_\psi^*) \in \mathcal{L}$, that is,

$$\hat{\delta}^2(q_\sigma, k_\psi^*) \leq \frac{\lambda_{q_\sigma}^{(12)}}{n}. \quad \hat{\delta}^2(q_\sigma, k_\psi^*) \leq \frac{\lambda_{k_\psi^*}^{(21)}}{p},$$

then $(q_\sigma, k_\psi) \notin \mathcal{L}_R$. From the previous result presented in (23), $\hat{\delta}^2(q_\sigma, k_\psi^*) \xrightarrow{p} \delta^2$. This and (26) indicate that

$$Pr(\hat{\delta}^2(q_\sigma, k_\psi^*) \leq \lambda_{k_\psi^*}^{(21)}/p) \to 1 \tag{27}$$

for (i)–(iii). Moreover, it follows from (18) with $a = np - nq_\sigma - k_\psi^* p$, $b = \mathrm{tr}(X_{22}X_{22}^\top) +$

$\sum_{s>q_\sigma} \lambda_s^{(12)} + \sum_{t>k_\psi^*} \lambda_t^{(21)}$, $c = (k_\psi - k_\psi^*)p$ and $d = \sum_{k_\psi<t\le k_\psi^*} \lambda_t^{(21)}$ that

$$\frac{b}{a} = \hat{\delta}^2(q_\sigma, k_\psi^*) \le \frac{\lambda_{k_\psi^*}^{(21)}}{p} \le \frac{\sum_{k_\psi<t\le k_\psi^*} \lambda_t^{(21)}}{(k_\psi - k_\psi^*)p} = \frac{d}{c}$$
$$\Rightarrow \hat{\delta}^2(q_\sigma, k_\psi^*) = \frac{b}{a} \le \frac{b+d}{a+c} = \hat{\delta}^2(q_\sigma, k_\psi).$$

Hence, (27) implies that $Pr(\hat{\delta}^2(q_\sigma, k_\psi^*) \le \hat{\delta}^2(q_\sigma, k_\psi)) \to 1$ for (i)–(iii). Since the assumption $(q_\sigma, k_\psi) \in \mathcal{L}$ indicates that $\hat{\delta}^2(q_\sigma, k_\psi) \le \lambda_{q_\sigma}^{(12)}/n$, the inequality $\hat{\delta}^2(q_\sigma, k_\psi^*) \le \hat{\delta}^2(q_\sigma, k_\psi)$ yields $\hat{\delta}^2(q_\sigma, k_\psi^*) \le \lambda_{q_\sigma}^{(12)}/n$. From (27), it follows that

$$Pr(\hat{\delta}^2(q_\sigma, k_\psi^*) \le \lambda_{q_\sigma}^{(12)}/n) \to 1 \tag{28}$$

for (i)–(iii). It is established from (27) and (28) that

$$Pr(q_\sigma \ge q_\sigma^*, k_\psi < k_\psi^* \Rightarrow (q_\sigma, k_\psi) \notin \mathcal{L}_R) \to 1,$$

for (i)–(iii). Thus, the proof is completed. □

Proof of Lemma 2

Proof. Fix $q_\sigma \ge q_\sigma^*$ and $k_\psi \ge k_\psi^*$. Recall that

$$\hat{\Sigma}(q_\sigma, k_\psi) = (A^\top A)^{-1/2}\{\hat{\Sigma}_*(q_\sigma, k_\psi) - \hat{\delta}^2(q_\sigma, k_\psi)I_q\}(A^\top A)^{-1/2},$$
$$\hat{\Sigma}_*(q_\sigma, k_\psi) = Q_{12}^\top \mathrm{diag}\{\lambda_1^{(12)}/n, \dots, \lambda_{q_\sigma}^{(12)}/n, \hat{\delta}^2(q_\sigma, k_\psi), \dots, \hat{\delta}^2(q_\sigma, k_\psi)\}Q_{12},$$

where Q_{12} is an orthogonal matrix, which is used to diagonalize: $n\hat{\Sigma}_*(q,k) = (\tilde{\Sigma}+\delta^2 I_q)^{1/2} W_{12} (\tilde{\Sigma}+\delta^2 I_q)^{1/2} = Q_{12}^\top \mathrm{diag}\{\lambda_1^{(12)}, \dots, \lambda_q^{(12)}\}Q_{12}$. Note that $\tilde{\Sigma} = (A^\top A)^{1/2}\Sigma(A^\top A)^{1/2}$ and $W_{12} \sim$

$W_q(I_q, n-k)$ defined in (20). Here, $\hat{\Sigma}(q_\sigma, k_\psi) - \Sigma$ is evaluated by

$$
\begin{aligned}
\hat{\Sigma}(q_\sigma, k_\psi) - \Sigma &= (A^\top A)^{-1/2}\hat{\Sigma}_*(q_\sigma, k_\psi)(A^\top A)^{-1/2} - \hat{\delta}^2(q_\sigma, k_\psi)(A^\top A)^{-1} - \Sigma \\
&= (A^\top A)^{-1/2}\{\hat{\Sigma}_*(q_\sigma, k_\psi) - \hat{\Sigma}_*(q, k)\}(A^\top A)^{-1/2} \\
&\quad + (A^\top A)^{-1/2}\hat{\Sigma}_*(q, k)(A^\top A)^{-1/2} - \Sigma - \delta^2(A^\top A)^{-1} \\
&\quad + \{\delta^2 - \hat{\delta}^2(q_\sigma, k_\psi)\}(A^\top A)^{-1}.
\end{aligned}
$$

Therefore, an upper bound of $\|\hat{\Sigma}(q_\sigma, k_\psi) - \Sigma\|_2$ can be obtained as follows:

$$
\begin{aligned}
\|\hat{\Sigma}(q_\sigma, k_\psi) - \Sigma\|_2 \leq \max_{q_\sigma < s \leq q} &|\hat{\delta}^2(q_\sigma, k_\psi) - \lambda_s^{(12)}/n| \|(A^\top A)^{-1}\|_2 \\
&+ \|(A^\top A)^{-1/2}\hat{\Sigma}_*(q, k)(A^\top A)^{-1/2} - \Sigma - \delta^2(A^\top A)^{-1}\|_2 \\
&+ |\delta^2 - \hat{\delta}^2(q_\sigma, k_\psi)| \|(A^\top A)^{-1}\|_2.
\end{aligned}
$$

For (i) and (ii), the limits $\hat{\delta}^2(q_\sigma, k_\psi) \xrightarrow{p} \delta^2$ $(q_\sigma \geq q_\sigma^*,\ k_\psi \geq k_\psi^*)$ and $\lambda_s^{(12)}/n \xrightarrow{p} \delta^2$ $(s > q_\sigma^*)$ have been shown in the proof of Lemma 1. Moreover, because $W_{12}/n \xrightarrow{p} I_q$, it holds that

$$
\begin{aligned}
&\|(A^\top A)^{-1/2}\hat{\Sigma}_*(q, k)(A^\top A)^{-1/2} - \Sigma - \delta^2(A^\top A)^{-1}\|_2 \\
&\leq \|\Sigma + \delta^2(A^\top A)^{-1}\|_2 \|W_{12}/n - I_q\|_2 \xrightarrow{p} 0.
\end{aligned}
$$

Hence, for (i) and (ii), $\|\hat{\Sigma}(q_\sigma, k_\psi) - \Sigma\|_2 \xrightarrow{p} 0$.

In a similar way to the above presentation, we can verify the convergence of $\hat{\Psi}(q_\sigma, k_\psi)$ to Ψ for (i) and (iii). Hence, the proof is completed. $\square$