Estimation for an additive growth curve model with orthogonal design matrices

JIANHUA HU[1], GUOHUA YAN[2] and JINHONG YOU[1]

[1]*School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, P.R. China. E-mail: frank.jianhuahu@gmail.com; johnyou07@gmail.com*
[2]*Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, E3B 5A3 Canada. E-mail: gyan@unb.ca*

An additive growth curve model with orthogonal design matrices is proposed in which observations may have different profile forms. The proposed model allows us to fit data and then estimate parameters in a more parsimonious way than the traditional growth curve model. Two-stage generalized least-squares estimators for the regression coefficients are derived where a quadratic estimator for the covariance of observations is taken as the first-stage estimator. Consistency, asymptotic normality and asymptotic independence of these estimators are investigated. Simulation studies and a numerical example are given to illustrate the efficiency and parsimony of the proposed model for model specifications in the sense of minimizing Akaike's information criterion (AIC).

Keywords: AIC; asymptotic normality; consistent estimator; growth curve model; quadratic estimator; two-stage generalized least squares

1. Introduction

In a variety of areas, observations are measured over multiple time points on a particular characteristic to investigate the temporal pattern of change on the characteristic. The observations of repeated measurements are usually analyzed by the growth curve model (GCM), initiated by Potthoff and Roy [14]. Since then, parameter estimation, hypothesis testing and prediction of future values have been investigated by numerous researchers, generating a substantial amount of literature, including [2,3,7,8,10,11,15,16]. The basic idea of the growth curve model is to introduce some known functions, usually polynomial functions, so as to capture patterns of change for time-dependent measurements. We shall generalize the growth curve model to the case where observations of time-dependently repeated measurements may have polynomial functions with different degrees rather than polynomial functions with a common degree. In this article, different profile forms mean polynomial functions with different degrees and a profile form means polynomial functions with a common degree.

To motivate it, let us look at the following situation. We have many groups of animals, with each group being subjected to a different treatment. Animals in all groups are measured at the same p time points and assumed to have the same covariance matrix Σ. The growth curve associated with the ith group is $\theta_{i0} + \theta_{i1}t + \theta_{i2}t^2 + \cdots + \theta_{iq_i}t^{q_i}$, implying that the growth curves may have different profiles, say k profiles, not necessarily one profile. There are m_i groups that

have the same profile form with index i and n_i individuals in total. Here $n = \sum_{i=1}^{k} n_i$. The simplest situation is that each group has a different profile form. Assume that there are k groups of individuals and p observing time points such that $k + p \leq n$. For $i = 1, 2, \ldots, k$, put

$$
Z_i = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{q_i-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{q_i-1} \\ . & . & . & \cdots & . \\ 1 & t_p & t_p^2 & \cdots & t_p^{q_i-1} \end{bmatrix},
$$
$$
\Theta_i = (\theta_{i0}, \theta_{i1}, \theta_{i2}, \ldots, \theta_{iq_i-1}),
$$

and

$$
X_i = (x_{i1}, x_{i2}, \ldots, x_{in})' \in \mathbb{R}^n,
$$

where $x_{i(p_i+j)} = 1$ for $j = 1, 2, \ldots, n_i$ with $p_0 = 0$, $p_i = \sum_{j=1}^{i-1} n_j$ and other x'_{ij}'s are 0.

Generalizing the above situation we propose the following additive growth curve model

$$
Y = \sum_{i=1}^{k} X_i \Theta_i Z_i' + \mathcal{E}, \qquad \mathcal{E} \sim \mathcal{G}(\mathbf{0}, I \otimes \Sigma), \tag{1}
$$

with orthogonal design matrices or mutually orthogonal column spaces of design matrices, defined as

$$
\operatorname{rank}(X_i) + p \leq n \quad \text{and} \quad \mathscr{C}(X_i) \perp \mathscr{C}(X_j) \quad \text{or} \quad X_i' X_j = \mathbf{0} \text{ for any distinct } i, j, \tag{2}
$$

where Y is an $n \times p$ matrix of observations; X_i, Z_i ($1 \leq i \leq k$) are known $n \times m_i$ ($n > m_i$) full-rank design matrices and $p \times q_i$ ($p > q_i$) full-rank profile matrices, respectively; Θ_i ($1 \leq i \leq k$) are unknown $m_i \times q_i$ matrices of the regression coefficients; $\mathscr{C}(X)$ denotes the column space of the matrix X; $\mathcal{G}$ is a general continuous type distribution function; observations on individuals are independent; and the rows of the random error matrix $\mathcal{E}$ are independent and identically distributed with mean zero and a common unknown covariance matrix Σ of order p.

The model (1) subject to (2) will be demonstrated to have an advantage that it fits data in a more parsimonious way than the traditional growth curve model in the situation where model specification is needed. In the above stated example of animals, the traditional growth curve model assumes that all observations have the same profile form, which may cause the model misspecification, underfitting or overfitting.

On the other hand, Kollo and von Rosen [9], in Chapter 4, investigated an additive growth curve model with nested column spaces generalized by design matrices, that is, constraint $\mathscr{C}(X_1) \supseteq \mathscr{C}(X_2) \supseteq \cdots \supseteq \mathscr{C}(X_k)$ with $\operatorname{rank}(X_1) + p \leq n$, usually called the extended growth curve model. Obviously, there is not an inclusion relationship between the extended growth curve model and the proposed model (1) with (2) because the constraint of nested column spaces and the constraint of orthogonal column spaces have no inclusion relationship. An extension of the growth curve model proposed in [17] did not include the proposed model (1) with (2), either.

This paper will investigate estimation of parameters and properties of the corresponding estimators in the proposed model (1) with (2), including consistency and asymptotic normality.

The organization of the paper is as follows. Two-stage generalized least-squares estimators of the regression coefficients are obtained in Section 2. Both the consistency of the estimators for the regression coefficients and a quadratic estimator for the unknown covariance are investigated in Section 3, while their asymptotic normalities under certain conditions are investigated in Section 4. Simulation studies are given in Section 5. A numerical example is explored to illustrate our techniques in Section 6. Finally, brief concluding remarks are stated in Section 7.

Throughout this paper, the following notations are used. $\mathscr{M}_{n \times p}$ denotes the set of all $n \times p$ matrices over real set $\mathbb{R}$ with trace inner product $\langle , \rangle$ and $\| \cdot \|$ denotes the trace norm on the set $\mathscr{M}_{n \times p}$. $\mathrm{tr}(A)$ denotes the trace of matrix A and I_n denotes the identity matrix of order n. For an $n \times p$ matrix Y, we write $Y = [\mathbf{y}'_1, \ldots, \mathbf{y}'_n]'$, $\mathbf{y}'_i \in \mathbb{R}^p$, where $\mathbb{R}^p$ is the p-dimensional real space and $\mathrm{vec}(Y)$ denotes np-dimensional vector $[\mathbf{y}_1, \ldots, \mathbf{y}_n]'$. Here the vec operator transforms a matrix into a vector by stacking the rows of the matrix one underneath another. $Y \sim \mathcal{G}(\boldsymbol{\mu}, I \otimes \Sigma)$ means that Y follows a general continuous type distribution $\mathcal{G}$ with $\mathrm{E}(Y) = \boldsymbol{\mu}$ and $\mathrm{E}(Y - \boldsymbol{\mu})(Y - \boldsymbol{\mu})' = I \otimes \Sigma$. The Kronecker product $A \otimes B$ of matrices A and B is defined to be $A \otimes B = (a_{ij}B)$. Then we have $\mathrm{vec}(ABC) = (A \otimes C') \mathrm{vec}(B)$. Let A^+ denote the Moore–Penrose inverse of A. $P_X = X(X'X)^- X'$ denotes the orthogonal projection onto the column space $\mathscr{C}(X)$ of a matrix X. $M_X = I - X(X'X)^- X'$ is the orthogonal projection onto the orthogonal complement $\mathscr{C}(X)^\perp$ of $\mathscr{C}(X)$.

2. Two-stage generalized least-squares estimators

Recall that the regression coefficients, $\Theta_1, \ldots, \Theta_k$, in the model (1) are defined before a design is planned and observation Y is obtained. Thus the rows of the design matrices, $X_1, \ldots, X_k$, are added one after another and the profile forms, $Z_1, \ldots, Z_k$, do not depend on the sample size n. So, we shall only consider the case of full-rank X_i's and Z_i's in the present paper.

Set

$$\boldsymbol{\mu} = \sum_{i=1}^{k} X_i \Theta_i Z'_i. \tag{3}$$

Equation (3) is said to be the mean structure of the model (1).

A statistic $\hat{\boldsymbol{\mu}}_{\mathrm{gls}}(Y)$ is said to be the generalized least-squares (GLS) estimator of parameter matrix $\boldsymbol{\mu}$ if the minimum value of function $\langle Y - \boldsymbol{\mu}, Y - \boldsymbol{\mu} \rangle$ is attained at the point $\mathbf{u} = \hat{\boldsymbol{\mu}}_{\mathrm{gls}}(Y)$, where the inner product $\langle , \rangle$ or the trace norm $\| \cdot \|$ associated with the covariance $I \otimes \Sigma$ of $Y : \langle \mathbf{w}_1, \mathbf{w}_2 \rangle = \mathrm{vec}(\mathbf{w}_2)'(I \otimes \Sigma)^{-1} \mathrm{vec}(\mathbf{w}_1)$ with $\| \mathbf{w} \| = \langle \mathbf{w}, \mathbf{w} \rangle^{1/2}$ and $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in \mathscr{M}_{n \times p}$.

Generally speaking, we actually know nothing or very little about the covariance Σ of observations of repeated measurements before we measure these observations. So, alternatively, a two-stage estimation is used to find an estimator of $\boldsymbol{\mu}$, denoted by $\hat{\boldsymbol{\mu}}_{\mathrm{2sgls}}(Y)$. The two-stage estimation procedure is as follows: First, based on the observation Y, find a first-stage estimator $\widetilde{\Sigma}$ of Σ. Second, replace the unknown Σ with the first-stage estimator $\widetilde{\Sigma}$ and then find $\widehat{\boldsymbol{\mu}}_{\mathrm{2sgls}}(Y)$ through the GLS method. For convenience, we shall omit the subscript of $\hat{\boldsymbol{\mu}}_{\mathrm{2sgls}}(Y)$.

In order to get a good first-stage estimator $\widetilde{\Sigma}$ of Σ, let us have a close look at the following quadratic statistic (a quadratic form without associating with parameters):

$$\widehat{\Sigma}(Y) = Y'WY, \qquad W \equiv \frac{1}{r}\left(I - \sum_{i=1}^{k} P_{X_i}\right), \tag{4}$$

where $r = n - \sum_{i=1}^{k} \text{rank}(X_i)$.

(1) The statistic $\widehat{\Sigma}(Y)$ is easily proven to be positive definite with probability 1; see Theorem 3.1.4 of [13]. So, $\widehat{\Sigma}^{-1}(Y)$ exists with probability 1.

(2) Under the assumption of normality, the quadratic estimator $\widehat{\Sigma}(Y)$ given by equation (4) follows a Wishart distribution; see [4].

(3) $\widehat{\Sigma}(Y)$ is an unbiased invariant estimator of Σ; see [5]. A similar result for the growth curve model was obtained by Žežula [18].

It follows from the above properties that the statistic $\widehat{\Sigma}(Y)$ seems to be a very good candidate for the first-stage estimator. As a consequence, it will be taken as the first-stage estimator $\widetilde{\Sigma}$ in our subsequent discussion.

For $i = 1, \ldots, k$, let

$$H_i(Y) \equiv \widehat{\Sigma}^{-1}(Y)Z_i(Z_i'\widehat{\Sigma}^{-1}(Y)Z_i)^{-1}Z_i' = \widehat{\Sigma}^{-1}(Y)(P_{Z_i}\widehat{\Sigma}^{-1}(Y)P_{Z_i})^{+}. \tag{5}$$

Then, we easily see that

$$Z_i'H_i(Y) = Z_i', \qquad i = 1, \ldots, k. \tag{6}$$

When $\widehat{\Sigma}(Y)$ is taken as the first-stage estimator, the following lemma provides the explicit expression of the two-stage GLS estimators both for mean matrix $\boldsymbol{\mu}$ and the regression coefficients, $\Theta_1, \ldots, \Theta_k$. Furthermore, under certain conditions, these estimators are unbiased.

Theorem 2.1. *Consider $\Sigma = \widehat{\Sigma}(Y)$ for the model* (1) *subject to* (2). *The following statements hold.*

(1) *The two-stage GLS estimator $\hat{\boldsymbol{\mu}}(Y)$ of $\boldsymbol{\mu}$ is given by*

$$\hat{\boldsymbol{\mu}}(Y) = \sum_{i=1}^{k} P_{X_i} Y \Sigma^{-1}(Y)(P_{Z_i}\Sigma^{-1}(Y)P_{Z_i})^{+} = \sum_{i=1}^{k} P_{X_i} Y H_i(Y). \tag{7}$$

(2) *The two-stage GLS estimator $\widehat{\Theta}_i(Y)$ of Θ_i is given by*

$$\widehat{\Theta}_i(Y) = (X_i'X_i)^{-1}X_i'Y H_i(Y)Z_i(Z_i'Z_i)^{-1}. \tag{8}$$

(3) *If the distribution of $\mathcal{E}$ is symmetric about the origin $\boldsymbol{0}$, the statistic $\hat{\boldsymbol{\mu}}(Y)$ is an unbiased estimator of mean $\boldsymbol{\mu}$. Moreover, for each i, the statistic $\widehat{\Theta}_i(Y)$ is an unbiased estimator of the regression coefficients Θ_i.*

The proof of Theorem 2.1 is deferred to the Appendix.

3. Consistency

Since Y is associated with the sample size n, we shall use Y_n to replace Y in (4)–(8) and then investigate the consistency of the estimator $\widehat{\Sigma}(Y_n)$ and the estimators, $\widehat{\Theta}_1(Y), \ldots, \widehat{\Theta}_k(Y)$, as the sample size n tends to infinity. Note that X and $\mathcal{E}$ are also associated with the sample size n.

Regarding the consistency of the quadratic estimator $\widehat{\Sigma}(Y_n)$, we have the following result.

Theorem 3.1. *For the model* (1) *subject to* (2), *the statistic* $\widehat{\Sigma}(Y_n)$ *defined by* (4) *is a consistent estimator of the covariance matrix* Σ.

Proof. It follows from the invariance of statistic $\widehat{\Sigma}(Y)$ that $\widehat{\Sigma}(Y) = \widehat{\Sigma}(\mathcal{E})$. And $\widehat{\Sigma}(Y)$ can be rewritten as

$$\widehat{\Sigma}(Y_n) = \widehat{\Sigma}(\mathcal{E}) = \frac{n}{n-m}\left(\frac{1}{n}\sum_{l=1}^{n}\mathcal{E}_l\mathcal{E}_l' - \frac{1}{n}\mathcal{E}'\sum_{i=1}^{k}P_{X_i}\mathcal{E}\right), \tag{9}$$

where $m = \sum_{i=1}^{k}\mathrm{rank}(X_i)$ and $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_n)' \sim \mathcal{G}(\mathbf{0}, I_n \otimes \Sigma)$.

Note that $(\mathcal{E}_l\mathcal{E}_l')_{l=1}^{n}$ is a random sample from a population with mean $\mathrm{E}(\mathcal{E}_l\mathcal{E}_l') = \Sigma$. Kolmogorov's strong law of large numbers tells us that

$$\frac{1}{n}\sum_{l=1}^{n}\mathcal{E}_l\mathcal{E}_l' \text{ converges almost surely to } \Sigma. \tag{10}$$

Let $\varepsilon > 0$. By Chebyshev's inequality and $\mathrm{E}(\mathcal{E}'\mathcal{E}) = \mathrm{tr}(I)\Sigma$, we have

$$P\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}P_{X_i}\mathcal{E}\right\| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}\mathrm{E}\left[\mathrm{tr}\left(\mathcal{E}'\sum_{i=1}^{k}P_{X_i}\mathcal{E}\right)\right] = \frac{1}{n\varepsilon^2}\mathrm{tr}\left(\mathrm{E}[\mathcal{E}\mathcal{E}']\sum_{i=1}^{k}P_{X_i}\right)$$

$$= \frac{1}{n\varepsilon^2}\mathrm{tr}\left(I_n\,\mathrm{tr}(\Sigma)\sum_{i=1}^{k}P_{X_i}\right) = \frac{1}{n\varepsilon^2}\mathrm{tr}\left(\sum_{i=1}^{k}P_{X_i}\right)\mathrm{tr}(\Sigma).$$

Since $\mathrm{tr}(\sum_{i=1}^{k}P_{X_i}) = \sum_{i=1}^{k}\mathrm{rank}(X_i)$ is a constant, $P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^{k}P_{X_i}\mathcal{E}\| \geq \varepsilon)$ tends to $\mathbf{0}$ as the sample size n tends to infinity. So

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{k}P_{X_i}\mathcal{E} \text{ converges in probability to } \mathbf{0}. \tag{11}$$

Since convergence almost surely implies convergence in probability, by (10) and (11), we obtain from (9) that $\widehat{\Sigma}(Y_n)$ converges in probability to Σ, which completes the proof. $\square$

Assumption 1. *For* $l = 1, \ldots, k$,

$$\lim_{n\to\infty} n^{-1}X_l'X_l = R_l, \tag{12}$$

where R_l is positive definite.

For convenience, we restate Lemma 3.2 of [6] as follows.

Lemma 3.2. *For $i \in \{1, \ldots, k\}$, $H_i(Y_n)$ converges in probability to $H_i \equiv \Sigma^{-1}(P_{Z_i}\Sigma^{-1}P_{Z_i})^{+}$.*

On the consistency of the estimators of the regression coefficients $\Theta_i(Y_n)$s, we obtain the following theorem.

Theorem 3.3. *For any fixed $i \in \{1, \ldots, k\}$, with Assumption 1, the statistic $\widehat{\Theta}_i(Y_n)$ is a consistent estimator of the regression coefficient Θ_i.*

Proof. Fix $i \in \{1, \ldots, k\}$. By equation (8), we obtain the following equation:

$$\widehat{\Theta}_i(Y_n) = \Theta_i + S_i \mathcal{E} H_i(Y_n) K_i, \tag{13}$$

where $S_i = (X_i'X_i)^{-1}X_i'$ and $K_i = Z_i(Z_i'Z_i)^{-1}$. The second term of the right side in (13) can be rewritten as

$$S_i \mathcal{E} H_i(Y_n) K_i = n(X_i'X_i)^{-1}\left(\frac{1}{\sqrt{n}}X_i'\right)\left(\frac{1}{\sqrt{n}}P_{X_i}\mathcal{E}\right)H_i(Y_n)K_i.$$

By condition (12), $X_i'/\sqrt{n}$ are bounded. In fact, the elements of $X_i'/\sqrt{n}$ are at most of order $n^{-1/2}$ (see the proof of Lemma 4.1 below). So by (11), (12), Lemma 3.2 and Theorem 11.2.12 of [12], the second term of the right side in (13) converges in probability to $\mathbf{0}$. Thus, $\widehat{\Theta}_i(Y_n)$ converges in probability to Θ_i, which completes the proof. $\square$

In order to prove the consistency of the estimators $\widehat{\Theta}_1(Y_n), \widehat{\Theta}_2(Y_n), \ldots, \widehat{\Theta}_k(Y_n)$, the conditions $\lim_{n\to\infty} n^{-1}X_l'X_l = R_l$ for $l = 1, \ldots, k$ have been used in Theorem 3.3. We imagine that for each new observation, a new row is added to the matrices X_l and that the earlier rows remain intact in such a way that, for $l = 1, 2, \ldots, k$, the $m_l \times m_l$ elements of $X_l'X_l$ are O(n). In addition, we exclude the possibility that the limits of $n^{-1}X_l'X_l$s are singular.

4. Asymptotic normality

We have investigated the consistency of the estimators $\widehat{\Sigma}(Y)$ and $\widehat{\Theta}_i(Y_n)$ in the preceding section. In this section, we shall investigate the asymptotic normality of $\sqrt{n}[\widehat{\Theta}_i(Y_n) - \Theta_i]$ and $\sqrt{n}[\widehat{\Sigma}(Y_n) - \Sigma]$ under certain conditions.

We need the following lemma in the proof of the subsequent results.

Lemma 4.1. *Let $S_i = (X_i'X_i)^{-1}X_i' \equiv (\mathbf{s}_{i1}, \mathbf{s}_{i2}, \ldots, \mathbf{s}_{in})_{m_i \times n}$, where $\mathbf{s}_{ij}$ is the jth column of X_i. Then, under condition (12), the m_i elements of $\sqrt{n}\mathbf{s}_{ij}$ are O($n^{-1/2}$) for any $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, n\}$.*

The proof of Lemma 4.1 is deferred to the Appendix.

Theorem 4.2. *Under Assumption* 1, *the random matrix* $\sqrt{n}S_i\mathcal{E}$ *converges in distribution to* $\mathcal{N}_{m_i \times p}(\mathbf{0}, R_i^{-1} \otimes \Sigma)$ *for any* $i \in \{1, \dots, k\}$.

Also, the proof of Theorem 4.2 is deferred to the Appendix.

Finally, by Theorem 4.2 and Slutsky's theorem, we obtain our main result on the asymptotic normality of $\sqrt{n}[\widehat{\Theta}_i(Y_n) - \Theta_i]$.

Theorem 4.3. *Under Assumption* 1, *the statistic* $\sqrt{n}[\widehat{\Theta}_i(Y_n) - \Theta_i]$ *converges in distribution to* $\mathcal{N}_{m_i \times q_i}(\mathbf{0}, R_i \otimes (Z_i'\Sigma Z_i)^{-1})$ *for any* $i \in \{1, \dots, k\}$.

Next, we shall investigate the asymptotic normality of the $\widehat{\Sigma}(Y)$. The fourth-order moment of the error matrix will be needed in the following discussion.

Assumption 2. $\mathrm{E}(\mathcal{E}_1) = \mathbf{0}$, $\mathrm{E}(\mathcal{E}_1\mathcal{E}_1') = \Sigma > \mathbf{0}$, $\mathrm{E}(\mathcal{E}_1 \otimes \mathcal{E}_1\mathcal{E}_1') = \mathbf{0}_{p^2 \times p}$ *and* $\mathrm{E}\|\mathcal{E}_1\|^4 < \infty$, *where* $\mathcal{E}_1'$ *is the first row vector of the error matrix* $\mathcal{E}$.

Theorem 4.4. *Under Assumptions* 1 *and* 2, *the following probability statements hold*:

(a) $\sqrt{n}(\widehat{\Sigma}(Y) - \Sigma)$ *converges to* $\mathcal{N}(\mathbf{0}, \mathrm{Cov}(\mathcal{E}_1' \otimes \mathcal{E}_1'))$ *in distribution.*
(b) *For each* i, $\sqrt{n}(\widehat{\Sigma}(Y) - \Sigma)$ *and* $\sqrt{n}(\widehat{\Theta}_i(Y) - \Theta_i)$ *are asymptotically independent.*
(c) *For any distinct* i, j, $\sqrt{n}(\widehat{\Theta}_i(Y) - \Theta_i)$ *and* $\sqrt{n}(\widehat{\Theta}_j(Y) - \Theta_j)$ *are independent.*

Proof. (a) $\sqrt{n}(\widehat{\Sigma}(Y) - \Sigma)$ can be decomposed into

$$\sqrt{n}(\widehat{\Sigma}(Y) - \Sigma) = \Delta_1 + \Delta_2 + \Delta_3,$$

where

$$\Delta_1 = \sqrt{n}\left(\frac{1}{n}\sum_{l=1}^{n}\mathcal{E}_l\mathcal{E}_l' - \Sigma\right),$$

$$\Delta_2 = \frac{m}{\sqrt{n}(n-m)}\sum_{l=1}^{n}\mathcal{E}_l\mathcal{E}_l',$$

$$\Delta_3 = -\frac{\sqrt{n}}{n-m}\mathcal{E}'\sum_{i=1}^{k}P_{X_i}\mathcal{E}.$$

Similar to the proof of conclusions (10) and (11) in Theorem 3.1, we easily obtain that Δ_2 and Δ_3 converges to $\mathbf{0}$ in probability 1.

Also by assumptions 1 and 2, Δ_1 converges to $\mathcal{N}(\mathbf{0}, \Phi_2)$ in distribution, where $\Phi_2 = \mathrm{Cov}(\mathcal{E}_1' \otimes \mathcal{E}_1')$. Thus, we have

$$\sqrt{n}\,\mathrm{vec}(\widehat{\Sigma}(Y) - \Sigma) = \mathrm{vec}(\Delta_1) + \mathrm{o}_P(\mathbf{1}).$$

Hence, $\sqrt{n}(\widehat{\Sigma}(Y) - \Sigma)$ converges to $\mathcal{N}(\mathbf{0}, \mathrm{Cov}(\mathcal{E}_1' \otimes \mathcal{E}_1'))$ in distribution.

(b) By equation (13), it suffices to prove the asymptotic independence between $\frac{1}{\sqrt{n}} \text{vec}(X_i'\mathcal{E})$ and $\sqrt{n} \, \text{vec}(\widehat{\Sigma}(Y) - \Sigma)$.

Let $Q_n = X_i'\mathcal{E} = (\mathbf{x}_1^i, \dots, \mathbf{x}_n^i)(\mathcal{E}_1, \dots, \mathcal{E}_n)'$. Then

$$\text{Cov}\left(\left(\frac{1}{\sqrt{n}} X_i'\mathcal{E}\right), \sqrt{n}(\widehat{\Sigma} - \Sigma)\right) = \text{Cov}\left(\left(\sum_{l=1}^{n} \mathbf{x}_l^i \mathcal{E}_l'\right), \left(\frac{1}{n}\sum_{l=1}^{n} \mathcal{E}_l \mathcal{E}_l' - \Sigma\right)\right) + o_P(1)$$

$$= \text{E}\left(\left(\sum_{l=1}^{n} \mathbf{x}_l^i \otimes \mathcal{E}_l'\right)\left(\frac{1}{n}\sum_{j=1}^{n} \mathcal{E}_j \otimes \mathcal{E}_j' - \Sigma\right)\right) + o_P(1).$$

According to Assumption 2, $\text{Cov}((\frac{1}{\sqrt{n}} X_i'\mathcal{E}), \sqrt{n}(\widehat{\Sigma}(Y) - \Sigma))$ converges to $\mathbf{0}$ in probability 1. It follows that the vectors $\frac{1}{\sqrt{n}} \text{vec}(X_i'\mathcal{E})$ and $\sqrt{n} \, \text{vec}(\widehat{\Sigma}(Y) - \Sigma)$ are asymptotically independent. Therefore, $\sqrt{n}(\widehat{\Sigma}(Y) - \Sigma)$ and $\sqrt{n}(\widehat{\Theta}_i(Y) - \Theta_i)$ also are asymptotically independent.

(c) For any distinct i, j, it follows from condition (2) that

$$\text{Cov}\left(\sqrt{n}\left(\widehat{\Theta}_i(Y) - \Theta_i\right), \sqrt{n}\left(\widehat{\Theta}_j(Y) - \Theta_j\right)\right) = \mathbf{0}.$$

We have completed the proofs of statements (A)–(C). $\qquad\square$

Sometimes, it is necessary to consider hypothesis tests of the form

$$\text{H}_i : C\Theta_i V' = \mathbf{0},$$

where C and V are, respectively, $s \times m_i$ and $t \times q_i$ constant matrices. In this case, Theorem 4.3 and Slutsky's theorem are explored to understand the asymptotic behavior of $\sqrt{n}(C\widehat{\Theta}_i(Y)V' - C\Theta_i V')$.

Corollary 4.5. *Under Assumption* 1, *if matrices* $C(X'X)^{-1}C'$ *and* $V(Z'\widehat{\Sigma}^{-1}(Y)Z)^{-1}V'$ *are non-singular, then the statistic*

$$(Cn(X'X)^{-1}C')^{-1/2}\sqrt{n}(C\widehat{\Theta}_i(Y)V')(V(Z'\widehat{\Sigma}^{-1}(Y)Z)^{-1}V')^{-1/2}$$

under H_i *converges in distribution to* $\mathcal{N}_{s\times t}(\mathbf{0}, I)$.

Therefore, if it is necessary to make the statistical inference about certain Θ_i in the model (1), we can take the normal distribution $\mathcal{N}_{s\times t}(0, I)$ as an approximate distribution of the statistic

$$(Cn(X'X)^{-1}C')^{-1/2}\sqrt{n}(C\widehat{\Theta}_i(Y)V')(V(Z'\widehat{\Sigma}^{-1}(Y)Z)^{-1}V')^{-1/2}$$

under H_i if the sample size is large. Moreover, due to (c) of Theorem 4.4, H_i and H_j can be considered independently.

5. Simulation studies

In this section, we shall use simulation to investigate the efficiency and parsimony of the model (1) subject to constraint (2), compared with the traditional growth curve model $Y = X\Theta Z' + \mathcal{E}$.

We take an example as follows. Suppose n patients are divided into two groups with numbers of patients n_1 and n_2, respectively. A certain measurement in an active drug trial is made on each of the n_1 patients taking a placebo and the n_2 patients taking the active drug at time points $t_1 = -1, t_2 = -0.5, t_3 = 0.5$ and $t_4 = 1$. Assume that the first n_1 observations come from normal distribution $\mathcal{N}(\boldsymbol{\mu}_1, \Sigma_0)$, where

$$\boldsymbol{\mu}_1 = (4 + 2t_1, 4 + 2t_2, 4 + 2t_3, 4 + 2t_4)$$

and

$$\Sigma_0 = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}, \qquad 0 < \rho < 1.$$

(The model with this Σ_0 is called the simplest serial correlation model in literature.) It means that the n_1 observations have a linear profile form of time points. The remaining n_2 observations come from normal distribution $\mathcal{N}(\boldsymbol{\mu}_2, \Sigma_0)$, where

$$\boldsymbol{\mu}_2 = (3 + 2t_1 + t_1^2 - t_1^3, 3 + 2t_2 + t_2^2 - t_2^3, 3 + 2t_3 + t_3^2 - t_3^3, 3 + 2t_4 + t_4^2 - t_4^3),$$

implying that the n_2 observations have a cubic polynomial profile form of time points.

Let

$$Z_1' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ t_1 & t_2 & t_3 & t_4 \end{pmatrix} \quad \text{and} \quad Z_2' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ t_1 & t_2 & t_3 & t_4 \\ t_1^2 & t_2^2 & t_3^2 & t_4^2 \\ t_1^3 & t_2^3 & t_3^3 & t_4^3 \end{pmatrix}$$

and

$$B_1 = (4 \quad 2), \qquad B_2 = (3 \quad 2 \quad -3 \quad 2),$$

then

$$\boldsymbol{\mu}_1 = B_1 Z_1' \quad \text{and} \quad \boldsymbol{\mu}_2 = B_2 Z_2'.$$

In real experiments, with observations Y, model specification is a challenging task.

Without loss of generality, we shall consider three approaches using the growth curve model to fit data of repeated measurements from the above synthetic example.

The first approach is to regard all observations of repeated measurements as having linear profile forms over multiple time points. In this scenario, model underfitting has occurred. The underfitted model is denoted by ψ_u,

$$\text{Model } \psi_u : Y = X\Theta_u Z_1' + \mathcal{E},$$

where $X = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \end{pmatrix}$ and $\Theta_u = \begin{pmatrix} \theta_{11}^u & \theta_{12}^u \\ \theta_{21}^u & \theta_{22}^u \end{pmatrix}$ to fit the n observations. By (b) of Lemma 2.1, the estimator $\widehat{\Theta}_u = (X'X)^{-1} X'Y \widehat{\Sigma}^{-1}(Y) Z_1 (Z_1' \widehat{\Sigma}^{-1}(Y) Z_1)^{-1}$.

The second approach is to regard all observations of repeated measurements as following cubic polynomial profile forms over multiple time points. In this case, model overfitting has occurred. The overfitted model is denoted by ψ_o,

$$\text{Model } \psi_o : Y = X \Theta_o Z_2' + \mathcal{E},$$

where $\Theta_o = \begin{pmatrix} \theta_{11}^o & \theta_{12}^o & \theta_{13}^o & \theta_{14}^o \\ \theta_{21}^o & \theta_{22}^o & \theta_{23}^o & \theta_{24}^o \end{pmatrix}$ to fit the n observations. The estimators of regression coefficients are $\widehat{\Theta}_o = (X'X)^{-1} X'Y \widehat{\Sigma}^{-1}(Y) Z_2 (Z_2' \widehat{\Sigma}^{-1}(Y) Z_2)^{-1}$.

The third approach is to regard the first n_1 observations as having a linear profile form of time points and the remaining n_2 observations as having a cubic polynomial profile form over multiple time points. In this case, model misspecification may not occur. The additive model is denoted by ψ_a,

$$\text{Model } \psi_a : Y = X_1 \Theta_1 Z_1' + X_2 \Theta_2 Z_2' + \mathcal{E},$$

where $X_1 = \begin{pmatrix} \mathbf{1}_{n_1} \\ \mathbf{0} \end{pmatrix}$, $X_2 = \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{n_2} \end{pmatrix}$, $\Theta_1 = \begin{pmatrix} \theta_{11} & \theta_{12} \end{pmatrix}$ and $\Theta_2 = \begin{pmatrix} \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} \end{pmatrix}$, to fit the n observations. Based on the above assumption, ψ_a actually is the true model. The estimators $\widehat{\Theta}_i = (X_i'X_i)^{-1} X_i'Y \widehat{\Sigma}^{-1}(Y) Z_i (Z_i' \widehat{\Sigma}^{-1}(Y) Z_i)^{-1}$ for $i = 1, 2$.

The model specification starts with residuals. We shall use a residual matrix R, defined as the difference between the observation Y and fitted mean $\widehat{Y}$, that is, $R = Y - \widehat{Y}$, to discuss the model specification for our example.

The residual matrix sum of squares (RMSS) is defined as the trace of $R'R$

$$\text{RMSS} = \|R\|^2 = \text{tr}\big((Y - \widehat{Y})'(Y - \widehat{Y})\big). \tag{14}$$

Usually, overfitting of model specification can provide a smaller RMSS as well as use more parameters. The residual matrix sum of squares and the number of parameters are two trade-off issues in model specification. Here, Akaike's information criterion (AIC) – see [1] – is explored to reward the decreasing RMSS and penalize overparametrization. Akaike's information criterion formula is given by

$$\text{AIC} = n \ln(\text{RMSS}) + 2(p + 1) - n \ln(n).$$

Specially, we have the following three AICs for the above chosen three models.

$$\text{AIC}_u = n \ln\big(\text{tr}\big((Y - X\widehat{\Theta}_u Z_1')'(Y - X\widehat{\Theta}_u Z_1')\big)\big) + 2(p_u + 1) - n \ln(n),$$
$$\text{AIC}_o = n \ln\big(\text{tr}\big((Y - X\widehat{\Theta}_o Z_2')'(Y - X\widehat{\Theta}_o Z_2')\big)\big) + 2(p_o + 1) - n \ln(n)$$

and

$$\text{AIC}_a = n \ln\big(\text{tr}\big((Y - X_1\widehat{\Theta}_1 Z_1' - X_2\widehat{\Theta}_2 Z_2')'(Y - X_1\widehat{\Theta}_1 Z_1' - X_2\widehat{\Theta}_2 Z_2')\big)\big) + 2(p_a + 1) - n \ln(n),$$
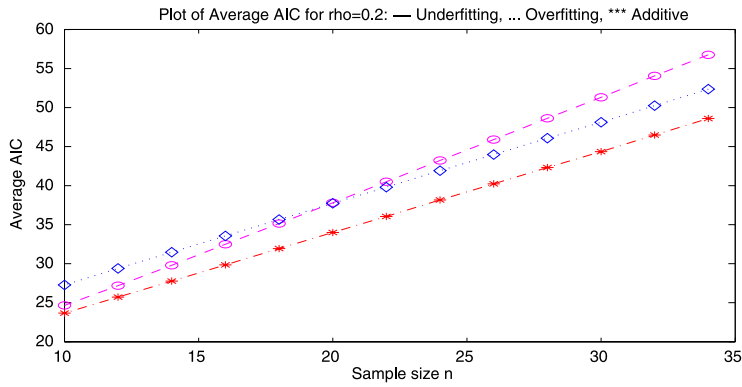
Figure 1. AIC(ψ_u), AIC(ψ_o), AIC(ψ_a) and sample size n for $\rho = 0.2$.

where p_u, p_o are the numbers of parameters in Θ_u, Θ_o, respectively, and p_a is sum of the numbers of parameters in Θ_1 and Θ_2.

In our simulation, consider $n_1 = n_2 = n/2$, replication times $N = 10\,000$ and $\rho = 0.2, 0.5$ and 0.8, respectively.

With N replication times, the average values of AIC_u, AIC_o and AIC_a are denoted by

$$\text{AIC}(\psi_u) = \frac{1}{N} \sum_{i=1}^{N} \text{AIC}_u^i, \qquad \text{AIC}(\psi_o) = \frac{1}{N} \sum_{i=1}^{N} \text{AIC}_o^i, \qquad \text{AIC}(\psi_a) = \frac{1}{N} \sum_{i=1}^{N} \text{AIC}_a^i.$$

The relations between the sample size n and AIC(ψ_u), AIC(ψ_o) and AIC(ψ_a) are illustrated in Figures 1–3 for $\rho = 0.2, 0.5$ and 0.8, respectively. We can make the following conclusions from these curves:
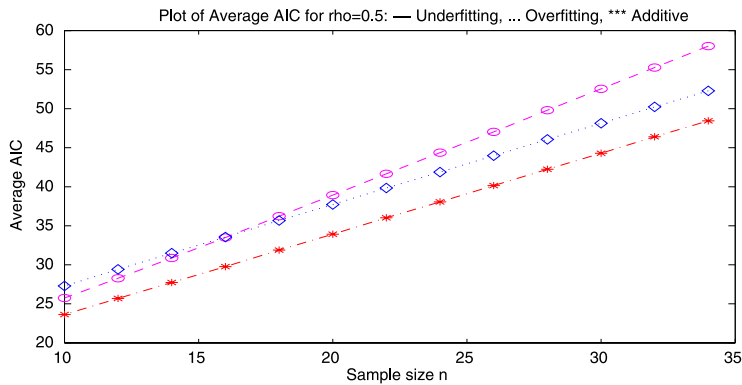

Figure 2. AIC(ψ_u), AIC(ψ_o), AIC(ψ_a) and sample size n for $\rho = 0.5$.
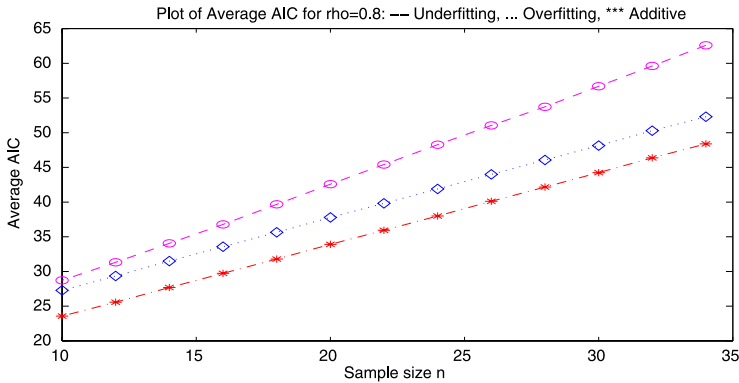
Figure 3. AIC(ψ_u), AIC(ψ_o), AIC(ψ_a) and sample size n for $\rho = 0.8$.

(1) Akaike's information criterion of the true model ψ_a remains to be uniformly smallest for all cases of $\rho = 0.2, 0.5$ and 0.8. The trend becomes particularly obvious as the sample size n increases. We believe that the conclusion is true for all $\rho \in (0, 1)$.

(2) The curve for AIC of the true model ψ_a and the curve for AIC of the overfitted model ψ_o are parallel. It implies that the difference between AIC for the true model ψ_a and AIC of the overfitted model ψ_o is a constant. The constant is due to the penalty for overparametrization. This shows that it is not significant for the difference between the RMSS for the true model ψ_a and the RMSS for the overfitted model. Overfitting gets a penalty for overparametrization and leads to a bigger AIC.

(3) The underfitted model ψ_u seems to have a bigger AIC than the overfitted model. It means that underfitting incurs more loss than overfitting does in the terms of AIC. The loss becomes larger and larger as the sample size increases or ρ is closer and closer to 1.

(4) The curve for AIC of the underfitted model becomes a little bit steeper as ρ is gradually close to 0, while the curve for AIC of the overfitted model and the curve for AIC of the true model seem to be unrelated to ρ.

In conclusion, using the additive growth curve model (1) with orthogonal design matrices has an obvious advantage over using the traditional growth curve model in model specification and then in parameter estimation.

6. A numerical example

The numerical example, stated in [14], about a certain measurement in a dental study on 11 girls and 16 boys at 4 different ages – 8, 10, 12 and 14 – is employed here (see Table 1) to illustrate the ideas and techniques stated in the paper.

Prior to making the model specification, we do not know whether the distances, in millimeters, from the center of the pituitary to the pteryo-maxillary fissure of these girls and boys follow two polynomial functions of time t with a same degree. So we assume that the distances for girls

Table 1. Measurements on 11 girls and 16 boys, at 4 different ages –8, 10, 12, 14

Girls	8	10	12	14	Boys	8	10	12	14
1	21	20	21.5	23	1	26	25	29	31
2	21	21.5	24	25.5	2	21.5	22.5	23	26.5
3	20.5	24	24.5	26	3	23	22.5	24	27.5
4	23.5	24.5	25	26.5	4	25.5	27.5	26.5	27
5	21.5	23	22.5	23.5	5	20	23.5	22.5	26
6	20	21	21	22.5	6	24.5	25.5	27	28.5
7	21.5	22.5	23	25	7	22	22	24.5	26.5
8	23	23	23.5	24	8	24	21.5	24.5	25.5
9	20	21	22	21.5	9	23	20.5	31	26
10	16.5	19	19	19.5	10	27.5	28	31	31.5
11	24.5	25	28	28	11	23	23	23.5	25
					12	21.5	23.5	24	28
					13	17	24.5	26	29.5
					14	22.5	25.5	25.5	26
					15	23	24.5	26	30
					16	22	21.5	23.5	25
Mean	21.18	22.23	23.09	24.09	Mean	22.87	23.81	25.72	27.47

and for boys follow two polynomial functions of time t with different degrees g and b (set $1 \leq g, b \leq 3$).

Based on the model (1), we think of these observations as realizations of the following model:

$$Y = X_1 \Theta_1 Z'_g + X_2 \Theta_2 Z'_b + \mathcal{E},$$

where

$$X_1 = \begin{pmatrix} \mathbf{1}_{11} \\ \mathbf{0} \end{pmatrix}, \qquad \Theta_1 = (\theta_{11} \quad \cdots \quad \theta_{1g}), \qquad Z'_g = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ t_1^g & t_2^g & t_3^g & t_4^g \end{pmatrix} \quad \text{for } 1 \leq g \leq 3,$$

and

$$X_2 = \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_{16} \end{pmatrix}, \qquad \Theta_2 = (\theta_{21} \quad \cdots \quad \theta_{2b}), \qquad Z'_b = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ t_1^b & t_2^b & t_3^b & t_4^b \end{pmatrix} \quad \text{for } 1 \leq b \leq 3.$$

We should trade the effect of the RMSS from the simple "true" model and the loss from overparameterization. Due to setting $1 \leq g, b \leq 3$, we can structure nine models for selection. The corresponding AICs of the nine models are displayed in Table 2.

The best model is the one with the minimum AIC. Based on AIC, the model with parameter pairs $(1, 1)$ is best, that is, the growth curves for girls and boys are two linear equations of time t. Our conclusion of model specification is consistent with the chosen model of [14].

Table 2. Parameter pair (g, b) and AICs for 9 models

(g, b)	AIC	(g, d)	AIC	(g, b)	AIC
(1, 1)	90.4011*	(1, 2)	92.2497	(1, 3)	94.1817
(2, 1)	92.4009	(2, 2)	94.2495	(2, 3)	96.1815
(3, 1)	94.3972	(3, 2)	96.2458	(3, 3)	98.1777

7. Concluding remarks

When observations of a repeated measurement at multiple time points follow polynomial functions with different degrees rather than the same degree, using the traditional growth curve model may cause underfitting or overfitting. To avoid these troubles, we proposed an additive growth curve model (1) with orthogonal design matrices that allows us to fit the data and then estimate parameters in a more parsimonious and efficient way than the traditional growth curve model. Obviously, the proposed additive growth curve model can not be included in the extended growth curve models investigated by Kollo and von Rosen [9], Chapter 4, and Verbyla and Venables [17].

In the paper, we explored the least-squares approach to derive two-stage GLS estimators for the regression coefficients, where an invariant and unbiased quadratic estimator for the covariance of observations is taken as the first-stage estimator. We investigated the properties of these estimators, including unbiasedness, consistency and asymptotic normality.

Simulation studies and a numerical example are given to illustrate the efficiency and parsimony of the proposed model for model specification in the sense of minimizing AIC compared to the traditional growth curve model. It follows that our additive growth curve model and the least-squares estimation for regression coefficients are competitive alternatives to the traditional growth curve model.

Appendix

Proof of Theorem 2.1. Put $T = (X_1 \otimes Z_1, \ldots, X_k \otimes Z_k)$ and $\boldsymbol{\beta} = ((\mathrm{vec}(\Theta_1))', \ldots, (\mathrm{vec}(\Theta_k))')'$. Then the model (1) can be rewritten as $\mathrm{vec}(\boldsymbol{\mu}) = T\boldsymbol{\beta}$ where $\mathscr{C}(T) = \mathscr{C}(X_1 \otimes Z_1) + \cdots + \mathscr{C}(X_k \otimes Z_k)$.

(a) To obtain the two-stage generalized least square estimate of $\boldsymbol{\mu}$ from the data Y is equivalent to applying the ordinary least square method to the following model

$$\mathrm{vec}(Z) = \mathrm{vec}(\boldsymbol{v}) + \left(I \otimes \Sigma^{-1/2}(Y)\right) \mathrm{vec}(\mathcal{E}), \tag{15}$$

where $\mathrm{vec}(Z) = (I \otimes \Sigma^{-1/2}(Y)) \mathrm{vec}(Y)$ and $\mathrm{vec}(\boldsymbol{v}) = (I \otimes \Sigma^{-1/2}(Y))T\boldsymbol{\beta}$. So the ordinary least square estimator of $\mathrm{vec}(\boldsymbol{v})$ is

$$\mathrm{vec}(\hat{\boldsymbol{v}}_{\mathrm{ols}}(Z)) = P_{(I \otimes \Sigma^{-1/2}(Y))T} \, \mathrm{vec}(Z). \tag{16}$$

Thus by equations (15) and (16)

$$\text{vec}(\hat{\boldsymbol{\mu}}(Y)) = \left(I \otimes \Sigma^{1/2}(Y)\right) P_{(I \otimes \Sigma^{-1/2}(Y))T} \left(I \otimes \Sigma^{-1/2}(Y)\right) \text{vec}(Y). \tag{17}$$

Since

$$P_{(I \otimes \Sigma^{-1/2}(Y))T} = \left(I \otimes \Sigma^{-1/2}(Y)\right) T \left(T'\left(I \otimes \Sigma(Y)\right)^{-1} T\right)^{+} T' \left(I \otimes \Sigma^{-1/2}(Y)\right), \tag{18}$$

by equations (17) and (18), we obtain

$$\text{vec}(\hat{\boldsymbol{\mu}}(Y)) = T \left(T'\left(I \otimes \Sigma(Y)\right)^{-1} T\right)^{+} T' \left(I \otimes \Sigma(Y)\right)^{-1} \text{vec}(Y).$$

By Kronecker product operations and (7), $\text{vec}(\hat{\boldsymbol{\mu}}(Y))$ reduces to

$$\text{vec}(\hat{\boldsymbol{\mu}}(Y)) = \sum_{i=1}^{k} \{X_i (X_i' X_i)^{-} X_i' \otimes Z_i (Z_i' \Sigma^{-1}(Y) Z_i)^{+} Z_i' \Sigma^{-1}(Y)\} \text{vec}(Y). \tag{19}$$

Since $(Z_i (Z_i' Z_i)^{-} Z_i' \Sigma^{-1}(Y) Z_i (Z_i' Z_i)^{-} Z_i')^{+} = Z_i (Z_i' \Sigma^{-1}(Y) Z_i)^{+} Z_i'$, in matrix language, we obtain the expression (7) of $\hat{\boldsymbol{\mu}}(Y)$ by rewriting (19).

(b) It follows from equation (3) and the condition (2).

(c) To prove the unbiasedness of Θ_i's, by (4) and (7), it suffices to show that $\hat{\boldsymbol{\mu}}(Y)$ is an unbiased estimator of $\boldsymbol{\mu}$.

Since $\widehat{\Sigma}(Y) = \widehat{\Sigma}(\mathcal{E}) = \widehat{\Sigma}(-\mathcal{E})$, $H_i(-\mathcal{E}) = H_i(\mathcal{E})$ and $\text{E}(\mathcal{E} H_i(\mathcal{E})) = \mathbf{0}$. By (5), $\hat{\boldsymbol{\mu}}(Y)$ can be expressed as

$$\hat{\boldsymbol{\mu}}(Y) = \sum_{i=1}^{k} P_{X_i} Y H_i(Y) = \sum_{i=1}^{k} X_i \Theta_i Z_i' + \sum_{i=1}^{k} P_{X_i} \mathcal{E} H_i(\mathcal{E}).$$

And

$$\text{E}(\hat{\boldsymbol{\mu}}(Y)) = \sum_{i=1}^{k} X_i \Theta_i Z_i' + \sum_{i=1}^{k} P_{X_i} \text{E}(\mathcal{E} H_i(\mathcal{E})) = \sum_{i=1}^{k} X_i \Theta_i Z_i' = \boldsymbol{\mu},$$

completing the proof.   □

Proof of Lemma 4.1. Some subscript i's are ignored in the following statements. Write $V = \frac{1}{\sqrt{n}} X_i' = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$. The transpose of $\mathbf{v}_j$ is an m_i-element row vector as follows,

$$\mathbf{v}_j' = \left(\frac{1}{\sqrt{n}} a_{j1}, \ldots, \frac{1}{\sqrt{n}} a_{jm}\right),$$

where $X_i = [a_{lj}]_{n \times m_i}$. By (12), $VV' = n^{-1} X_i' X_i$ converges to a positive definite matrix R_i. So the elements of $VV' = \mathbf{v}_1 \mathbf{v}_1' + \cdots + \mathbf{v}_n \mathbf{v}_n'$ are bounded. We claim that, for any $j \in \{1, \ldots, n\}$, the m_i elements of $\mathbf{v}_j$ are all $\text{O}(n^{-1/2})$.

If this is not true, we can assume without loss of generality that one element of $\mathbf{v}_n$ is $\text{O}(n^{p-1/2})$ with $p > 0$. Then one element of $\mathbf{v}_n \mathbf{v}_n'$ would be $\text{O}(n^{2p-1})$. Hence, the corresponding element in

matrix $VV' = \mathbf{v}_1\mathbf{v}_1' + \cdots + \mathbf{v}_n\mathbf{v}_n'$ would be $O(n^{2p})$, which is not bounded. This is a contradiction to condition (12).

Since

$$\left(\sqrt{n}\mathbf{s}_{i1}, \ldots, \sqrt{n}\mathbf{s}_{in}\right) = \sqrt{n}(X_i'X_i)^{-1}X_i' = n(X_i'X_i)^{-1}\frac{1}{\sqrt{n}}X_i'$$

$$= n(X_i'X_i)^{-1}[\mathbf{v}_1, \ldots, \mathbf{v}_n],$$

namely, for $j = 1, \ldots, n$, $\sqrt{n}\mathbf{s}_{ij} = n(X_i'X_i)^{-1}\mathbf{v}_j$. Thus, for $j = 1, \ldots, n$, the m_i elements of $\sqrt{n}\mathbf{s}_{ij}$ are also $O(n^{-1/2})$, completing the proof. $\qquad\square$

Proof of Theorem 4.2. Fix i. Let $\Gamma_i = S_i\mathcal{E} \in \mathcal{M}_{m_i \times p}$. Then Γ_i can be rewritten as

$$\Gamma_i = \sum_{j=1}^{n} \mathbf{s}_{ij}\mathcal{E}_j',$$

where $\mathbf{s}_{ij}$ is the jth column vector of $\mathbf{S}_i$ and $\mathcal{E}_j'$ is the jth row vector of the matrix $\mathcal{E}$ with $\mathcal{E} \sim \mathcal{G}(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$.

Since $\{\mathcal{E}_j'\}_{j=1}^{n}$ are independent and identically distributed, for $\mathbf{t} \in \mathcal{M}_{m_i \times p}$, the characteristic function $\Psi_n(\mathbf{t})$ of $\sqrt{n}\Gamma_i$ is given by

$$\Psi_n(\mathbf{t}) = E\left(\exp\{i\operatorname{tr}(\sqrt{n}\mathbf{t}'\Gamma_i)\}\right) = E\left(\exp\left\{i\operatorname{tr}\left(\sqrt{n}\mathbf{t}'\sum_{j=1}^{n}\mathbf{s}_{ij}\mathcal{E}_j'\right)\right\}\right)$$

$$= E\left(\exp\left\{i\operatorname{tr}\left(\sqrt{n}\sum_{j=1}^{n}\mathbf{t}'\mathbf{s}_{ij}\mathcal{E}_j'\right)\right\}\right) = \prod_{j=1}^{n}\Phi(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}),$$

where $\Phi(\cdot)$ is the characteristic function of $\mathcal{E}_j'$.

Recall that for u in the neighborhood of 0,

$$\ln(1 - u) = -u + f(u) \qquad \text{with } f(u) = \frac{1}{2}u^2 + o(u^2). \tag{20}$$

Write $p(u) = f(u)/u$, then from (20),

$$p(u) = o(u) \qquad \text{as } u \to 0. \tag{21}$$

And

$$\Phi(\mathbf{x}) = 1 - \frac{1}{2}\mathbf{x}'\Sigma\mathbf{x} + g(\mathbf{x}) \qquad \text{for } \mathbf{x} \in \mathbb{R}^{m_i} \quad \text{and}$$

$$g(\mathbf{x}) = o(\|\mathbf{x}\|^2) \qquad \text{as } \mathbf{x} \to \mathbf{0}. \tag{22}$$

For $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that

$$|g(\mathbf{x})| < \varepsilon \|\mathbf{x}\|^2 \qquad \text{as } 0 < \|\mathbf{x}\| < \delta(\varepsilon). \tag{23}$$

By (20) and (22),

$$\begin{aligned}
\ln\big(\Phi\big(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\big)\big) &= \ln\left(1 - \frac{n}{2}\mathbf{s}'_{ij}\mathbf{t}\Sigma\mathbf{t}'\mathbf{s}_{ij} + g\big(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\big)\right) \\
&= -\frac{1}{2}n\mathbf{s}'_{ij}\mathbf{t}\Sigma\mathbf{t}'\mathbf{s}_{ij} + g\big(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\big) + f\left(\frac{1}{2}n\mathbf{s}'_{ij}\mathbf{t}\Sigma\mathbf{t}'\mathbf{s}_{ij} - g\big(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\big)\right).
\end{aligned}$$

Therefore, the characteristic function of $\sqrt{n}\Gamma_n$ can be decomposed as

$$\Psi_n(\mathbf{t}) = \exp\left\{\sum_{j=1}^{n} \ln\big(\Phi\big(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\big)\big)\right\} \equiv \exp\left\{-\frac{1}{2}\alpha_n + \boldsymbol{\beta}_n + \eta_n\right\}, \tag{24}$$

where

$$\alpha_n = \sum_{j=1}^{n} n\mathbf{s}'_{ij}\mathbf{t}\Sigma\mathbf{t}'\mathbf{s}_{ij} = \text{tr}\left(\sum_{j=1}^{n} n\mathbf{s}'_{ij}\mathbf{t}\Sigma\mathbf{t}'\mathbf{s}_{ij}\right),$$

$$\boldsymbol{\beta}_n = \sum_{j=1}^{n} g\big(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\big)$$

and

$$\eta_n = \sum_{j=1}^{n} f\left(\frac{1}{2}n\mathbf{s}'_{ij}\mathbf{t}\Sigma\mathbf{t}'\mathbf{s}_{ij} - g\big(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\big)\right).$$

Note that

$$\sum_{j=1}^{n} \mathbf{s}_{ij}\mathbf{s}'_{ij} = (X'_i X_i)^{-1}. \tag{25}$$

For α_n, by (25), we have

$$\alpha_n = \text{tr}\left(n\mathbf{t}\Sigma\mathbf{t}'\sum_{j=1}^{n}\mathbf{s}_{ij}\mathbf{s}'_{ij}\right) = \text{tr}(n\Sigma\mathbf{t}'(X'_i X_i)^{-1}\mathbf{t}). \tag{26}$$

By (12),

$$\lim_{n\to\infty} \alpha_n = \text{vec}(\mathbf{t})'(R_i^{-1} \otimes \Sigma)\,\text{vec}(\mathbf{t}). \tag{27}$$

For $\boldsymbol{\beta}_n$, by Lemma 4.1 and the continuity of $\mathbf{t}'\mathbf{s}_{ij}$, for the $\delta(\varepsilon) > 0$ in (23), there is an integer $N(\varepsilon) > 0$ such that for $n > N(\varepsilon)$,

$$0 < \left\| \sqrt{n}\mathbf{t}'\mathbf{s}_{ij} \right\| < \delta(\varepsilon) \qquad \text{for all } j = 1, \ldots, n. \tag{28}$$

Take $n > N(\varepsilon)$, then by (23) and (28),

$$\left| g\left(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\right) \right| < \left\| \sqrt{n}\mathbf{t}'\mathbf{s}_{ij} \right\|^2 \varepsilon. \tag{29}$$

By (25),

$$|\boldsymbol{\beta}_n| < \sum_{j=1}^n \left\| \sqrt{n}\mathbf{t}'\mathbf{s}_{ij} \right\|^2 \varepsilon = \varepsilon n \sum_{j=1}^n \mathrm{tr}(\mathbf{t}'\mathbf{s}_{ij}\mathbf{s}'_{ij}\mathbf{t}) = \varepsilon \, \mathrm{tr}(\mathbf{t}'n(X_i X_i)^{-1}\mathbf{t}). \tag{30}$$

So by (12), $\limsup_{n\to\infty} |\boldsymbol{\beta}_n| \le \varepsilon \, \mathrm{tr}(\mathbf{t}' R_i^{-1}\mathbf{t})$. Since $\varepsilon > 0$ is arbitrary, we obtain

$$\lim_{n\to\infty} \boldsymbol{\beta}_n = 0. \tag{31}$$

And for η_n, let

$$\lambda_j = \frac{1}{2}\left(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\right)' \Sigma \left(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\right) - g\left(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\right).$$

Thus, by (29),

$$|\lambda_j| < \frac{1}{2}\left(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\right)' \Sigma \left(\sqrt{n}\mathbf{t}'\mathbf{s}_{ij}\right) + \left\| \sqrt{n}\mathbf{t}'\mathbf{s}_{ij} \right\|^2 \varepsilon. \tag{32}$$

Take $n > N(\varepsilon)$, by Lemma 4.1, the continuity of $\mathbf{t}'\mathbf{s}_{ij}$ and (21), increasing $N(\varepsilon)$ if necessary, we may suppose that for all j, $|p(\lambda_j)| < \varepsilon$. Since $f(\lambda_j) = p(\lambda_j)\lambda_j$,

$$|\eta_n| \le \sum_{j=1}^n |f(\lambda_j)| = \sum_{j=1}^n |p(\lambda_j)||\lambda_j| \le \sum_{j=1}^n \varepsilon|\lambda_j|.$$

By (32),

$$|\eta_n| \le \sum_{j=1}^n \left[\frac{\varepsilon}{2} \mathrm{tr}\left(\sqrt{n}\mathbf{s}'_{ij}\mathbf{t}\Sigma\mathbf{t}'\sqrt{n}\mathbf{s}_{ij}\right) + \left\| \sqrt{n}\mathbf{t}\mathbf{s}_{ij} \right\|^2 \varepsilon^2 \right].$$

Then, taking the same operations as (26) and (30), we obtain the following inequality

$$|\eta_n| \le \left[\frac{\varepsilon}{2} \mathrm{tr}(n(X_i' X_i)^{-1}\mathbf{t}\Sigma\mathbf{t}') + \varepsilon^2 \, \mathrm{tr}(\mathbf{t}'n(X_i' X_i)^{-1}\mathbf{t}) \right],$$

namely, by (12)

$$|\eta_n| \le \frac{\varepsilon}{2} \mathrm{tr}\left(\sum_{l=1}^i R_i^{-1}\mathbf{t}\Sigma\mathbf{t}'\right) + \varepsilon^2 \, \mathrm{tr}(\mathbf{t}' R_i^{-1}\mathbf{t}). \tag{33}$$

Due to arbitrary of ε and (33),

$$\lim_{n\to\infty} \eta_n = 0. \tag{34}$$

By (27), (31) and (34), we obtain from (24),

$$\lim_{n\to\infty} \Psi_n(\mathbf{t}) = \exp\left\{-\frac{1}{2}\operatorname{vec}(\mathbf{t})'(R_i^{-1}\otimes\Sigma)\operatorname{vec}(\mathbf{t})\right\}. \tag{35}$$

So by Levy's continuity theorem, $\sqrt{n}\,\Gamma_i$ converges in distribution to $\mathcal{N}_{m_i\times p}(\mathbf{0}, R_i^{-1}\otimes\Sigma)$, completing the proof of the desired result. $\qquad\square$

Acknowledgements

References

[1] Akaike, H. (1973). Information theory and in extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (B.N. Petrov and B.F. Csaki, eds.) 267–281. Academiai Kiado: Budapest. MR0483125

[2] Chaganty, N.R. (2003). Analysis of growth curves with patterned correlation matrices using quasi-least squares. *J. Statist. Plann. Inference* **117** 123–139. MR2001145

[3] Grizzle, J.E. and Allen, D.M. (1969). Analysis of growth and dose response curves. *Biometrics* **25** 357–381.

[4] Hu, J. (2008). Wishartness and independence of matrix quadratic forms in a normal random matrix. *J. Multivariate Anal.* **99** 555–571. MR2396979

[5] Hu, J. (2010). Properties of the explicit estimators in the extended growth curve model. *Statistics* **44** 477–492.

[6] Hu, J. and Yan, G. (2008). Asymptotic normality and consistency of a two-stage generalized least squares estimator in the growth curve model. *Bernoulli* **14** 623–636. MR2537805

[7] Khatri, C.G. (1966). A note on a MANOV model applied to problems in growth curves. *Ann. Inst. Statist. Math.* **18** 75–86. MR0219181

[8] Kollo, T., Roos, A. and von Rosen, D. (2007). Approximation of the distribution of the location parameter in the growth curve model. *Scand. J. Statist.* **34** 499–510. MR2368795

[9] Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statsitics with Matrics. Mathematicals and its Applications (New York)* **579**. Dordrecht: Springer.

[10] Lang, N. and Laird, N.M. (1988). The effect of covariance structure on variance estimation in balanced growth curve models with random parameters. *J. Amer. Statist. Assoc.* **84** 241–247. MR0999684

[11] Lee, J.C. (1988). Prediction and estimation of growth curves with special covariance structures. *J. Amer. Statist. Assoc.* **83** 432–440. MR0971369

[12] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*. New York: Springer. MR2135927

[13] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley. MR0652932

[14] 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. MR0181062

[15] Rao, C.R. (1965). The theory of last squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52** 447–458. MR0205392

[16] Rao, C.R. (1987). Prediction of future observations in growth curve models (with comments). *Statist. Sci.* **4** 434–471. MR0933738

[17] Verbyla, A.P. and Venables, W.N. (1988). An extension of the growth curve model. *Biometrika* **75** 129–138. MR0932826

[18] Žežula, I. (1993). Covariance components estimation in the growth curve model. *Statistics* **24** 320–330. MR1241625