Efficient Estimation of General Linear Growth Curve Models

Eugene Demidenko and Thérèse A. Stukel
Section of Epidemiology and Biostatistics
Dartmouth Medical School

November 19, 1998

Submitted to

Journal of Statistical Planning and Inference

ABSTRACT

A general linear growth curve model is proposed where only part, or more generally, a linear combination of growth curve parameters, growth curve chracteristic is modelled in the second stage model. The formula for generalized least squares estimator under constraints is derived when variance parameters are known. When the variance parameters are unknown, three methods of estimation for general linear growth curve model are proposed: full maximum likelihood, second stage maximum likelihood and the two-stage estimator. A consistent and unbiased estimator for the covariance matrix of the random effects is derived based on the method of moments. The two-stage estimator has the form of a generalized least squares estimator under constraints with variance parameters estimated by the method of moments. It is shown that all three estimators are consistent, asymptotically normally distributed, efficient and equivalent when the number of individuals goes to infinity. The properties of estimators are investigated under model misspecification. It is shown that one might come to a severe bias using standard growth curve model. The suggested approach is illustrated on pancreatic tumor growth data example.

KEY WORDS: Linear model; Mixed effects, Random effects; Repeated measures, Longitudinal analysis, Maximum likelihood; Moment estimator.

1. Introduction

Linear mixed-effects models have received considerable attention both from the theoretical and applied points of view. Recently released books on linear and nonlinear mixed effects models provide a comprehensive description and introduction to the topic, .Davidian , Vonesh. This paper concerns an important type of linear mixed-effects model, the linear growth curve model (Potthof and Roy (1964), Ware (1985)). The present paper further develops a generalized class of linear growth curve models introduced in Stukel and Demidenko (1987), where only a linear combination of parameters is specified.

The linear growth curve model is expressed in two stages. The first stage is written as a linear regression with random coefficients,

$$y_i = X_i a_i + \epsilon_i; \qquad i = 1; \ldots; N \tag{1.1}$$

where y_i is the $n_i \times 1$ vector of serial measurements on individual i; X_i is a fixed $n_i \times m$ within-individual design matrix; in the present paper it is assumed $\operatorname{rank}(X_i) = m \leq n$: a_i is an $m \times 1$ vector of random coefficients, and ϵ_i is independent error term with $E(\epsilon_i) = 0$ and $\operatorname{cov}(\epsilon_i) = \sigma^2 I$: The second stage model describes the random coefficients a_i through population covariates,

$$a_i = Z_i \beta + b_i \tag{1.2}$$

where Z_i is a known $m \times k$ population design matrix; β is a $k \times 1$ vector of population parameters of interest, and b_i is the vector of random effects with $E(b_i) = 0$ and $\operatorname{cov}(b_i) = \Omega$; independent of ε_i. Vector comprises σ^2 and the distinct elements of the matrix Ω is denoted θ and called variance parameter. It is assumed that observations on different individuals are independent. Assuming normal errors, estimation for model (1.1, 1.2) has been based on unrestricted (ML) or restricted (RML) maximum likelihood methods (Laird and Ware (1982), Jennrich and Schluchter (1986), Laird, Lange and Ware (1987), Lindstrom and Bates (1988)). Vonesh and Carter (1987) suggested a noniterative three-step procedure for models where $Z_i = (I \otimes z_i)$; i.e. z_i is the common vector of covariates for elements of vector a_i.

We underscore that in the model (1.2) we have to specify the relationship for all elements of vector a_i: However, often investigator is interested only in a subset of the elements of a_i; and consequently, only for these elements he can provide a justified relationship in the form of the second stage model. The goal of the present paper is to suggest new estimators for the general class of growth curve model introduced in Stukel and Demidenko (1987) and investigate their asymptotic properties.

The structure of the paper is as follows. In the next section generalized least squares estimator under constraints for general growth curve model is derived assuming variance parameters are known. In section 3 likelihood based methods are described when variance paramater is unknown. variance parameters are estimated by the method of moments. In section 4 moment estimator for the variance parameter is derived. In section 5 asymptotic properties of considered estimators are formulated. In section 6 we investigate properties of estimators under model misspecification. Finally, in section 7 the general growth curve model is illustrated on a pancreatic tumor growth example.

2. General linear growth curve model

The first stage of the general linear growth curve model (1.1) remains the same, but the second stage specifies only some characteristics of the curve parameters, namely

$$Ca_i = Z_i \beta + \delta_i \tag{2.1}$$

where C is a known $r \times m$ matrix of full rank $r \leq m$ and $\mathrm{cov}(\delta_i) = C - C'$. Z_i is a known $r \times k$ design matrix such that $\sum Z_i' Z_i$ is of full rank. Also, in this paper it is assumed that $m \leq n_i$ and matrix X_i has full rank. When $r = m$ model (2.1) is equivalent to the usual growth curve model (1.2).

For estimation purposes, it is convenient to rewrite model (1.1, 2.1) in its marginal form:

$$y_i = X_i \alpha_i + \zeta_i, \qquad E(\zeta_i) = 0, \quad \mathrm{cov}(\zeta_i) = V_i = \sigma^2 I + X_i - X_i' \tag{2.2}$$

with parameters subject to linear constraints

$$C\alpha_i = Z_i \beta \tag{2.3}$$

where $\alpha_i = E(a_i)$, $i = 1, \ldots, N$. Thus marginal means α_i are not specified completely unless $r = m$. In particular, the number of parameters increases with the number of individuals.

2.1. Variance parameters are known

We begin with the assumption that the variance parameters σ^2 and Ω are known. Based on model (2.2), the generalized least squares estimator of β would minimize the following sum of wighted squares:

$$\min_{\alpha_i} \sum_i (y_i - X_i\alpha_i)' V_i^{-1} (y_i - X_i\alpha_i) \tag{2.4}$$

under constraints (2.3). To find the solution to this quadratic form, we introduce the Lagrangian function:

$$L = \sum_i (y_i - X_i\alpha_i)' V_i^{-1} (y_i - X_i\alpha_i) + 2\sum \lambda_i'(C_i\alpha_i - Z_i\beta)$$

where λ_i is the $r \times 1$ Lagrangian vector of coefficients. Differentiating L with respect to α_i and β we obtain the necessary conditions for the minimum,

$$-\frac{1}{2}\partial L/\partial\alpha_i = X_i' V_i^{-1}(y_i - X_i\alpha_i) + C'\lambda_i = 0; \qquad i = 1, \ldots, N \tag{2.5}$$

and

$$-\frac{1}{2}\partial L/\partial\beta = \sum Z_i'\lambda_i = 0. \tag{2.6}$$

Define

$$R_i = X_i' V_i^{-1} X_i; \qquad M_i = C R_i^{-1} C'; \qquad a_i^o = (X_i' V_i^{-1} X_i)^{-1} X_i' V_i^{-1} y_i; \tag{2.7}$$

and notice that matrices R_i and M_i are positive definite. Using the matrix identity $(P'(I + PQP')^{-1}P)^{-1} = (P'P)^{-1} + Q$; we obtain

$$R_i^{-1} = \sigma^2 (X_i' X_i)^{-1} + \Omega; \quad M_i = \sigma^2 C (X_i' X_i)^{-1} C' + B. \tag{2.8}$$

Solving (2.5) and (2.6) in terms of β we find the solution to (2.4) as

$$\hat{\beta} = \left(\sum Z_i' M_i^{-1} Z_i\right)^{-1} \left(\sum Z_i M_i^{-1} C R_i^{-1} X_i V_i^{-1} y_i\right)$$

which is called the Generalized Least Squares estimator under Constraints (GLSC). Applying Theorem 2.4.1 of Anderson (1971) we can simplify the above formula for the GLSC as

$$\hat{\beta} = \left(\sum Z_i' M_i^{-1} Z_i\right)^{-1} \left(\sum Z_i M_i^{-1} C a_i^o\right) \tag{2.9}$$

where $a_i^o = (X_i' X_i)^{-1} X_i' y_i$ is the ordinary least squares estimator of the ith regression (1.1). The covariance matrix $\mathrm{cov}(\hat{\beta}) = (\sum Z_i' M_i^{-1} Z_i)^{-1}$. As follows from the derived formula $\hat{\beta}$ can be calculated in two steps: (i) estimate each curve (1.1) individually by OLS, (ii) apply the weighted least squares to (2.1) substituting Ca_i by Ca_i^o with the weight matrix M_i defined in (2.8). This way of calculation provides the basis for the two-stage estimator (Stukel and Demidenko 1987). Applying standard theory for linear models with restrictions on parameters (Rao (1973)), it follows that when the variance parameters are known $\hat{\beta}$ is the best linear unbiased estimator (BLUE).

In practice, however, the variance parameters are unknown, and in order to use the GLSC we need to estimate σ^2 and D. From formula (2.8), the GLSC depends on $B = CDC'$, and consequently we need only estimators for σ^2 and B. Those estimators are considered in the following two sections.

3. Likelihood-based estimates

We can apply maximum likelihood (ML) theory to estimate the general linear growth curve model assuming normal distribution. Two types of likelihood estimates might be considered. The first is called as full ML and the second as second stage likelihood.

3.1. Full maximum likelihood estimation

If the variances parameters are known the MLE for β is defined by (2.9). Otherwise we have to estimate β and the variance parameters simultineously based on the log-likelihood function for (2.2):

$$l = -\frac{1}{2}\left\{\sum_{i=1}^{N}\left[\ln\left|\sigma^2 I + X_i D X_i'\right| + (y_i - X_i\alpha_i)'(\sigma^2 I + X_i D X_i')^{-1}(y_i - X_i\alpha_i)\right]\right\} \tag{3.1}$$

where β and $\alpha_1, \ldots, \alpha_N$ are subject to constraints (2.3). Our aim is to find the concentrated log-likelihood function exluding nuisance parameters that expresses l in terms of β and the variance parameters. Thus we minimize (3.1) under constraints (2.3) assiming β is fixed. After some algebra we come to the concentrated log-likelihood function

$$\begin{aligned} l_c(\beta; \sigma^2; D) \quad = \quad & -\frac{1}{2}\sum (n_i - m)\ln\sigma^2 \\ & -\frac{1}{2}\sum\left[\ln\left|\sigma^2(X_i'X_i)^{-1} + D\right| + (c_i - Z_i\beta)'(\sigma^2 P_i + CDC')^{-1}(c_i - Z_i\beta)\right] \end{aligned}$$

where $c_i = Ca_i^o$ and $P_i = C(X_i'X_i)^{-1}C'$. Maximization of l_c can be accomplished by any algorithm applied to the maximization of the log-likelihood function for standard linear mixed effect models, described by Laird and Ware (1982), Jennrich and Schluchter (1986), Lindstrom and Bates (1988), since the c_i are fixed.

3.2. Second stage maximum likelihood estimation

Taking $c_i = Ca_i^o$ as known we can estimate β using the second stage model (2.1) alone. It is easy to show that $\mathrm{cov}(c_i) = \sigma^2 P_i + B$. Thus, given $\{P_i\}$ the second stage model is equivalent to $c_i \sim N(Z_i\beta, \sigma^2 P_i + B)$. There exists an unbiased and consistent estimator of σ^2 as the pooled variance (Vonesh and Carter 1992),

$$\hat\sigma^2 = \frac{1}{\sum (n_i - m)} \sum (y_i - X_i a_i^o)'(y_i - X_i a_i^o). \tag{3.2}$$

Then, given c_i and $\hat\sigma^2$ the log-likelihood function for the second stage model is

$$l_2(\beta; B) = -\frac{1}{2}\left\{\sum_{i=1}^{N}\left[\ln\left|\hat\sigma^2 P_i + B\right| + (c_i - Z_i\beta)'(\hat\sigma^2 P_i + B)^{-1}(c_i - Z_i\beta)\right]\right\}. \tag{3.3}$$

A simple algorithm for maximizing (3.3) in the case of one characteristic based on the fix-point theorem was suggested by Pockock et al. (1981). Berke and Laird (1986) extended this algorithm to the multivariate case.

4. Moment estimators of the variance parameters and the two-stage estimator

A moment estimator for B was derived by Reinsel (1985) for the standard growth curve model (1.1, 1.2) in the case the Z_i are constant. We generalize that estimator to the general growth curve model with arbitrary $\{Z_i\}$. A natural estimate of B is based on the second stage model (2.1), once α_i are estimated as a_i^o. Then the OLS-estimate for β is $\hat\beta_o = (\sum Z_i'Z_i)^{-1}\sum Z_i'Ca_i^o$, and the empirical residual for the second stage model is $Ca_i^o - Z_i\hat\beta_o$. Following the idea of the method of moments, we construct the matrix of cross-products of residuals

$$\frac{1}{N}\sum (Ca_i^o - Z_i\hat\beta_o)(Ca_i^o - Z_i\hat\beta_o)'. \tag{4.1}$$

The moment estimator for B is found by equating (4.1) to its mathematical expectation (see the Appendix for details):

$$\mathrm{vech}(\hat B) = J^{-1}\mathrm{vech}\Big(\frac{1}{N}\sum_{i=1}^{N}\{(Ca_i^o - Z_i\hat\beta_o)(Ca_i^o - Z_i\hat\beta_o)' - \hat\sigma^2[P_i - P_iQ_i - Q_iP_i + Q_iP_iQ_i]\}\Big); \tag{4.2}$$

where

$$J = D^{+}\left[I_{r^2} - \frac{1}{N}\sum_{i=1}^{N}\left(Q_i \otimes I_r + I_r \otimes Q_i - Q_i \otimes Q_i'\right)\right] D \tag{4.3}$$

and $Q_i = Z_i(\sum_j Z_j'Z_j)^{-1}Z_i'$. Here D is the $r^2 \times r(r+1)/2$ duplication matrix and D^+ is its general inverse (Lancaster and Tismenetsky 1985, Fuller 1987, Magnus 1988). Let us recall that vech is the $r(r+1)/2$ vector function of a $r \times r$ symmetric matrix and vec is the $r^2 \times 1$ vector function. These functions are linearly related to each other as vec = Dvech and vech = D^+vech . As is shown in the Appendix, matrix J is nonsingular if $2r < N$ which is a very mild condition. Since $\hat{B}$ is a linear function of (4.1) and $\hat{\sigma}^2$, the moment estimator is unbiased. When N goes to infinity, terms in (4.2) containing Q_i vanish (see the Appendix), and we come to a simplified estimator

$$\hat{B}_s = \frac{1}{N}\sum_{i=1}^{N}\left[(Ca_i^o - Z_i\hat{\beta}_o)(Ca_i^o - Z_i\hat{\beta}_o)' - \hat{\sigma}^2 P_i\right]. \tag{4.4}$$

This estimator has a clear variance components interpretation from the formula $\mathrm{var}(Ca_i^o - Z_i\beta) = \mathrm{var}(Ca_i^o) + \mathrm{var}(\delta_i)$. The asymptotic properties and variances for these estimators are provided below.

4.1. Two-stage estimator

Since σ^2 and B are well estimated by the method of moments we can substitute them into formula for GLSEUC (2.9). Similar idea was implemented by Vonesh and Carter (1992) for standard growth curve model and a special structure for Z_i. This estimator is called "two-stage" (TS): at the first step σ^2 and B are estimated as (3.2) and (4.2) or (4.4), and at the second step β is estimated by generalized least squares.

5. Asymptotic properties of estimators

The asymptotic properties of the estimators are formulated in the following theorem. In order to prove asymptotic properties we have to make some assumptions when $N \to \infty$.

In particular, the following Asymptotic Boundedness assumptions (AB-assumptions) will be used: there exist positive constants D_0, d_X, D_4 and D_Z such that:

(a) $n_i \leq D_0$; (b) $d_X I_m \leq X_i'X_i$;

$$\text{(c) } E\varepsilon_{ij}^4 < D_4;\ E(a_{il} - E(a_{il}))^4 \le D_4; \qquad \text{(d) } |Z_{il}| \le D_Z$$

for $i = 1, 2, \ldots, N \to \infty;\ j = 1, \ldots, n_i;\ l = 1, \ldots, m$. First, we formulate the asymptotic properties of the estimates for the variances parameters.

Theorem 5.1. Under the AB-assumptions:

1. the estimators $\hat{B}$ and $\hat{B}_s$ are asymptotically equivalent,
2. the moment estimators are unbiased and consistent,
3. the asymptotic covariance matrix for vech$(\hat{B})$ and vech$(\hat{B}_s)$ is

$$\frac{2}{N} D^{+} \left(B \otimes B + \frac{\sigma^4}{N} \sum_{i=1}^{N} P_i \otimes P_i \right) D^{+\prime}. \tag{5.1}$$

Proof is in the Appendix.

In the case of one characteristic($r = 1$) the moment estimators (4.2) and (4.4) collapse to

$$\hat{B} = \frac{\frac{1}{N} \sum \left((Ca_i^0 - Z_i \hat{\beta}_o)^2 - \hat{\sigma}^2 p_i (1 - q_i)^2 \right)}{1 - \frac{1}{N}(2k - \sum q_i^2)}; \quad \hat{B}_s = \frac{1}{N} \sum \left((Ca_i^0 - Z_i \hat{\beta}_o)^2 - \hat{\sigma}^2 p_i \right)$$

where $q_i = Q_i$ and $p_i = P_i$ are scalars, with asymptotic variance

$$\mathrm{var}(\hat{B}_s) = \mathrm{var}(\hat{B}) = \frac{2}{N} \left(B^2 + \sigma^4 \frac{\sum p_i^2}{N} \right) \tag{5.2}$$

The expression (5.1) can be interpreted in terms of (4.4): it is the sum of the corresponding variances since, as is shown in the Appendix, these terms are asymptotically uncorrelated. In particular, formula (5.1) can be useful in testing the significance of the random effects by the Wald test. Note, that we did not used the assumption on normality of error terms.

In the next theorem, we show that all three estimators for β considered above are asymptotically equivalent.

Theorem 5.2. Under the AB-assumptions, the full and two-stage maximum likelihood estimators and the second-stage estimators are asymptotically normally distributed as $\sqrt{N}(\hat{\beta} - \beta) \simeq N(0; H)$; where

$$H = \lim_{N \to \infty} \frac{1}{N} \sum Z_i^{\prime} \left(\sigma^2 C (X_i^{\prime} X_i)^{-1} C^{\prime} + B \right)^{-1} Z_i \tag{5.3}$$

exists and is a positive definite matrix, i.e. are consistent, asymptotically efficient and equivalent. The normal distribution assumption is not neccassary for the two–stage estimator.

Proof is in the Appendix.

Thus, for the standard linear growth curve model (1.1, 1.2), i.e., when $C = I$ and errors are normally distributed, the TS-estimator is asymptotically equivalent to the usual MLE.

6. Properties of the estimators under model misspecification

In general linear growth curve model unlike standard setting parameters are not specified completely. A natural question is what we lose when we impose extra restrictions on parameters when they are in fact free. Vise versa, what happens when standard growth curve model takes place and we let parameters unspecified, assuming general setting.

This findings correspond to our previous simulation reported in Stukel and Demidenko (1997).

The aim of this section is to find out consequences of accepting the standard approach in modelling growth curves while only a part of parameters can be specified correctly, i.e. general growth curve model takes place and vice versa. In this section we restrict ourselves by modelling only the first element of $\mathbf{a}_i$

$$a_{i1} = \mathbf{z}_i \boldsymbol{\gamma} + b_{i1} \qquad (6.1)$$

where $\mathbf{z}_i$ is the $1 \times k$ vector of covariate, $\boldsymbol{\gamma}$ is the $k \times 1$ parameter of interest, and $E(b_{i1}) = 0$, $\mathrm{var}(b_{i1}) = \Omega_{11}$. The standard growth curve setting (1.2) forces us to make additional, irrelevant assumptions on the rest of the elements of $\mathbf{a}_i$. Apparently, there are three options:

1. Parameters of the growth curve not covered by the specification (6.1) are fixed and unknown, i.e.,

$$a_{ij} = \tau_j, \qquad \mathrm{var}(a_{ij}) = 0, \quad j = 2, \ldots, m. \qquad (6.2)$$

2. The remaining parameters are random with unknown means:

$$a_{ij} = \tau_j + b_{ij}, \qquad j = 2, \ldots, m. \qquad (6.3)$$

where $b_i = (b_{i1}, \ldots, b_{im})'$ and $\mathrm{cov}(b_i) = \Omega$.

3. Model (6.1) is applied to all elements of a_i:

$$a_i = (I \otimes z_i)\gamma + b_i, \tag{6.4}$$

where $\gamma = (\gamma'^{\,}, \tau')'$ is the $mk \times 1$ vector of parameters, $\tau = (\gamma_2', \ldots, \gamma_m')'$, and $\mathrm{cov}(b) = \Omega$.

Thus, one can apply standard growth curve model even when only part of parameter vector is of interest. Clearly, it leads to the model misspecification, and the aim of this section is to find out what is the price of the overspecification due to (6.2), (6.3) or (6.4). Vice versa, what happens if the true model is (1.1) and we model only the first component of a_i, i.e., what is the effect of underspecification?

We start with the overspecification. Thus, let us assume that the true model is (6.1), i.e., $z_i'\gamma = \alpha_{i1}$ but the rest of parameters $\alpha_{i2}, \ldots, \alpha_{im}$ are left unspecified and considered as nuisance parameters. Three models (6.2), (6.3), (6.4) are compared to the general linear growth curve model (2.1) with $C = (1, 0, \ldots, 0)$. We wish to determine bias induced by imposing an irrelevant structure on parameters $\alpha_{i2}, \ldots, \alpha_{im}$ when in fact there is none. In general, all three models lead to biased estimates in contrast to the GLSC which is unbiased.

To simplify, the variance parameters are assumed known in this section.

As follows from (2.9) the GLS-estimator for all three models can be written in the form

$$\begin{bmatrix} \hat{\beta} \\ \hat{\tau} \end{bmatrix} = \left(\sum Q_i' M_i^{-1} Q_i\right)^{-1} \left(\sum Q_i' M_i^{-1} a_i^o\right), \tag{6.5}$$

where $Q_i = \mathrm{diag}(z_i, T_i)$, and $T_i = I_{m-1}$ for models (6.1) and (6.2), and $T_i = (I_{m-1} \otimes z_i)$ for model (6.3), and τ is the vector of nuisance parameters. It is easy to see that GLSEUC is

$$\hat{\beta} = \left(\sum \frac{z_i'z_i}{\sigma^2 (X_i'X_i)^{-1}_{11} + \Omega_{11}}\right)^{-1} \left(\sum \frac{z_i' a_{i1}^o}{\sigma^2 (X_i'X_i)^{-1}_{11} + \Omega_{11}}\right), \tag{6.6}$$

and is unbiased.

In the orthogonal case, when $X_i'X_i$ and Ω are diagonal, all four estimators coincide. For balanced data when $X_i = X$ estimator (6.5) does not depend on the variance parameters coincides with the OLS-estimator (e.g., Laird, Lange and Ware

1987), so that (6.5) and (6.6) coincide. However, in the general case, all three models lead to biased estimates. To show this we partition α_i' as $(\alpha_{i1}, \alpha_{i2}')$ where $\alpha_{i2}' = (\alpha_{i2}, \ldots, \alpha_{im})$, and the other quantities are partitioned correspondingly. Then from (6.5) we find the expectation

$$E\begin{bmatrix}\hat{\gamma}\\ \hat{\boldsymbol{\alpha}}\end{bmatrix} = \begin{bmatrix}\sum M_{i11}^{-1} z_i' z_i & \sum z_i' M_{i12}^{-1} T_i \\ \sum T_i' M_{i12}^{-1\prime} z_i & \sum T_i' M_{i22}^{-1} T_i\end{bmatrix}^{-1} \begin{bmatrix}(\sum M_{i11}^{-1} z_i' z_i)\gamma + \sum z_i' M_{i12}^{-1}\alpha_{i2} \\ (\sum T_i' M_{i12}^{-1\prime} z_i)\gamma + \sum T_i' M_{i22}^{-1}\alpha_{i2}\end{bmatrix}$$

where, for instance, M_{i11}^{-1} is the $(1,1)$th element of matrix M_i^{-1}. Denote $U_{11} = \sum M_{i11}^{-1} z_i' z_i$, and $U_{12} = \sum z_i' M_{i12}^{-1} T_i$, and $U_{22} = \sum T_i' M_{i22}^{-1} T_i$; then using the formula for the partial inverse we come to the following formula for the bias

$$E(\hat{\gamma}) - \gamma = F \sum z_i' M_{i12}^{-1}\alpha_{i2} - U_{11}^{-1} U_{12} E^{-1} \sum T_i' M_{i22}^{-1}\alpha_{i2},$$

where $E = U_{22} - U_{12}' U_{11}^{-1} U_{12}$, $F = U_{11}^{-1} + U_{11}^{-1} U_{12} E^{-1} U_{12}' U_{11}^{-1}$. Further, it is easy to check that $F \sum z_i' M_{i12}^{-1} T_i \boldsymbol{\alpha} = A^{-1} B E^{-1} \sum T_i' M_{i22}^{-1} T_i \boldsymbol{\alpha}$ for any $\boldsymbol{\alpha}$. Thus, the above bias can be expressed in terms of the deviation of α_{i2} from its "modelled" value $T_i \boldsymbol{\alpha}$:

$$E(\hat{\gamma}) - \gamma = \sum_{i=1}^{N} (F z_i' M_{i12}^{-1} - U_{11}^{-1} U_{12} E^{-1} T_i' M_{i22}^{-1})(\alpha_{i2} - T_i \boldsymbol{\alpha}). \tag{6.7}$$

Thus, the estimate of γ based on the standard model (1.2) is unbiased if either $E(a_{i2}) = T_i \boldsymbol{\alpha}$ for some $\boldsymbol{\alpha}$, or if $F z_i' M_{i12}^{-1} = U_{11}^{-1} U_{12} E^{-1} T_i' M_{i22}^{-1}$ for $i = 1, \ldots, N$. Notice, that the latter case occurs when $z_i' M_{i12}^{-1}$ and $T_i' M_{i22}^{-1}$ are collinear. In particular, for the random coefficients model when $z_i = 1$ and $T_i = I$, the magnitude of the bias is related to how unbalance the data is. Therefore in general, specifying structure for the α_{i2} leads to a bias in the estimate.

We can even compute the maximum bias in terms of the deviation of α_{i2} from $T_i \boldsymbol{\alpha}$. Applying the matrix inequality $X'Y(Y'Y)^{-1}Y'X \le X'X$ (Graybill 1983) to (6.7) we obtain

$$\| E(\hat{\gamma}) - \gamma \|^2 \le \sum (\alpha_{i2} - T_i \boldsymbol{\alpha})'(\alpha_{i2} - T_i \boldsymbol{\alpha}) \sum \mathrm{tr}(R_i' R_i)$$

where $C_i = F z_i' M_{i12}^{-1} - A^{-1} B E^{-1} T_i' M_{i22}^{-1}$.

In contrast, if the true model has a fixed structure for the rest of parameters like in (6.2), (6.3) or (6.4) and we model only the first element of a_i the GLSC remains unbiased, however with some loss of efficiency. To illustrate the different consequences of model misspecification, we use the following simple example.

Example. We consider a simple regression model with random coefficients, $y_{ij} = \alpha_{i1}x_{ij} + \alpha_{i2} + \varepsilon_{ij};\ j = 1,\ldots,n_i;\ i = 1,\ldots,N$ with second stage model $\alpha_{i1} = \gamma$; i.e., $z_i = 1$. For all three models (6.2), (6.3), (6.4), $T_i = 1$. We assume the true model for the intercept term is $a_{i2} = \alpha_{i2} + b_{i2}$; i.e., no restrictions are imposed. We attempt to find the bias induced by imposing structures such as (6.2) or (6.3). For these models

$$X_i'X_i = \begin{bmatrix} \sum_j x_{ij}^2 & \sum_j x_{ij} \\ \sum_j x_{ij} & n_i \end{bmatrix}; \quad M_i = \begin{bmatrix} \sigma^2(X_i'X_i)^{-1}_{11} + \tau_{11} & \sigma^2(X_i'X_i)^{-1}_{12} + \tau_{12} \\ \sigma^2(X_i'X_i)^{-1}_{12} + \tau_{12} & \sigma^2(X_i'X_i)^{-1}_{22} + \tau_{22} \end{bmatrix};$$

where $\tau_{12} = \tau_{22} = 0$ for model (6.2), and $(X_i'X_i)^{-1}_{11}$, for instance, denotes the $(1,1)$th element of the inversed matrix. We obtain

$$\begin{bmatrix} \sum M^{-1}_{i11} z_i' z_i & \sum z_i' M^{-1}_{i12} T_i \\ \sum T_i' M^{-1}_{i12} z_i & \sum T_i' M^{-1}_{i22} T_i \end{bmatrix} = \sum_{i=1}^{N} \frac{1}{m_{i11}m_{i22} - m_{i12}^2} \begin{bmatrix} m_{i22} & -m_{i12} \\ -m_{i12} & m_{i11} \end{bmatrix} \tag{6.8}$$

where m_{ijk} is the (j,k)th element of matrix M_i. Denoting the $(1,1),(1,2)$ and $(2,2)$th elements of matrix (6.8) as l_{11}, l_{12} and l_{22}, and $\Delta = l_{11}l_{22} - l_{12}^2$, the bias from (6.7) can be written as

$$E(\hat{\gamma}) - \gamma = \frac{1}{\Delta}\sum_{i=1}^{N} \frac{(m_{i12}l_{22} - m_{i11}l_{12})(\alpha_{i2} - \bar{\alpha})}{m_{i11}m_{i22} - m_{i12}^2}.$$

The maximum bias, as follows from the Cauchy inequality, is

$$\max_{\|\alpha_{i2} - \bar{\alpha}\| = \delta} |E(\hat{\gamma}) - \gamma| = \frac{\delta}{\Delta}\sqrt{\sum \left(\frac{(M_{i12}l_{22} - M_{i11}l_{12})}{M_{i11}M_{i22} - M_{i12}^2}\right)^2}. \tag{6.9}$$

Simulations, however, show that when classical models are applied to general growth curve with unknown variance parameters, the bias might be negligable (Stukel and Demidenko 1997). This can be explained in the following way, a simple case with two parameters is considered $(m = 2)$. Then, letting $\tau_{12} = \rho \to \infty$ and $\tau_{12} = \rho^2$ we obtain

$$M_i^{-1} = \frac{1}{(h_{11i} + \tau_{11})(h_{22i} + \rho^2)} \begin{bmatrix} h_{22i} + \rho^2 & -(h_{12i} + \rho) \\ (h_{12i} + \rho) & h_{11i} + \tau_{11} \end{bmatrix}$$

where $h_{kli} = \sigma^2(X_i'X_i)^{-1}_{kl}$. Letting $\rho \to \infty$ we obtain

$$M_i^{-1} \to \begin{bmatrix} (\sigma^2(X_i'X_i)^{-1}_{11} + \tau_{11})^{-1} & 0 \\ 0 & 0 \end{bmatrix};$$

and we come to the estimate (6.6). Thus, a perculiar compensation effect takes place: when in the growth curve model in the standard setting (1.1,1.2) the second stage model for the first parameter is correctly specified and misspecified for the rest of parameters, one still obtains almost unbiased estimates of betas for the first parameter, however variances of random effects will be overestimated, and a result the variances for beta-estimates will be underestimated.

We now investigate underspecification of the second stage model, i.e., when the true model is one of (6.2), (6.3) or (6.4), but we do not specify $\{\alpha_{i2}\}$ leaving them unrestriced. As was pointed out above, the GLSC remains unbiased. However there may be an efficiency loss we are to determine.The variance of GLSC is $(\sum M_{i11}^{-1})^{-1}.$ Denoting $\Delta_i = M_{i11}M_{i22} - M_{i12}^2$ as the determinant of the 2×2 matrix M_i; the variance of the GLS-estimator for γ is written as

$$\begin{aligned} \mathrm{var}(\hat{\gamma}) &= \Delta^{-1} = \left[\sum M_{i22}\Delta_i^{-1} - \left(\sum M_{i12}\Delta_i^{-1}\right)^2\left(\sum M_{i11}\Delta_i^{-1}\right)^{-1}\right]^{-1} \\ &= \left\{\sum M_{i11}^{-1} + \left(\sum M_{i12}^2 M_{i11}^{-1}\Delta_i^{-1} - \left(\sum M_{i12}\Delta_i^{-1}\right)^2\left(\sum M_{i11}\Delta_i^{-1}\right)\right)^{-1}\right\}^{-1}. \end{aligned}$$

We show that

$$0 \le \sum M_{i12}^2 M_{i11}^{-1} d_i^{-1} - \left(\sum M_{i12} d_i^{-1}\right)^2\left(\sum M_{i11} d_i^{-1}\right) \le \sum M_{i12}^2 M_{i11}^{-1} d_i^{-1}. \quad (6.10)$$

The right inequality is obvious, and the left one follows from the Cauchy inequality:

$$\left(\sum M_{i12} d_i^{-1}\right)^2 = \left(\sum M_{i12} M_{i11}^{-1/2} d_i^{-1/2} \cdot M_{i11}^{1/2} d_i^{-1/2}\right)^2 \le \sum M_{i12}^2 M_{i11}^{-1} d_i^{-1} \cdot \left(\sum M_{i11} d_i^{-1}\right).$$

Based on (6.10) one can find the best and the worst situation for the GLSC. Indeed, there will be no loss in efficiency if and only if for some λ we have $M_{i12} = \lambda M_{i22}$ for all $i = 1, \ldots, N$. For instance, it is true for balanced data. In contrast, maximum loss of efficiency is when $\sum M_{i12} d_i^{-1} = 0$.

7. Effect of sex hormones on pancreatic tumor growth

Pathologists performed an experiment on laboratory rats to assess whether sex hormones modulate the growth rate of pancreatic carcinoma during the early stages (Sumi, Brinck-Johnsen, and Longnecker, 1989). The inhibitory effect of estrogen was expected to cause these cancers to grow less rapidly in females. Pancreatic tumors were transplanted subcutaneously into 22 rats. The tumor was usually palpable within two to three weeks. Tumor diameters were measured with calipers in three orthogonal directions every two to five days for several weeks and were used to calculate the tumor volume assuming that tumors were ellipsoid. Rats were killed and autopsied when the tumor reached 30 to 40 cm^3. This resulted in 3 to 13 observations per tumor (Figure 1). While unperturbed tumors are expected to increase continuously in volume, there were occasional decreases due to the probable rupturing of cysts. Since rats with slower growing tumors were followed longer, the censoring mechanism was dependent on the response, so that the missing data was non-ignorable. We fitted the linear growth curve to the logarithm of the tumor volumes $\log(V_{ij}+1)$ using a simple linear trend model

$$\log(V_{ij}+1) = a_{i1}t_{ij} + a_{i2} + \epsilon_{ij}; \quad j = 1;\ldots;n_i;\ i = 1;\ldots;N = 22 \tag{7.1}$$

since this was consistent with known patterns of early growth. We assessed the effect of host sex on the rate of growth 20 days post transplantation using the second-stage model for the slope

$$a_{i1} = \beta_1 + \beta_2 \mathrm{sex}_i + b_{i1} \tag{7.2}$$

where sex = 1 for males, 0 for females. For the general model, the intercept term, a_{i2}; which can be accociated with the individual immunity, is left unspecified. By contrast, in the classical growth curve model requires specification of the intercept term: (i) a_{i2} is a fixed unknown parameter, $a_{i2} = \beta_3$; (ii) a_{i2} is random with unknown mean, $a_{i2} = \beta_3 + b_{i2}$; (iii) a_{i2} is modelled via sex, $a_{i2} = \beta_3 + \beta_4 \mathrm{sex}_i + b_{i2}$:

Table 1. Results of estimation based on four models

	Two-step		Standard Growth Curve Formulation					
	a_{i2} are not specified		$a_{i2} = \beta_3$		$a_{i2} = \beta_3 + b_{i2}$		$a_{i2} = \beta_3 + \beta_4 \mathrm{sex}_i + b_{i2}$	
	Estimate	%Bias	Estimate	%Bias	Estimate	%Bias	Estimate	%Bias
β_1	0.0376	0.0	0.0311	-10.8	0.0377	6.7	0.0362	18.4
(SE)	(0.00214)		(0.0028)		(0.0040)		(0.0054)	
β_2	0.00457	0.0	0.0011	-103.7	0.924×10^{-5}	-135.7	0.00293	-62.6
(SE)	(0.00284)		(0.0028)		0.0032		(0.0073)	
SE(b_{i1})	0.0112	NA	0.0061	NA	0.0147	NA	0.0147	NA

We start with estimating (7.1) by the OLS applying it individually to each equation of (7.2). There is a slight positive slope in regressing a^o_{i1} on sex. The results of estimation based on four models are presented in Table 1. The bias is calculated based on the assumption that $\alpha_{i2} = a^o_{i2}$: As the reader can see, there is a significant bias when one imposes a fixed structure on the intercept term when, in fact, it does not exist. As the reader can see, accepting the general growth curve setting, suggested in the present paper, leads to much more distinguished e®ect of sex.

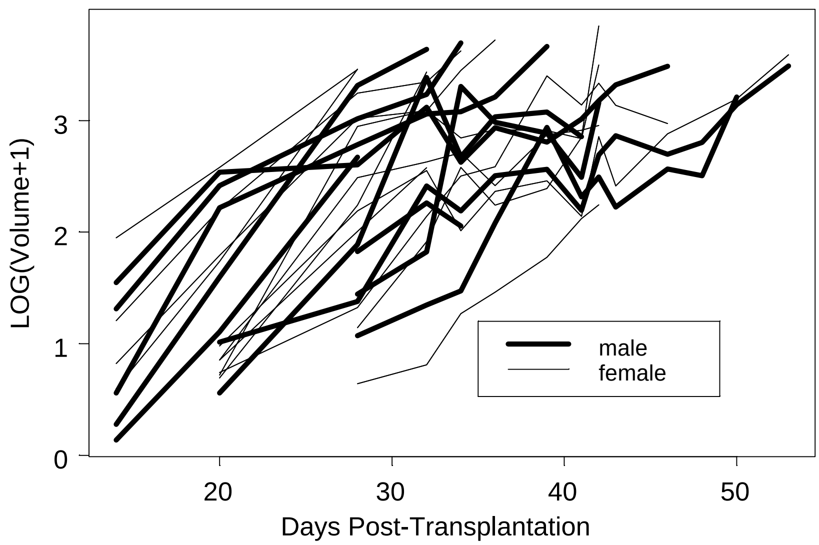

Figure 1. Individual longitudinal data for the tumor volume of 22 rats.

8. Conclusion

Frequently, only part of parameters or, generally, linear combinations of parameters can be modeled with certain degree of con¯dence. Applying the standard approach of linear growth curve models the researches is forced to model the rest of parameters. One can set the mean to constant or use the same sets of explanatory variables. As it is shown in this paper, when the data are unbalanced, the misspeci¯cation in the rest of parameters can lead to a severe bias of the parameters of interest. Principally, this phenomena is known from the multivariate regression: omitting a variable leads to a bias, in contrast, introducing an extra variable leaves the OLS unbiased, however, with some loss of e±ciency. Summarizing, the general growth curve approach suggested in this paper makes the model more robust to misspeci¯cation. The moment estimators of variance parameters developed in this paper are unbiased and consistent under mild conditions. Substituting them into the GLS-under restriction estimator gives us an asymptotically e±cient estimator of parameters of interest.

9. Appendix

9.1. Deriviation of moment estimator for B

Let us introduce a $r \times 1$ random vector $u_i = Ca_i^o - Z_i\beta = C(X_i'X_i)^{-1}X_i'\varepsilon_i$ where β is the true parameter vector. Then, it is easy to see that matrix (4.1) times N can be rewritten as follows:

$$\sum_i u_iu_i' - \sum_i u_i(\sum_j u_j'Z_j)HZ_i' - \sum_i Z_iH(\sum_j Z_j'u_j)u_i' + \sum_i Z_iH(\sum_j Z_j'u_j)(\sum_j u_j'Z_j)HZ_i'. \tag{9.1}$$

where $H = (\sum Z_i'Z_i)^{-1}$. Our aim is to find the expectation of (9.1). Since u_i and u_j are independent for $i \neq j$, we obtain

$$\begin{aligned} E\sum_i u_iu_i' &= NB + \sigma^2\sum P_i;\\ E(\sum_i u_i(\sum_j u_j'Z_j)HZ_i') &= E(\sum_i u_iu_i'Z_iHZ_i') = \sigma^2\sum_i P_iQ_i + B\sum Q_i;\\ E(\sum_i Z_iH(\sum_j Z_j'u_j)u_i') &= \sigma^2\sum_i Q_iP_i + \sum Q_iB;\\ E(\sum_i Z_iH(\sum_j Z_j'u_j)(\sum_j u_j'Z_j)HZ_i') &= E(\sum_i Z_iHZ_i'u_iu_i'Z_iHZ_i')\\ &= \sigma^2\sum Q_iP_iQ_i + \sum Q_iBQ_i. \end{aligned}$$

Combining these results we come to

$$\begin{aligned} E\sum(Ca_i^o - Z_i\hat\beta_o)(Ca_i^o - Z_i\hat\beta_o)' &= NB + \sigma^2\sum P_i - \sigma^2\sum_i P_iQ_i - B\sum Q_i\\ &- \sigma^2\sum_i Q_iP_i - \sum Q_iB + \sigma^2\sum Q_iP_iQ_i + \sum Q_iBQ_i. \end{aligned} \tag{9.2}$$

By definition, $\hat B$ is the solution to the matrix equation

$$B - B\frac{1}{N}\sum Q_i - \frac{1}{N}\sum Q_iB + \frac{1}{N}\sum Q_iBQ_i = L \tag{9.3}$$

where σ^2 is substituted by the pooled variance (3.2), and

$$L = \frac{1}{N}\sum_{i=1}^{N}\{(Ca_i^o - Z_i\hat\beta_o)(Ca_i^o - Z_i\hat\beta_o)' - \hat\sigma^2[P_i - P_iQ_i - Q_iP_i + Q_iP_iQ_i]\}. \tag{9.4}$$

In order to solve (9.3) for B we employ the vec and vech functions (Fuller (1987), Magnus (1988)). Hence, using the formula $\mathrm{vec}(ABF) = (F' \otimes A)\mathrm{vec}(B)$ we obtain

$$\begin{aligned} \mathrm{vech}(B\frac{1}{N}\sum Q_i) &= \frac{1}{N}D^+\mathrm{vec}(I_kB\sum Q_i) = \frac{1}{N}D^+(\sum Q_i \otimes I_k)D\,\mathrm{vech}(B);\\ \mathrm{vech}(\frac{1}{N}\sum Q_iB) &= \frac{1}{N}D^+\mathrm{vec}(\sum Q_iBI_k) = \frac{1}{N}D^+(I_k \otimes \sum Q_i)D\,\mathrm{vech}(B);\\ \mathrm{vech}(\frac{1}{N}\sum Q_iBQ_i) &= \frac{1}{N}D^+\sum\mathrm{vec}(Q_iBQ_i) = \frac{1}{N}D^+\sum(Q_i \otimes Q_i)D\,\mathrm{vech}(B) \end{aligned}$$

that leads to the estimator (4.2). To be certain that the solution (4.2) exists we need to show that matrix J is nonsingular. This is true if $2r < N$. Indeed, let U be any $r \times r$ matrix. Without loss of generality we can assume it is normalized, i.e., $\mathrm{tr}(U'U) = 1$. Since matrices D and D^+ are of full rank to prove that matrix J is nonsingular it suffices to show that

$$\mathrm{vec}'(U)\left[I_{r^2} - \frac{1}{N}\sum_{i=1}^{N}\left(Q_i \otimes I_k + I_k \otimes Q_i - Q_i \otimes Q_i'\right)\right]\mathrm{vec}(U) > 0. \qquad (9.5)$$

Using the formula $\mathrm{vec}'(U)(A \otimes B)\mathrm{vec}(U) = \mathrm{tr}(UAU'B')$ and applying the inequality $\mathrm{tr}(AB) \le \mathrm{tr}(A)\mathrm{tr}(B)$ for any nonnegative definite matrices A and B one obtains:

$$\frac{1}{N}\left(\mathrm{vec}'(U)\left(\sum_i Q_i \otimes I\right)\mathrm{vec}(U)\right) = \frac{1}{N}\left(\mathrm{vec}'(U)\left(\sum_i Q_i \otimes I\right)\mathrm{vec}(U)\right) = \frac{1}{N}\sum_i \mathrm{tr}(UQ_iU')$$
$$= \frac{1}{N}\sum_i \mathrm{tr}(Q_iU'U) \le \frac{1}{N}\sum_i \mathrm{tr}(Q_i) = \frac{1}{N}\sum_i \mathrm{tr}(Z_iHZ_i') = \frac{1}{N}\sum_i \mathrm{tr}(HZ_i'Z_i) = \frac{r}{N}.$$

Since $\mathrm{vec}'(U)(Q_i \otimes Q_i)\mathrm{vec}(U) \ge 0$, the left-hand side of the inequality (9.5) is more than $1 - 2r/N$. Therefore, if $2r < N$ matrix J is nonsingular.

9.2. Proof of Theorem 3.1

To prove the theorem the following results will be used.

Lemma 1. The following holds:

1. Let $U_1, \ldots, U_N$ be a sequence of $r \times r$ random matrices, then $\mathrm{plim}_{N\to\infty} U_N = U_*$ if and only if for any nonrandom $r \times 1$ vector of unit length, t we have $\mathrm{plim}_{N\to\infty} t'U_Nt = t'U_*t$.

2. Let $U_1, \ldots, U_N$ be a sequence of $r \times r$ random symmetric nonnegative definite matrices and $\lim_{N\to\infty} E(\mathrm{tr}(U_N)) = 0$, then $\mathrm{plim}_{N\to\infty}(U_N) = 0$.

3. Let u be a $r \times 1$ random vector with $E(u) = 0$ and $\mathrm{cov}(u) = C$, $E(u_i^4) \le \lambda$. If A is a $r \times r$ symmetric matrix, then $E(u'Au) = \mathrm{tr}(AC)$ and $E(u'Au)^2 \le \lambda(\sum |A_{ij}|)^2 \le r\lambda\,\mathrm{tr}(A^2)$.

4. Let $\{u_i\}$ be $r \times 1$ independently distributed vectors with $E(u_i) = 0$, and $\mathrm{cov}(u_i) = C_i$. Let A_i and B_i be any fixed $m \times r$ matrices, $i = 1, \ldots, N$. Then

$$\mathrm{p}\lim_{N\to\infty} \frac{1}{N}\left(\sum_i A_iu_i\right)'\left(\sum_i B_iu_i\right) = 0 \quad \text{if} \quad \lim_{N\to\infty}\frac{1}{N^2}\mathrm{tr}\left(\sum_i A_i'A_iC_i\right)\mathrm{tr}\left(\sum_i B_i'B_iC_i\right) = 0.$$

Proof. The proof of 1 is obvious. The proof of 2 follows from the Chebyshev inequality. To prove 3 we apply the Cauchy inequality twice:

$$E(u'Au)^2 = E\sum_{i,j,k,l}(A_{ij}u_iu_j)(A_{kl}u_ku_l) \le \sum_{i,j,k,l}|A_{ij}||A_{kl}|\sqrt{E(u_iu_j)^2}\sqrt{E(u_ku_l)^2}$$

$$\leq \sum_{i,j,k,l} |A_{ij}||A_{kl}| \sqrt{\sqrt{E(u_i^4)E(u_j^4)E(u_k^4)E(u_l^4)}} = \lambda\Big(\sum_{i,j} |A_{ij}|\Big)^2.$$

The last inequality in 3 follows from $(\sum_{i,j} |A_{ij}|)^2 \leq r(\sum_{i,j} A_{ij}^2) = r \cdot tr(A^2)$. To prove 4 we first apply the Cauchy inequality, and then use 3,

$$\begin{aligned}
var\left(\Big(\sum_i A_i u_i\Big)'\Big(\sum_i B_i u_i\Big)\right) &\leq E\left(\Big(\sum_i A_i u_i\Big)'\Big(\sum_i B_i u_i\Big)\right)^2 \\
&= E\left(\Big(\sum_i A_i u_i\Big)'\Big(\sum_i A_i u_i\Big)\right) E\left(\Big(\sum_i B_i u_i\Big)'\Big(\sum_i B_i u_i\Big)\right) \\
&= \sum tr(A_i A_i' C_i) \sum tr(B_i B_i' C_i).
\end{aligned}$$

Proof of the Theorem. The consistency of $\hat{\sigma}^2$. Using 3 of the above Lemma and assumptions AB(a) and AB(c) we obtain

$$\begin{aligned}
var(\hat{\sigma}^2) &= \frac{1}{(\sum(n_i - m_i))^2} \sum var\Big(y_i'(I - X_i(X_i'X_i)^{-1}X_i')y_i\Big) \\
&= \frac{1}{(\sum(n_i - m_i))^2} \sum var\Big(\varepsilon_i'(I - X_i(X_i'X_i)^{-1}X_i')\varepsilon_i\Big) \\
&\leq \frac{D_4}{(\sum(n_i - m_i))^2} \sum n_i(n_i - m_i) \leq \frac{D_4 D_0}{\sum(n_i - m_i)} \to 0, \quad N \to \infty.
\end{aligned}$$

The equivalence of (4.2) and (4.4). We start with proving that

$$\lim \frac{1}{N}\sum P_i Q_i = \lim \frac{1}{N}\sum Q_i P_i = \lim \frac{1}{N}\sum Q_i P_i Q_i = 0, \qquad N \to \infty. \tag{9.6}$$

Since matrices P_iQ_i, Q_iP_i and $Q_iP_iQ_i$ have nonnegative eigenvalues it suffices to show that the trace of matrices in (9.6) vanish. Using $tr(AB) \leq tr(A)tr(B)$ for any nonnegative definite matrices A and B, and the assumption AB(b) we obtain

$$\begin{aligned}
tr\Big(\sum P_i Q_i\Big) &= \sum tr(P_i Q_i) \leq \sum tr(P_i)tr(Q_i) \leq d_X^{-1} tr(CC') \sum tr(Q_i) \\
&= d_X^{-1} tr(CC') tr\Big(\sum (Z_i'Z_i)^{-1} \sum Z_i'Z_i\Big) = d_X^{-1} k \cdot tr(CC');
\end{aligned}$$

and consequently, the first two limits in (9.6) are zero. To prove the third limit we proceed similarly:

$$tr\Big(\sum Q_i P_i Q_i\Big) \leq d_X^{-1} tr(CC') \sum tr(Q_i^2) \leq d_X^{-1} tr(CC')\Big(\sum tr(Q_i)\Big)^2 = d_X^{-1} k^2 tr(CC').$$

To prove the equivalence of $\hat{B}_s$ to $\hat{B}$ it is left to show that $J \to I_{r^2}$. Indeed,

$$\lim \frac{1}{N}\sum (Q_i - I) = \lim \frac{1}{N}\sum (I - Q_i) = \lim \frac{1}{N}\sum (Q_i - Q_i) = 0. \tag{9.7}$$

which follows from the fact $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$.

Consistency of (4.4). As follows from 1 of the above Lemma it suffices to consider a quadratic form of (9.1) times $1/N$. Thus, let t be a fixed $r \times 1$ vector, $\| t \| = 1$. For the first term of (9.1), using AB(b) and AB(c) and 3 of the Lemma, we obtain when $N \to \infty$

$$
\begin{aligned}
\mathrm{var}[\frac{1}{N}t'(\sum_i u_i u_i')t] &= \frac{1}{N^2}\sum \mathrm{var}(\varepsilon_i' X_i(X_i'X_i)^{-1}C'tt'C(X_i'X_i)^{-1}X_i'\varepsilon_i) \\
+\frac{1}{N^2}\sum \mathrm{var}(\delta_i' t t' \delta_i) &\le \frac{D_4}{N^2}\sum \left(t'C(X_i'X_i)^{-1}X_i'X_i(X_i'X_i)^{-1}C't\right)^2 \\
+\frac{D_4}{N^2}\sum (t't)^2 &\le \frac{D_4}{N}\left(d_X^{-2}\mathrm{tr}^2(CC') + 1\right) \to 0.
\end{aligned}
$$

For the fourth term in (9.1) we apply 2 of the Lemma:

$$
\begin{aligned}
& E\left(\mathrm{tr}\frac{1}{N}\sum_i Z_i H(\sum_j Z_j'u_j)(\sum_j u_j'Z_j)HZ_i'\right) \\
= {} & \frac{1}{N}\mathrm{tr}\sum_i Z_i H\left(\sum_j Z_j' E(u_j u_j')Z_j\right)HZ_i' \\
\le {} & \frac{1}{N}\sigma^2 d_X^{-1}\mathrm{tr}(CC' + B)\mathrm{tr}\left(\sum_i Z_i H(\sum_j Z_j'Z_j)HZ_i'\right) \\
= {} & \frac{1}{N}\sigma^2 d_X^{-1}\mathrm{tr}(CC' + B)\mathrm{tr}\left(H\sum_i Z_i'Z_i\right) = \sigma^2 d_X^{-1} k N^{-1}\mathrm{tr}(CC' + B) \to 0.
\end{aligned}
$$

For the second term of (9.1) after multiplying by t' and t we obtain

$$
t'\sum_i u_i(\sum_j u_j'Z_j)HZ_i't = (\sum_i t'u_i t Z_i)(H\sum_i Z_i'u_i) = (\sum_i Z_i't'tu_i)'(\sum_i HZ_i'u_i).
$$

Now we apply 4 of the Lemma noting that

$$
\begin{aligned}
& \frac{1}{N^2}\sum \mathrm{tr}(Z_i't't t'Z_i(\sigma^2 C(X_i'X_i)^{-1}C' + B))\sum \mathrm{tr}(Z_i HHZ_i'(\sigma^2 C(X_i'X_i)^{-1}C' + B)) \\
\le {} & \frac{\sigma^4 d_X^{-2}\mathrm{tr}^2(CC' + B)}{N^2}\sum \mathrm{tr}(Z_i'Z_i)\sum \mathrm{tr}(Z_i HHZ_i') = \frac{\sigma^4 d_X^{-2}\mathrm{tr}^2(CC' + B)k^2}{N^2} \to 0.
\end{aligned}
$$

Finally, since the variances of all four terms of (9.1) times $1/N$ converge to zero we infer that $\hat{B}$ and $\hat{B}_s$ are consistent.

Asymptotic covariance matrix for $\hat{B}$ and $\hat{B}_s$. It is easy to see that $u_i = \delta_i + C(X_i'X_i)^{-1}X_i'\varepsilon_i = \delta_i + T_i\varepsilon_i$, where $T_i = C(X_i'X_i)^{-1}X_i'$, and $\hat{\sigma}^2 = \sum \varepsilon_i' W_i \varepsilon_i$ where

$W_i = (I - X_i(X_i'X_i)^{-1}X_i')/\sum(n_i - m)$. Therefore, asymptotically, $\hat{B}$ is equivalent to

$$\sum\left[(\delta_i + T_i\varepsilon_i)(\delta_i + T_i\varepsilon_i)' - P_i\varepsilon_i'W_i\varepsilon_i\right],$$

and

$$\operatorname{cov}(\operatorname{vech}(\hat{B})) \simeq \frac{1}{N^2}D^{+}\sum \operatorname{cov}\left\{\operatorname{vec}\left((\delta_i + T_i\varepsilon_i)(\delta_i + T_i\varepsilon_i)' - P_i\varepsilon_i'W_i\varepsilon_i\right)\right\}D^{+\prime},$$

where

$$\begin{aligned} &\operatorname{vec}\left((\delta_i + T_i\varepsilon_i)(\delta_i + T_i\varepsilon_i)' - \varepsilon_i'W_i\varepsilon_iP_i\right) \\ = \;& (\delta_i \otimes \delta_i) - (T_i\varepsilon_i \otimes \delta_i + \delta_i \otimes T_i\varepsilon_i) + M_i(\varepsilon_i \otimes \varepsilon_i) \qquad (9.8) \end{aligned}$$

and $M_i = T_i \otimes T_i - \operatorname{vec}(P_i)\operatorname{vec}'(W_i)$. Now we will use the following fact: let u be a normal $r \times 1$ vector, then $\operatorname{cov}(u \otimes u) = 2N(C_u \otimes C_u)$ where $N = (I + K)/2$ and K is the $r^2 \times r^2$ commutation matrix and $\operatorname{cov}(u) = C_u$ (e.g. Magnus 1988, p. 164). We consider three terms in (9.8). For the first term $\operatorname{cov}(\delta_i \otimes \delta_i) = 2(B \otimes B)N$. The second term is uncorrelated with other two. In fact, due to the independence of δ_i and ε_i,

$$\begin{aligned} E(T_i\varepsilon_i \otimes \delta_i + \delta_i \otimes T_i\varepsilon_i) &= E(T_i\varepsilon_i \otimes \delta_i) + E(\delta_i \otimes T_i\varepsilon_i) \\ &= (ET_i\varepsilon_i) \otimes E(\delta_i) + E(\delta_i) \otimes E(T_i\varepsilon_i) = 0, \end{aligned}$$

and therefore,

$$\operatorname{cov}(\delta_i \otimes \delta_i, T_i\varepsilon_i \otimes \delta_i + \delta_i \otimes T_i\varepsilon_i) = E(\delta_i\varepsilon_i'T_i' \otimes \delta_i\delta_i') + E(\delta_i\delta_i' \otimes \delta_i\varepsilon_i'T_i').$$

However, again, due to the independence, it is easy to see that $E(\delta_i\varepsilon_i'T_i' \otimes \delta_i\delta_i')$ and $E(\delta_i\delta_i' \otimes \delta_i\varepsilon_i'T_i')$ vanish. Thus, for the third term of (9.8) we have $\operatorname{cov}(M_i(\varepsilon_i \otimes \varepsilon_i)) = 2\sigma^4 N M_i M_i'$. Now we have

$$\begin{aligned} M_iM_i' &= (T_i \otimes T_i - \operatorname{vec}(P_i) \otimes \operatorname{vec}'(W_i))(T_i' \otimes T_i' - \operatorname{vec}(W_i) \otimes \operatorname{vec}'(P_i)) \\ &= T_iT_i' \otimes T_iT_i' + \operatorname{vec}(P_i)\operatorname{vec}'(W_i)\operatorname{vec}(W_i)\operatorname{vec}'(P_i) \\ &\quad - \operatorname{vec}(P_i)\operatorname{vec}'(W_i)(T_i' \otimes T_i') - (T_i \otimes T_i)\operatorname{vec}W_i\operatorname{vec}'(P_i). \qquad (9.9) \end{aligned}$$

The third and the fourth terms vanish because, for instance, for the third term

$$\operatorname{vec}'(W_i)(T_i' \otimes T_i') = ((T_i \otimes T_i)\operatorname{vec}(W_i))' = (\operatorname{vec}(T_iW_iT_i'))' = 0$$

because $T_i W_i = 0$. Similarly we can show that the fourth term in (9.9) vanishes. Now, for the second term of (9.9) we have

$$\begin{aligned} \mathrm{vec}(P_i)\mathrm{vec}'(W_i)\mathrm{vec}(W_i)\mathrm{vec}'(P_i) &= \mathrm{vec}(P_i)\mathrm{vec}'(P_i)\mathrm{tr}(W_i^2) \\ &= \frac{n_i - m}{(\sum (n_i - m))^2}\mathrm{vec}(P_i)\mathrm{vec}'(P_i) \end{aligned}$$

and, noting that $T_i T_i' = C(X_i' X_i)^{-1} C' = P_i$, finally obtain

$$\begin{aligned} &\mathrm{cov}(\mathrm{vech}(\hat{B})) \\ &= \frac{2}{N} D^+ N \left((B \otimes B) + \frac{\sigma^4}{N} \sum_{i=1}^{N} \left(P_i \otimes P_i + \frac{n_i - m}{(\sum (n_i - m))^2}\mathrm{vec}(P_i)\mathrm{vec}'(P_i) \right) \right) D^{+\prime}. \end{aligned}$$

At last we notice that $(n_i - m)(\sum (n_i - m))^{-2} = O(N^{-2})$ and the according term can be omitted. Then we use the property $D^+ N = D^+$; that finally leads to formula (5.1).

9.3. Proof of Theorem 4.1

Full maximum likelihood estimator. In order to prove asymptotic properties of the MLE we apply the maximum likelihood theory with large number of nuisance parameters developed by Andersen (1970). In the case of general growth curve model the nuisance parameters are $(\alpha_1, \ldots, \alpha_N)$ and the parameter of interest is β, or structural (incidental) parameter in the Anderson's terminology. Andersen proved that the MLE applied to conditional likelihood inherits its optimal statistical properties if there are N minimal sufficient statistics $t_1, \ldots, t_N$ not dependent on parameters that the conditional distribution $\{f(y_i \mid t_i), i = 1, \ldots, N\}$ does not depend of nuisance parameters. In our case, the minimal sufficient statistics are $\{Ca_i', i = 1, \ldots, N\}$. As follows from (??), the conditional distribution does not include $(\alpha_1, \ldots, \alpha_N)$; therfore the MLE is consistent, asymptotical normally distributed and efficient.

Second-stage maximum likelihood estimator. To establish asymptotic properties of the according estimator we combine the theory of M-estimation (Huber (1981, Schervish (1995)), or estimating equation approach (Carrol, Ruppert and Stefanski (1995)) with pseudo-maximum likelihood approach (Gong and Samaniego (1981), Parke (1986)). First, we show that l_2 maximization produces consistent estimates for matrix B. As follows from the theory of M-estimation or estimating equation approach it suffices to show that the expectation of the estimating equation for B is zero. Let us assume that in l_2 the true value of σ^2 is used. Then, the estimating equation for B is

$$\frac{\partial l_2}{\partial B} = -\frac{1}{2}\left\{\sum_{i=1}^{N}\left[(P_i+B)^{-1} - (P_i+B)^{-1}(c_i - Z_i\beta)(c_i - Z_i\beta)'(P_i+B)^{-1}\right]\right\} = 0. \tag{9.10}$$

Taking the expectation and recalling that $E(c_i - Z_i\beta)(c_i - Z_i\beta)' = P_i + B$ we see that the expectation of (9.10) is zero. Therefore, $\hat{B}$, as the solution to (9.10), remains consistent when σ^2 is substituted by a consistent estimator $\hat{\sigma}^2$ (e.g., Bierens 1994). Secondly, we apply the threory of pseudo-maximum likelihood (PML) approach . We notice that the expectation of the cross-product derivative of the likelihood function with respect to β and the variance parameter is zero. Therefore, as follows from PML theory, the generalized leat squares estimator (2.9) is efficient because $\hat{\sigma}^2$ and $\hat{B}$ are consistent.

Two-stage estimator. The following general fact will be used to show asymptotic properties of the TS-estimator (Bierens 1994).

Lemma 2. Let $T_N(\theta) = N^{-1/2}\sum_{i=1}^{N} F_i(\theta)\varepsilon_i$ where θ is a $q \times 1$ parameter vector, $F_i(\theta)$ is a $k \times r$ matrix such that $\| F_i(\theta) \| \leq A$; $\| \partial F_i/\partial\theta \| \leq A$ in some neighborhood of θ_0 for $i = 1, ..., N$; and ε_i is a random $r \times 1$ vector, $E(\varepsilon_i) = 0$; $\mathrm{cov}(\varepsilon_i) = B$. Let $\hat{\theta}_N$ be a consistent statistic for θ_0. Then $T_N(\hat{\theta}_N)$ has a normal asymptotic distribution $N(0, \lim N^{-1}\sum F_i(\theta_0) B F_i'(\theta_0))$, assuming the last limit exists, when $N \to \infty$.

Then, as follows from this Lemma, for the GLSC we obtain

$$\sqrt{N}(\hat{\beta} - \beta) = \left(N^{-1}\sum Z_i' M_i^{-1} Z_i\right)^{-1}(T_{N1} + T_{N1})$$

where

$$T_{N1} = N^{-1/2}\sum Z_i M_i^{-1} C\delta_i; \quad T_{N1} = N^{-1/2}\sum Z_i M_i^{-1} C(X_i'X_i)^{-1}X_i'\eta_i).$$

We apply Lemma 2 to show that vectors T_{N1} and T_{N2} are asymptotically normally distributed when $N \to \infty$. Let us start with T_{N1} denoting $F_i = Z_i M_i^{-1} C$. The uniform boundeness of the elements of F_i follows directly from the AB-assumptions. It is easy to see that the same is true for derivatives because

$$\frac{\partial F}{\partial \sigma^2} = -Z_i M_i^{-1} C (X_i'X_i)^{-1} C' M_i^{-1} C; \quad \frac{\partial F}{\partial B_{gh}} = -Z_i M_i^{-1} K_{gh} M_i^{-1} C$$

where K_{gh} is a $r \times r$ matrix which the $(g;h)$th and $(h;g)$th elements are one and zero elsewhere. Thus, the conditions of the Lemma 2 are fulfilled and T_{N1} is asymptotically normally distributed with the covariance matrix C_1; the same arguments apply to T_{N1} which has the covarianvce matrix C_2 where

$$C_1 = \lim N^{-1} \sum Z_i M_i^{-1} C - C' M_i^{-1} Z_i';$$
$$C_2 = \lim N^{-1} \sum Z_i M_i^{-1} (\sigma^2 C (X_i' X_i)^{-1} C') M_i^{-1} Z_i':$$

Since T_{N1} and T_{N2} are independent, the asymptotic covariance of $\sqrt{N}(\hat{\beta} - \beta)$ is the sum of covariances for the first and the second term, the theorem is proved.

References

[1] Andersen, E.B. (1970). "Asymptotic properties of conditional maximum likelihood estimators", Journal of the Royal Statistical Society, ser. B, 2, 283-301.

[2] Anderson, T.W. (1971). The Statistical Analysis of Time Series, New York, Wiley.

[3] Berke, C.S., and Laird, N.M. (1986). "Nonlinear growth curve analysis: estimating the population parameters", Annals of Human Biology,13, 111-128.

[4] Bierens H. J. (1994). Topics in Advanced Econometrics. Cambridge, Cambridge University Press.

[5] Carrol, R.J., Ruppert D., and Stefanski, L.A. (1995). Nonlinear Measurement Error Models, Chapman and Hall, New York.

[6] Gong, G., and Samaniego, F.J. (1981). "Pseudo maximum likelihood estimation: theory and applications", Annals of Statistics, 9, 861-869.

[7] Graybill, F.A. (1983). Matrices with Applications in Statistics, Belmont, Wadsworth Inc.

[8] Fuller, W.A. (1987). Measurement Error Models, New York, Wiley.

[9] Huber, P.J. (1981). Robust Statistics, New York, Wiley.

[10] Jennrich, R.I., and Schluchter, M.D. (1986). "Unbalanced repeated-measures models with structured covariance matrices", Biometrics, 42, 805-820.

[11] Laird, N.M., and Ware, J.H. (1982). "Random-e®ects models for longitudinal data". Biometrics, 38, 963-974.

[12] Laird, N., Lange, N., and Stram, D. (1987). "Maximum likelihood computations with repeated measures: application of the EM algorithm", Journal of American Statistical Association, 82, 97-105.

[13] Lancaster, P., and Tismenetsky, M. (1985). The Theory of Matrices (2nd ed.), New York, Academic Press.

[14] Lindstrom, M. J., and Bates, D.M. (1988). "Newton-Raphson and EM algorithms for linear mixed-e®ects models for repeated-measures data", Journal of American Statistical Association, 83, 1014-1022.

[15] Magnus, J.R. (1988). Linear Structures, Oxford University Press, London.

[16] Parke, W.R. (1986). "Pseudo-maximum likelihood estimation: the asymptotic distribution", Annals of Statistics, 14, 355-357.

[17] Pockock, S.J., Cook, D.G., and Beresford, S.A.A. (1981). "Regression of area mortality rates on explanatory variables: what weighting is appropriate?", Applied Statistics, 30, 286-295.

[18] Potthof, R.F., and Roy, S.N. (1964). "A generalized multivariate analysis of variance model useful especially for growth curve problems", Biometrika, 51, 313-326.

[19] Rao, C.R. (1973). Linear Statistical Inference and its Application, Second Edition, Wiley, Ney York.

[20] Reinsel, G.C. (1985). "Mean squared error properties of empirical Bayes estimators in a multivariate random e®ects general linear model", Journal of American Statistical Association, 80, 642-650.

[21] Schervish, M.J. (1995). Theory of Statistics, New York, Springer-Verlag.

[22] Stukel, T. A., and Demidenko, E. (1997)."E±cient estimation for general linear growth curve models". Biometrics 53, 340-348.

[23] Vonesh, E.F., and Carter, R.L. (1992). "Mixed-e®ects regression for unbalanced repeated measures", Biometrics, 48, 1-17.