E±cient Estimation of General Linear Growth Curve Models

Eugene Demidenko and Therese 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 characteristic 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, e±cient and equivalent when the number of individuals goes to in±nity. The properties of estimators are investigated under model misspeci±cation. 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 e®ects, Random e®ects; 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,

\begin{equation}
    y_i = X_i a_i + \varepsilon_i; \quad i = 1; \ldots; N
\end{equation}

(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 \( \text{rank}(X_i) = m \cdot n \); \( a_i \) is an \( m \times 1 \) vector of random coefficients, and \( \varepsilon_i \) is independent error term with \( \mathbb{E}(\varepsilon_i) = 0 \) and \( \text{cov}(\varepsilon_i) = \sigma^2 I \). The second stage model describes the random coefficients \( a_i \) through population covariates,

\begin{equation}
    a_i = Z_i \beta + b_i
\end{equation}

(1.2)

where \( Z_i \) is a known \( m \times k \) population design matrix; \( \beta \) is a \( k \times 1 \) vector of population parameters of interest, and \( b_i \) is the vector of random effects with \( \mathbb{E}(b_i) = 0 \) and \( \text{cov}(b_i) = \eta ; \) independent of \( \varepsilon_i \). Vector comprises \( \sigma^2 \) and the distinct elements of the matrix \( \eta \) is denoted \( \mu \) 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 \( Z_i = (I - z_i) \); 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 parameter 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

$$Ca_i = Z_i \bar{\gamma} + \pm_i$$

(2.1)

where $C$ is a known $r \times m$ matrix of full rank $r \cdot m$ and $\text{cov}(\pm_i) = C - C^0$. $Z_i$ is a known $r \times k$ design matrix such that $Z_i^0Z_i$ is of full rank. Also, in this paper it is assumed that $m \cdot n_i$ 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:

$$y_i = X_i \bar{\mu} + \varepsilon_i; \quad E(\varepsilon_i) = 0; \quad \text{cov}(\varepsilon_i) = V_i = \frac{\sigma^2}{2}I + X_i^0 X_i$$

(2.2)

with parameters subject to linear constraints

$$C \bar{\mu} = Z_i \bar{\gamma}$$

(2.3)

where $\bar{\mu} = E(a_i); i = 1; \ldots; N$. Thus marginal means $\bar{\mu}$ 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 $\sigma^2$ and $\sigma^2$ are known. Based on model (2.2), the generalized least squares estimator of $\bar{\theta}$ would minimize the following sum of weighted squares:

$$\min_{\hat{\theta}} X_i^2(y_i - X_i \hat{\theta})^2 V_i^{-1}(y_i - X_i \hat{\theta})$$

(2.4)

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

$$L = X_i^2(y_i - X_i \hat{\theta})^2 V_i^{-1}(y_i - X_i \hat{\theta}) + 2 \frac{1}{2} \sum_{i=1}^{N} (C_i \hat{\theta} - Z_i \bar{\theta})$$

where $\bar{\theta}$ is the $r \times 1$ Lagrangian vector of coefficients. Differentiating $L$ with respect to $\hat{\theta}$ and $\bar{\theta}$ we obtain the necessary conditions for the minimum,

$$\frac{1}{2} \frac{\partial L}{\partial \hat{\theta}} = X_i^2(y_i - X_i \hat{\theta}) + C_i \hat{\theta} - Z_i \bar{\theta} = 0; \quad i = 1; \ldots; N$$

(2.5)

and

$$\frac{1}{2} \frac{\partial L}{\partial \bar{\theta}} = X Z_i^2 \hat{\theta} - Z_i \bar{\theta} = 0; \quad i = 1; \ldots; N$$

(2.6)

Define

$$R_i = X_i^2 V_i^{-1} X_i; \quad M_i = C R_i^{-1} C^0; \quad a_i^0 = (X_i^0 V_i^{-1} X_i)^{-1} X_i^0 V_i^{-1} y_i; \quad i = 1; \ldots; N$$

(2.7)

and notice that matrices $R_i$ and $M_i$ are positive definite. Using the matrix identity

$$(P Q I + P Q P Q i)^{-1} = (P Q i)^{-1} + Q; \quad i = 1; \ldots; N$$

we obtain

$$R_i^{-1} = \frac{3}{2} (X_i^0 X_i)^{-1} + \cdot \cdot \cdot; \quad M_i = \frac{3}{2} C (X_i^0 X_i)^{-1} C^0 + B; \quad i = 1; \ldots; N$$

(2.8)

Solving (2.5) and (2.6) in terms of $\bar{\theta}$ we find the solution to (2.4) as

$$\hat{\bar{\theta}} = \frac{3}{X_i} Z_i^0 M_i^{-1} Z_i^{-1} \left( \frac{3}{X_i} X_i^0 M_i^{-1} \right)^{-1} X_i^0 V_i^{-1} y_i$$

(2.9)

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

$$\hat{\bar{\theta}} = \frac{3}{X_i} Z_i^0 M_i^{-1} Z_i^{-1} \left( \frac{3}{X_i} X_i^0 M_i^{-1} \right)^{-1} X_i^0 V_i^{-1} y_i$$

(2.9)

where $a_i^0 = (X_i^0 X_i)^{-1} X_i^0 y_i$ is the ordinary least squares estimator of the $i$th regression (1.1). The covariance matrix $\text{cov}(\hat{\theta}) = (P \cdot Z_i^0 M_i^{-1} Z_i^{-1})^{1/2}$ as follows from the derived
formula $\hat{\alpha}$ 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 $Ca_0^i$ 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 $b$ 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 $\frac{3}{2}$ and $b$. From formula (2.8), the GLSC depends on $B = C - C^0$, and consequently we need only estimators for $\frac{3}{2}$ 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 $\hat{\alpha}$ is defined by (2.9). Otherwise we have to estimate $\hat{\alpha}$ and the variance parameters simultaneously based on the log-likelihood function for (2.2):

$$l_i = \frac{1}{2} \left( \sum_{i=1}^{N} \left( \ln \frac{1}{2} \mathbf{I} + \mathbf{X}_{i-}^{T} \mathbf{P}_{i} \mathbf{X}_{i-} \right) \right) + \left( \mathbf{y}_{i} - \mathbf{X}_{i} \mathbf{\alpha}_{i} \right)^{T} \left( \frac{1}{2} \mathbf{P}_{i} + \mathbf{X}_{i}^{T} \mathbf{X}_{i} \right)^{-1} \left( \mathbf{y}_{i} - \mathbf{X}_{i} \mathbf{\alpha}_{i} \right)$$

where $\mathbf{X}_{i}$ and $\mathbf{\alpha}_{i}$; $\ldots; \mathbf{\alpha}_{N}$ are subject to constraints (2.3). Our aim is to find the concentrated log-likelihood function excluding nuisance parameters that expresses $l$ in terms of $\hat{\alpha}$ and the variance parameters. Thus we minimize (3.1) under constraints (2.3) assuming $\hat{\alpha}$ is fixed. After some algebra we come to the concentrated log-likelihood function

$$l_{c}(\hat{\alpha}; \frac{3}{2}; b) = \frac{1}{2} \sum_{i=1}^{N} \left( \mathbf{X}_{i}^{T} \mathbf{X}_{i} \right)^{1/2} \ln \frac{1}{2} \left( \mathbf{P}_{i} + \mathbf{X}_{i}^{T} \mathbf{X}_{i} \right)^{-1} \left( \mathbf{y}_{i} - \mathbf{X}_{i} \mathbf{\alpha}_{i} \right)^{T} \left( \mathbf{P}_{i} + \mathbf{X}_{i}^{T} \mathbf{X}_{i} \right)^{-1} \left( \mathbf{y}_{i} - \mathbf{X}_{i} \mathbf{\alpha}_{i} \right)$$

where $\mathbf{C}_{i} = Ca_{0}^i$ and $\mathbf{P}_{i} = C(X_{i}^{T}X_{i})^{-1}C^0$. 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 $c_i = Ca_i^0$ as known we can estimate $\theta$ using the second stage model (2.1) alone. It is easy to show that $\text{cov}(c_i) = \frac{3}{2}P_i + B$: Thus, given $fP_i g$ the second stage model is equivalent to $c_i \sim N(Zi^-; \frac{3}{2}P_i + B)$: There exists an unbiased and consistent estimator of $\frac{3}{2}$ as the pooled variance (Vonesh and Carter 1992).

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

$$l_2(\bar{\theta}; B) = i \frac{1}{2} \left( \text{ln} \theta^2 P_i + B + (c_i - Z_i^-)^{T} \theta^2 P_i + B \right) \ln \left( c_i - Z_i^- \right) \right)$$

(3.3)

A simple algorithm for maximizing (3.3) in the case of one characteristic based on the fixed-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 $fZ_i g$. A natural estimate of $B$ is based on the second stage model (2.1), once $a_i^0$ are estimated as $a_i^0$: Then the OLS-estimate for $\theta$ is $\hat{\theta}_o = (Z_i^0 Z_i)^{T} \theta^2 P_i + B$; and the empirical residual for the second stage model is $C_i a_i^0 - Z_i \hat{\theta}_o$: Following the idea of the method of moments, we construct the matrix of cross-products of residuals

$$\frac{1}{N} \sum_{i=1}^{X} (C_i a_i^0 - Z_i \hat{\theta}_o)(C_i a_i^0 - Z_i \hat{\theta}_o)^T.$$

(4.1)

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

$$\text{vech}(\hat{B}) = \sum_{i=1}^{X} f(C_i a_i^0 - Z_i \hat{\theta}_o)(C_i a_i^0 - Z_i \hat{\theta}_o)^T \theta^2 P_i + B \left( P_i Q_i + Q_i P_i \right) \left( P_i Q_i + Q_i P_i \right) \text{vech}(\hat{B})$$

(4.2)
where

\[ J = D^+ I_{r^2} i \left( \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i \right) Q_i - I_r + I_r \cdot Q_i \cdot \left( Q_i - Q_i^0 \right)^\# \ D \] (4.3)

and \( Q_i = Z_i(P_j Z_i^0 Z_i)^i Z_i^0 \). Here \( D \) is the \( r^2 \times (r + 1) \) duplication matrix and \( D^+ \) is its general inverse (Lancaster and Tismenetsky 1985, Fuller 1987, Magnus 1988). Let us recall that \( \text{vech} \) is the \( r \times (r + 1) \) vector function of a \( r \times r \) symmetric matrix and \( \text{vec} \) is the \( r^2 \times 1 \) vector function. These functions are linearly related to each other as \( \text{vec} = D\text{vech} \) and \( \text{vech} = D^+\text{vec} \). As is shown in the Appendix, matrix \( J \) is nonsingular if \( 2r < N \) which is a very mild condition. Since \( \mathbf{B} \) is a linear function of (4.1) and \( \mathbf{b}_2 \), 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

\[ \mathbf{B}_S = \frac{1}{N} \left( \sum_{i=1}^{N} \left( \mathbf{C}_i \mathbf{a}_i - Z_i \mathbf{b}_i \right) \left( \mathbf{C}_i \mathbf{a}_i - Z_i \mathbf{b}_i \right)^\# \right) \mathbf{P}_i \] (4.4)

This estimator has a clear variance components interpretation from the formula

\[ \text{var}(\mathbf{a}_i \mid Z_i) = \text{var}(\mathbf{a}_i) + \text{var}(\#) \].

The asymptotic properties and variances for these estimators are provided below.

4.1. Two-stage estimator

Since \( \mathbf{b}_2 \) and \( \mathbf{B} \) 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 \( \mathbf{b}_2 \) and \( \mathbf{B} \) are estimated as (3.2) and (4.2) or (4.4), and at the second step \( \mathbf{B} \) 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 \( N \rightarrow 1 \):

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

\begin{align*}
\text{(a) } n_i \cdot D_0; \\
\text{(b) } d_X \cdot X_i^0 X_i;
\end{align*}

7
(c) $E_{ij}^2 < D_4$; $E(a_{il} E(a_{lj}))^4 < D_4$; (d) $jZ_{il} j \cdot D_Z$

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

**Theorem 5.1.** Under the AB-assumptions:

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

$$
\frac{2}{N} D_{+} A - B + \frac{3}{N} \sum_{i=1}^{N} P_i - P_i D_{+}^0. \tag{5.1}
$$

**Proof** is in the Appendix.

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

$$
\bar{\theta} = \frac{1}{N} \sum_{i} \frac{1}{2k} \left( C a_i^0 Z_i b_o^0 \right)^2 \bar{b}_i \bar{p}(1 - q)^2
$$

$$
\bar{\theta}_s = \frac{1}{N} \sum_{i} \frac{1}{2k} \left( C a_i^0 Z_i b_o^0 \right)^2 \bar{b}_i \bar{p}
$$

where $q = Q_i$ and $p = P_i$ are scalars, with asymptotic variance

$$
\text{var}(\hat{\theta}_s) = \text{var}(\hat{\theta}) = \frac{2}{N} B + \frac{A}{N} \bar{p}^2. \tag{5.2}
$$

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 use the assumption on normality of error terms.

In the next theorem, we show that all three estimators for $\hat{\theta}$ 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 $\frac{1}{N} \left( \bar{b} \mid \hat{\theta} \right)^\prime \cdot N(0; H)$; where

$$
H = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \frac{3}{2} C(X_i^0 X_i)^{1/2} C^{0} + B \cdot \frac{1}{N} Z_i \tag{5.3}
$$
exists and is a positive definite matrix, i.e. are consistent, asymptotically efficient and equivalent. The normal distribution assumption is not necessary 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. Vice 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 modeling 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 modeling only the first element of \( a_i \)

\[
a_{1i} = z_i \cdot + b_1
\]  

(6.1)

where \( z_i \) is the \( 1 \times k \) vector of covariate, \( \cdot \) is the \( k \times 1 \) parameter of interest, and \( E(b_1) = 0; \text{var}(b_1) = I \). The standard growth curve setting (1.2) forces us to make additional, irrelevant assumptions on the rest of the elements of \( a_i \): Apparently, there are three options:

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

\[
a_{ij} = \xi_j; \quad \text{var}(a_{ij}) = 0; \quad j = 2, \ldots, m: \]

(6.2)

2. The remaining parameters are random with unknown means:

\[
a_{ij} = \xi_j + b_j; \quad j = 2, \ldots, m:
\]  

(6.3)
where \( b_i = (b_{i1}; \ldots; b_{im})^0 \) and \( \text{cov}(b_i) = -\cdot - \cdot \\

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

\[
a_i = (1 - z_i)^0 + b_i; \tag{6.4}
\]

where \( \omega = (\omega_0^0; \omega_0^0) \) is the \( mk \times 1 \) vector of parameters, \( \zeta = (\omega_0^0; \ldots; \omega_0^0) \) and \( \text{cov}(b) = -\cdot - \cdot \\

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., \( z_i^0 = \beta_1 \) but the rest of parameters \( \beta_2; \ldots; \beta_m \) 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 \( C = (1; \ldots; 0) \). We wish to determine bias induced by imposing an irrelevant structure on parameters \( \beta_2; \ldots; \beta_m \) 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

\[
\begin{align*}
\beta &= \frac{3X^0 \Omega M_i 1 \Omega_i 1 3X^0 \Omega M_i 1 \omega_0^0}{3X^0 (X_i^0 X_i)^{11} + \cdot \cdot \cdot} \\
\end{align*}
\]

where \( Q_i = \text{diag}(z_i; T_i) \); and \( T_i = I_{mi} - 1 \) for models (6.1) and (6.2), and \( T_i = (I_{mi} - z_i) \) for model (6.3), and \( \zeta \) is the vector of nuisance parameters. It is easy to see that GLSEUC is

\[
\begin{align*}
\beta &= \frac{\tilde{A}X^0 z_i^0 1 \tilde{A}X^0 z_i^0 1 \tilde{A}X^0 z_i^0 1}{3/4 (X_i ^0 X_i)^{11} + \cdot \cdot \cdot} \\
\end{align*}
\]

and is unbiased.

In the orthogonal case, when \( X_i ^0 X_i \) and \( \cdot \cdot \cdot \) 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 \( \Theta \) as \((\Theta_1; \Theta_2)\) where \( \Theta_2 = (\Theta_2; \ldots; \Theta_m) \); and the other quantities are partitioned correspondingly. Then from (6.5) we find the expectation

\[
(6.7) \quad \mathbb{E}(\Theta_2) = \mathbb{E}(\Theta_2) + \mathbb{E}(\Theta_1) - \mathbb{E}(\Theta_1) = \mathbb{E}(\Theta_2).
\]

Thus, the estimate of \( \Theta_2 \) based on the standard model (1.2) is unbiased if either \( \mathbb{E}(a_{i2}) = T_i \Theta \) for some \( \Theta \); or if \( F \mathbb{E}(M_{i2}^{1/2}) = U_{i1} U_{i2} E_i^{1/2} \mathbb{E}(M_{i2})^{1/2} \) for \( i = 1; \ldots; N \). Notice, that the latter case occurs when \( z_i \mathbb{E}(M_{i2})^{1/2} \) and \( T_i \mathbb{E}(M_{i2})^{1/2} \) are collinear. In particular, for the random coefficients model when \( z_i = 1 \) and \( T_i = 1 \); the magnitude of the bias is related to how unbalance the data is: Therefore in general, specifying structure for the \( \Theta_2 \) leads to a bias in the estimate.

We can even compute the maximum bias in terms of the deviation of \( \Theta_2 \) from \( T_i \Theta \). Applying the matrix inequality \( X^i Y (Y^i X)^i \geq X^i (Y^i X)^i \) (Graybill 1983) to (6.7) we obtain

\[
k \mathbb{E}(b_i) \leq k^2 \cdot \mathbb{E}(\Theta_2)^2 \mathbb{E}(\Theta_2)^X \mathbb{E}(\Theta_2)^T \mathbb{E}(R_i R_i)
\]

where \( C_i = F \mathbb{E}(M_{i2}^{1/2}) \mathbb{E}(M_{i2})^{1/2} \).

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, $y_{ij} = \theta_1 x_{ij} + \theta_2 + \eta_{ij}; j = 1; \cdots; n_i; i = 1; \cdots; N$ with second stage model $\theta_1 = \circ$; 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 $\alpha_2 = \theta_2 + b_2$; 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

$$X_i ^{X} = \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_m \end{pmatrix}^\top \begin{pmatrix} X_{ij} \end{pmatrix}; \quad M_i = \begin{pmatrix} \frac{3}{4} (X_{ij} X_{ij})_{11}^{\circ} - 11 & \frac{3}{4} (X_{ij} X_{ij})_{12}^{\circ} - 12 & \cdots & \frac{3}{4} (X_{ij} X_{ij})_{mn}^{\circ} - mn \end{pmatrix}$$

where $- 12 = - 22 = 0$ for model (6.2), and $(X_{ij} X_{ij})_{11}$; for instance, denotes the $(1; 1)$th element of the inverse matrix. We obtain

$$\begin{pmatrix} m_{112} & m_{1112} \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} \end{pmatrix}^\top = \begin{pmatrix} 1 \end{pmatrix}, \quad i = 1; \cdots; n_i$$

where $m_{ij}$ 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 $\xi = l_{11} l_{22} - l_{12}^2$; the bias from (6.7) can be written as

$$E(\xi) = \frac{1}{\xi} \begin{pmatrix} m_{112} & m_{1112} \end{pmatrix}(\theta_2 + \xi) \begin{pmatrix} m_{11} & m_{12} \end{pmatrix}.$$  

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

$$\max_{i \in [\hat{a}_1; \hat{a}_2]} \left\{ E(\xi) \right\} = \frac{\sqrt{\lambda}}{\xi} \begin{pmatrix} M_{112} & M_{1112} \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \end{pmatrix}^{\frac{1}{2}}.$$  

Simulations, however, show that when classical models are applied to general growth curve with unknown variance parameters, the bias might be negligible (Stukel and Demidenko 1997). This can be explained in the following way, a simple case with two parameters is considered ($m = 2$): Then, letting $- 12 = \frac{1}{2} \xi$ and $- 12 = \frac{1}{2}$ we obtain

$$M_i \begin{pmatrix} l_{11} & l_{12} \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix}, \quad \begin{pmatrix} h_{22i} & h_{22i} \end{pmatrix} = \begin{pmatrix} h_{12i} & h_{12i} \end{pmatrix}, \quad (h_{12i} + \frac{1}{2})$$

where $h_{kl} = \frac{3}{4} (X_{ij} X_{ij})_{kl}^{\circ}$: Letting $\frac{1}{2} \xi = \frac{1}{2}$ we obtain

$$M_i \begin{pmatrix} l_{11} & l_{12} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} (X_{ij} X_{ij})_{11}^{\circ} + \frac{1}{2} \xi \end{pmatrix}, \quad \begin{pmatrix} h_{22i} & h_{22i} \end{pmatrix} = \begin{pmatrix} h_{12i} & h_{12i} \end{pmatrix}, \quad (h_{12i} + \frac{1}{2})$$

and we come to the estimate (6.6). Thus, a peculiar compensation effect takes place: when in the growth curve model in the standard setting (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 as 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 $f_2$ leaving them unrestricted. 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 $(P_i M_i^{-1})^{-1}$: Denoting $\xi_i = M_i^{-1} M_i^{-1} M_i^{-1}$ as the determinant of the $2 \times 2$ matrix $M_i$, the variance of the GLS-estimator for $\beta$ is written as

$$\text{var}(b) = \xi_i^{-1} = hX M_{i12}^T (X M_{i12} \xi_i^{-1})^{-1} (X M_{i11} \xi_i^{-1})^{-1} = \xi_i^{-1} = M_{i12}^T M_{i11}^{-1} X^{-1} X M_{i11}^{-1} X^{-1} X M_{i12}^{-1} (X M_{i11} \xi_i^{-1})^{-1}.$$

We show that

$$0 \leq X M_{i12}^T M_{i11}^{-1} X^{-1} X M_{i11}^{-1} X^{-1} X M_{i12}^{-1} (X M_{i11} \xi_i^{-1})^{-1} = \xi_i^{-1} \leq X M_{i12}^T M_{i11}^{-1} X^{-1} X M_{i11}^{-1} X^{-1} X M_{i12}^{-1} (X M_{i11} \xi_i^{-1})^{-1}.$$

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

$$X M_{i12}^T M_{i11}^{-1} X^{-1} X M_{i11}^{-1} X^{-1} X M_{i12}^{-1} (X M_{i11} \xi_i^{-1})^{-1} = X M_{i12}^T M_{i11}^{-1} X^{-1} X M_{i11}^{-1} X^{-1} X M_{i12}^{-1} (X M_{i11} \xi_i^{-1})^{-1}.$$

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 $\varphi$ we have $M_{i12} = \varphi M_{i12}$ for all $i = 1, \ldots, N$. For instance, it is true for balanced data. In contrast, maximum loss of efficiency is when $P_i M_{i12}^T M_{i11}^{-1} X^{-1} X M_{i11}^{-1} X^{-1} X M_{i12}^{-1} (X M_{i11} \xi_i^{-1})^{-1} = 0$.

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-Johansen, 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 four 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 cm³. 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(Vᵢⱼ + 1) using a simple linear trend model

$$\log(Vᵢⱼ + 1) = aᵢ tᵢⱼ + aᵢ 2 + 2ᵢⱼ; \quad j = 1; \ldots; nᵢ; \quad i = 1; \ldots; N = 22 \tag{7.1}$$

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

$$a₁ = -₁ + -₂sexᵢ + b₁ \tag{7.2}$$

where sex = 1 for males, 0 for females. For the general model, the intercept term, a₂, which can be associated with the individual immunity, is left unspecified. By contrast, in the classical growth curve model requires specification of the intercept term: (i) a₂ is a fixed unknown parameter, a₂ = -₃; (ii) a₂ is random with unknown mean, a₂ = -₃ + b₂; (iii) a₂ is modelled via sex, a₂ = -₃ + -₄sexᵢ + b₂.

<table>
<thead>
<tr>
<th>Table 1. Results of estimation based on four models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Two-step</td>
</tr>
<tr>
<td>a₂ are not specified</td>
</tr>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>a₁</td>
</tr>
<tr>
<td>(SE)</td>
</tr>
<tr>
<td>-₂</td>
</tr>
<tr>
<td>(SE)</td>
</tr>
<tr>
<td>SE(b₁)</td>
</tr>
</tbody>
</table>

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 a₁ on sex. The results of estimation based on four models are presented in Table 1. The bias is calculated based on the assumption that a₂ = a₂: 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 effect 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 confidence. 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 misspecification 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 efficiency. Summarizing, the general growth curve approach suggested in this paper makes the model more robust to misspecification. 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 efficient estimator.
of parameters of interest.

9. Appendix

9.1. Derivation of moment estimator for B

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

$$\begin{align*}
X & \ u_i i \ X \ u_i (X u_i^0 Z_i) H Z_i X \ Z_i H (X Z_i^0 u_i) X \ Z_i H (X Z_i^0 u_i) (X u_i^0 Z_i) H Z_i^0. \\
& \ i \ j \ i \ j \ i \ j \ i \ j
\end{align*}$$

(9.1)

where $H = (P Z_i^0 Z_i)^{\dagger}$: Our aim is to find the expectation of (9.1). Since $u_i$ and $u_j$ are independent for $i \neq j$; we obtain

$$\begin{align*}
E P_i u_i^0 = \ N B + \frac{3}{2} P_i; \\
E (P_i u_i (P_i u_i^0 Z_i) H Z_i^0) = E (P_i u_i^0 Z_i H Z_i^0) = \frac{3}{2} P_i Q_i + B P_i Q_i; \\
E (P_i Z_i H (P_i Z_i^0 u_i) u_i) = \frac{3}{2} Q_i P_i + Q_i B Q_i.
\end{align*}$$

Combining these results we come to

$$\begin{align*}
E P_i (C a_i^o \ Z_i^0) (C a_i^o \ Z_i^0) P_i = N B + \frac{3}{2} P_i Q_i + B P_i Q_i. \\
& \ i \ j \ i \ j
\end{align*}$$

(9.2)

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

$$B_i \ B \frac{1}{N} X \ Q_i i \frac{1}{N} X \ Q_i B + \frac{1}{N} X \ Q_i B Q_i = L$$

(9.3)

where $\frac{3}{2}$ is substituted by the pooled variance (3.2), and

$$L = \frac{1}{N} \sum_{i=1}^{N} f(C a_i^o \ Z_i^0) (C a_i^o \ Z_i^0) Q_i P_i Q_i + Q_i P_i (Q_i P_i + Q_i B Q_i).$$

(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 $vec(ABF) = (F^0 - A) vec(B)$ we obtain

$$vec(B_i \frac{1}{N} X \ Q_i) = \frac{1}{N} D^+ vec(I_k B \frac{1}{N} X \ Q_i) = \frac{1}{N} D^+ (Q_i - I_k ) D vec(B);$$

$$vec(\frac{1}{N} X \ Q_i B) = \frac{1}{N} D^+ vec(\frac{1}{N} X \ Q_i B I_k) = \frac{1}{N} D^+ (I_k - \frac{1}{N} X \ Q_i) D vec(B);$$

$$vec(\frac{1}{N} X \ Q_i B Q_i) = \frac{1}{N} D^+ vec(Q_i B Q_i) = \frac{1}{N} D^+ (Q_i - Q_i) D vec(B).$$
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 $r \leq r$ matrix. Without loss of generality we can assume it is normalized, i.e., tr$(U^0U) = 1$: Since matrices $D$ and $D^+$ are of full rank to prove that matrix $J$ is nonsingular it suffices to show that

$$\text{vec}(U) \quad I_{r^2} \quad \frac{1}{N} \sum_{i=1}^{N} \# \quad Q_{i - 1} = I_{k - 1} + I_{k - 1}^Q = Q_i \quad \text{vec}(U) > 0: \quad (9.5)$$

Using the formula vec$(U)(A - B)\text{vec}(U) = \text{tr}(UAU^QB^Q)$ and applying the inequality tr$(AB) \cdot \text{tr}(A)\text{tr}(B)$ for any nonnegative definite matrices $A$ and $B$ one obtains:

$$\frac{1}{N}(\text{vec}(U)(P_i Q_i - 1)\text{vec}(U)) = \frac{1}{N}(\text{vec}(U)(P_i Q_i - 1)\text{vec}(U)) = \frac{1}{N} P_i \text{tr}(UQ_i U^Q) = \frac{1}{N} P_i \text{tr}(Q_i Z_i Z_i^Q) = \frac{1}{N} P_i \text{tr}(Z_i H Z_i^Q) = \frac{1}{N} P_i \text{tr}(H^T Z_i Z_i^Q) = \frac{1}{N} P_i.$$ 

Since vec$(U)(Q_i - 1)\text{vec}(U) \geq 0$; 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 $r \leq r$ random matrices, then \(p\lim_{N \downarrow 1} U_N = U_n\) if and only if for any nonrandom $r \leq r$ vector of unit length $t$ we have \(p\lim_{N \downarrow 1} t^T U_N t = t^T U_n t$.

2. Let $U_1; \ldots ; U_N$ be a sequence of $r \leq r$ random symmetric nonnegative definite matrices and \(\lim_{N \downarrow 1} E(\text{tr}(U_N)) = 0\); then \(\lim_{N \downarrow 1} (U_N) = 0$.

3. Let $u$ be a $r \leq r$ random vector with $E(u) = 0$ and $\text{cov}(u) = C$, $E(u^T) \cdot P$; $A$ is a $r \leq r$ symmetric matrix, then $E(u^T A u) = \text{tr}(A C)$ and $E(u^T A u)^2 \cdot j A_{ij} j^2 \cdot r_\ast \text{tr}(A^2)$.

4. Let $u_1; \ldots ; u_N$ be $r \leq r$ independently distributed vectors with $E(u_i) = 0$; and $\text{cov}(u_i) = C_i$: Let $A_i$ and $B_i$ be any $r \leq r$ matrices, $i = 1; \ldots ; N$. Then

$$\lim_{N \downarrow 1} \frac{1}{N} \sum_{i=1}^{N} A_i u_i^Q \sum_{i=1}^{N} B_i u_i = 0 \quad \lim_{N \downarrow 1} \frac{1}{N^2} \text{tr}(A_i^Q A_i C_i) \text{tr}(B_i^Q B_i C_i) = 0.$$ 

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:

$$E(u^T A u)^2 = \sum_{i,j,k,l} E(u_{ij} u_{ij}) (A_{kl} u_{kl} u_{kl}) \cdot \sum_{i,j,k,l} E(u_{ij} u_{ij}) A_{ij} j A_{ij} j A_{kl} j E(u_{ij} u_{ij}) E(u_{kl} u_{kl})$$

17
\[ \cdot \sum_{i,j,k,l} X_j A_{ij} j A_{kl} j \cdot \sum_{i,j,k,l} X_j A_{ij} j^2 = X \cdot \sum_{i,j} X_j A_{ij} j^2: \]

The last inequality in 3 follows from \((P_{ij} j A_{ij} j)^2 \cdot r(P_{ij} A_{ij}^2)^2 = r \text{etr}(A^2)\): To prove 4 we first apply the Cauchy inequality, and then use 3,

\[ \text{var} (X A_i u_i)^2 (X B_i u_i) \cdot E (X A_i u_i)^2 (X B_i u_i) \]

\[ \sum_{i,j} X_i A_{ij} Q_i \cdot \text{tr} \sum_{i,j} X_i A_{ij} Q_i \cdot \text{tr}(B_i B_{i}^j C_i) \]

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

\[ \text{var} (X A_i u_i)^2 (X B_i u_i) \cdot E (X A_i u_i)^2 (X B_i u_i) \]

\[ \sum_{i,j} X_i A_{ij} Q_i \cdot \text{tr} \sum_{i,j} X_i A_{ij} Q_i \cdot \text{tr}(B_i B_{i}^j C_i) \]

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

\[ \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N P_i Q_i = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N Q_i P_i = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N Q_i P_i Q_i = 0; \quad N \to \infty: \quad (9.6) \]

Since matrices \( P_i Q_i; Q_i P_i \) and \( Q_i P_i Q_i \) have nonnegative eigenvalues it suffices to show that the trace of matrices in (9.6) vanish. Using \( \text{tr}(AB) \cdot \text{tr}(A)\text{tr}(B) \) for any nonnegative definite matrices \( A \) and \( B \), and the assumption \( AB(b) \) we obtain

\[ \text{tr}(P_i Q_i) = \text{tr}(P_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \]

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

\[ \text{tr}(Q_i P_i Q_i) \cdot d_i \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \cdot \text{tr}(Q_i) \]

To prove the equivalence of \( \mathbf{B}_s \) to \( \mathbf{B} \) it is left to show that \( J = I \): Indeed,

\[ \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N (Q_i - I) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N (I - Q_i) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N (Q_i - Q_i) = 0; \quad (9.7) \]
which follows from the fact $\text{tr}(A - B) = \text{tr}(A) - \text{tr}(B)$:

Consistency of (4.4). As follows from 1 of the above Lemma it suffices to consider a quadratic form of (9.1) times $1^tN$. Thus, let $t$ be a fixed $1 \times 1$ vector, $k = 1$.

For the first term of (9.1), using $AB(b)$ and $AB(c)$ and 3 of the Lemma, we obtain when $N \to 1$

$$\text{var}\left[\frac{1}{N}t^tXu_iz_i\right] = \frac{1}{N^2} \text{var}(\frac{1}{X_i}X_i^t) + \frac{X^t}{N^2} \text{tr}(\frac{1}{X_i}X_i^t)$$

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

$$E @ \text{tr}\left[\frac{1}{N}XZ_i^tH(X_j^0u_j)(X_j^0u_j)^{i_1}H^0\right] = \frac{1}{N^2} \text{tr}(XZ_i^tH(X_j^0u_j)(X_j^0u_j)^{i_1}H^0)$$

For the second term of (9.1) after multiplying by $t^0$ and $t$ we obtain

$$t^0X_iu_i(t^0u_iZ_i^0)H^0Z_i^0 = (t^0u_iZ_i^0)H^0Z_i^0$$

Now we apply 4 of the Lemma noting that

$$\frac{1}{N^2} \text{tr}(Z_i^0t^0Z_i^0(3^0\text{C}(X_i^0X_i)^0 + B)) \text{ tr}(Z_i^0t^0Z_i^0(3^0\text{C}(X_i^0X_i)^0 + B))$$

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

Asymptotic covariance matrix for $\hat{\theta}$ and $\hat{\theta}_s$. It is easy to see that $u_i = \pm + C(X_i^0X_i)^0X_i^0 = \pm + T_i^2$; where $T_i = C(X_i^0X_i)^0X_i^0$ and $\hat{\theta}_s = \hat{\theta}_s^2$ where
\[ W_i = (I \cdot X_i \cdot (X_i^T X_i)^i \cdot 1)^P \cdot (n_i \cdot m) \] Therefore, asymptotically, \( \mathbb{E} \) is equivalent to
\[ \mathbb{E} \cdot (\pm + T_i^2) (\pm + T_i^2)^0 \cdot P_{i}^2 W_{i}^2 \]
and
\[ \text{cov}(\text{vec}(\mathbb{E})), \frac{1}{N^2} D^+ \mathbb{E} \cdot \text{cov} \cdot \text{vec}^i (\pm + T_i^2) (\pm + T_i^2)^0 \cdot P_{i}^2 W_{i}^2 \]
where
\[
\text{vec}^i (\pm + T_i^2) (\pm + T_i^2)^0 \cdot P_{i}^2 W_{i}^2 = (\pm - \pm) \cdot (T_i^2 - \pm + \pm - T_i^2) + M_i \]

and \( M_i = T_i - T_i \cdot \text{vec}(P_i) \cdot \text{vec}(W_i) \)

Now we will use the following fact: let \( u \) be a normal \( r \cdot 1 \) vector, then \( \text{cov}(u - u) = 2N(C_u - C_u) \) where \( N = (I + K) \approx 2 \) and \( K \) is the \( r^2 \cdot r^2 \) commutation matrix and \( \text{cov}(u) = C_u \) (e.g. Magnus 1988, p. 164):

We consider three terms in (9.8). For the first term \( \text{cov}(\pm - \pm) = 2(B - B)N \) The second term is uncorrelated with other two. In fact, due to the independence of \( \pm \) and \( T_i^2 \):
\[
E(T_i^2 - \pm + \pm - T_i^2) = E(T_i^2 - \pm) + E(\pm - T_i^2) = (ET_i^2) - E(\pm) + E(\pm) - E(T_i^2) = 0;
\]
and therefore,
\[
\text{cov}(\pm - \pm; T_i^2 - \pm + \pm - T_i^2) = E(\pm^2 \cdot T_i^2 - \pm \pm^0) + E(\pm \pm^0 - \pm^2 \cdot T_i^2);
\]
However, again, due to the independence, it is easy to see that \( E(\pm^2 \cdot T_i^2 - \pm \pm^0) \) and \( E(\pm \pm^0 - \pm^2 \cdot T_i^2) \) vanish. Thus, for the third term of (9.8) we have \( \text{cov}(M_i (\pm^2 - \pm^2)) = 2 \frac{1}{N^2} M_i M_i^0 \)
Now we have
\[
M_i M_i^0 = (T_i - T_i \cdot \text{vec}(P_i) - \text{vec}(W_i))(T_i^0 - T_i^0 \cdot \text{vec}(W_i) - \text{vec}(P_i))
\]
\[
= T_i T_i^0 - T_i T_i^0 + \text{vec}(P_i) \cdot \text{vec}(W_i) \cdot \text{vec}(W_i) \cdot \text{vec}(P_i)
\]
\[
\text{vec}(W_i)^0 (T_i^0 - T_i) (T_i - T_i \cdot \text{vec}(W_i) \cdot \text{vec}(P_i));
\]

The third and the fourth terms vanish because, for instance, for the third term
\[
\text{vec}(W_i)^0 (T_i^0 - T_i) = ((T_i - T_i \cdot \text{vec}(W_i))^0)^0 = (\text{vec}(T_i \cdot W_i \cdot T_i))^0 = 0
\]
because $T_iW_i = 0$: Similarly we can show that the fourth term in (9.9) vanishes.  

Now, for the second term of (9.9) we have

$$\text{vec}(P_i)\text{vec}(W_i)\text{vec}(W_i)\text{vec}(P_i) = \text{vec}(P_i)\text{vec}(P_i)\text{tr}(W_i^2)$$

$$= \frac{n_i m}{(n_i m)^2} \text{vec}(P_i)\text{vec}(P_i)$$

and, noting that $T_iT_i^0 = C(X_i^0X_i)^{1/2}C^0 = P_i$, finally obtain

$$\text{cov}(\text{vech}(B)) = \frac{2}{N} D^+N (B - B) + \frac{3/4}{N} \sum_{i=1}^{N} \text{vec}(P_i)\text{vec}(P_i) D^+0.$$ 

At last we notice that $(n_i m)^2 (n_i m)^2 = O(N^{-2})$ 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 $(\beta_1; \cdots; \beta_N)$ and the parameter of interest is its 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 $t_1; \cdots; t_N$ not dependent on parameters that the conditional distribution $f(y_{ij} \mid t_i); i = 1; \cdots; N$ does not depend of nuisance parameters. In our case, the minimal sufficient statistics are $fC a_i; i = 1; \cdots; N$ and as follows from (??), the conditional distribution does not include $(\beta_1; \cdots; \beta_N)$; therefore 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 $\frac{3}{2}$ is used. Then, the estimating equation for $B$ is

$$\frac{\partial \ell_2}{\partial B} = \frac{1}{2} \sum_{i=1}^{N} (P_i + B)^{\frac{1}{2}} (P_i + B)^{\frac{1}{2}} (c_i \cdot Z_i^{-}) (c_i \cdot Z_i^{-})^0 (P_i + B)^{\frac{1}{2}} = 0;$$

(9.10)

Taking the expectation and recalling that $E (c_i \cdot Z_i^{-}) (c_i \cdot Z_i^{-})^0 = P_i + B$ we see that the expectation of (9.10) is zero. Therefore, $\hat{B}$; as the solution to (9.10), remains consistent when $\frac{3}{2}$ is substituted by a consistent estimator $\hat{\theta}$ (e.g., Bierens 1994).

Secondly, we apply the theory of pseudo-maximum likelihood (PML) approach. We notice that the expectation of the cross-product derivative of the likelihood function with respect to $\vec{}$ and the variance parameter is zero. Therefore, as follows from PML theory, the generalized least squares estimator (2.9) is efficient because $\hat{\theta}$ 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 $T_N (\mu) = N \sum_{i=1}^{N} F_i (\mu) \cdot \gamma_i$, where $\mu$ is a $q \times 1$ parameter vector, $F_i (\mu)$ is a $k \times r$ matrix such that $k \cdot F_i (\mu) \cdot \gamma = A$; $k \cdot (\mu) \cdot \gamma = A$ in some neighborhood of $\mu_0$ for $i = 1; \ldots; N$; and $\gamma_i$ is a random $r \times 1$ vector, $E (\gamma_i) = 0$; $\text{cov} (\gamma_i) = B$. Let $\hat{\mu}_N$ be a consistent statistic for $\mu_0$. Then $T_N (\hat{\mu}_N)$ has a normal asymptotic distribution $N (0; \lim N \sum_{i=1}^{N} F_i (\mu_0) \cdot B \cdot F_i^0 (\mu_0))$; assuming the last limit exists, when $N \to \infty$.

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

$$P \left( \frac{\hat{\mu}_N}{\hat{\gamma}_N} \right) = N \sum_{i=1}^{N} Z_i \cdot M \cdot i_{t \cdot 1} Z_i \cdot i_{t \cdot 1} (T_{N_1} + T_{N_2})$$

where

$$T_{N_1} = N \sum_{i=1}^{N} X_i \cdot M \cdot i_{t \cdot 1} C \cdot i_{t \cdot 1}; \quad T_{N_2} = N \sum_{i=1}^{N} X_i \cdot M \cdot i_{t \cdot 1} C \cdot (X_i \cdot X_i) \cdot i_{t \cdot 1} X_i \cdot q_i.$$ 

We apply Lemma 2 to show that vectors $T_{N_1}$ and $T_{N_2}$ are asymptotically normally distributed when $N \to \infty$. Let us start with $T_{N_1}$ denoting $T_{N_1} = Z_i \cdot M \cdot i_{t \cdot 1} C$. 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

$$\frac{\partial \theta}{\partial \gamma} = i_{t \cdot 1} Z_i \cdot M \cdot i_{t \cdot 1} C \cdot (X_i \cdot X_i) \cdot i_{t \cdot 1} X_i \cdot q_i; \quad \frac{\partial \theta}{\partial \gamma} = i_{t \cdot 1} Z_i \cdot M \cdot i_{t \cdot 1} K_{gh} M \cdot i_{t \cdot 1} C$$
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 covariance matrix \( C_2 \) where

\[
C_1 = \lim_{N \to \infty} N^{-1} X_i^t M_i^t C^{-1} M_i^t Z_i^t C^{-1} Z_i^t;
\]
\[
C_2 = \lim_{N \to \infty} N^{-1} X_i^t M_i^t (X_i^t X_i)^{-1} C^{-1} M_i^t Z_i^t C^{-1} Z_i^t.
\]

Since \( T_{N1} \) and \( T_{N2} \) are independent, the asymptotic covariance of \( \hat{b}_{i \cdot} \) is the sum of covariances for the first and the second term, the theorem is proved.

References


