Estimation in a growth study with irregular measurement times

K. Poortema

Faculty of Applied Mathematics, University of Twente,
P.O. Box 217, 7500 AE Enschede, The Netherlands

R. Y. J. Tamminga

Division of Pediatric Oncology, University Hospital of Groningen,
P.O. Box 30.001, 9700 RB Groningen, The Netherlands

W. Schaafsma

Department of Mathematics and Computing Science, University of Groningen,
P.O. Box 800, 9700 AV Groningen, The Netherlands

Between 1982 and 1988 a growth study was carried out at the Division of Pediatric Oncology of the University Hospital of Groningen. A special feature of the project was that sample sizes are small and that ages at entry may be very different. In addition the intended design was not fully complied with. This paper highlights some aspects of the statistical analysis which is based on (1) reference scores, (2) statistical procedures allowing for an irregular pattern of measurement times caused by missing data and shifted measurement times.

Key Words & Phrases: growth curves, expectations and variances with linear structure, iterative generalized least squares.

1 Introduction

About 1980 N. M. Drayer (Division of Pediatric Endocrinology, University Hospital of Groningen) and J. A. de Vries (Division of Pediatric Oncology, University Hospital of Groningen) initiated the growth study to be discussed in this paper. To which extent growth of children may be disturbed by cancer and treatment was the subject of interest. Two subprojects were distinguished. Subproject A concerned growth retardation during treatment and subproject B concerned catch-up growth after completion of the treatment. The idea was that systematic and careful collection of a number of measurements would reveal specific growth patterns. With respect to anthropometric measurements, W. J. Gerver and N. M. Drayer were carrying out a project aimed at the assessment of means, standard deviations, and percentiles of the population of healthy children in the region where the hospital is located, see GERVER (1988). The statisticians J. C. Akkerboom and W. Schaafsma participated in the discussions about the design of the growth study. After many discussions it was decided: (i) to restrict attention to the anthropometric measurements investigated in the Gerver-Drayer project,

(ii) to measure the children 9 times with exactly 3 months between the different sessions, (iii) to refrain from measuring healthy children (It was argued that cross-sectional aspects could be inferred from the Gerver-Drayer data while longitudinal aspects could be derived from cohort analyses by VAN 'T HOF (1977) based on the Nijmegen growth study, cf. PRAHL-ANDERSEN et al. (1979)). Patient entry started in March 1982 and lasted till April 1986.

A special feature of the project is that sample sizes are small and that ages at entry may be very different. In order to account for differences in age and sex, the original scores are replaced by reference scores. If e.g. $h(t)$ is a height measurement of a boy (girl) at age t then, instead of $h(t)$, the score $x = (h(t) - \mu(t))/\sigma(t)$ is studied, where $\mu(t)$ and $\sigma(t)$ are (estimates of) respectively the mean height and the standard deviation of height, of boys (girls) at age t. The $\mu(t)$ and $\sigma(t)$ are computed on the basis of the data of GERVER (1988).

Unfortunately, but not unexpectedly, the intended design was not fully complied with. The height measurements of the entire group of subproject B $(n = 21)$ will be studied in Section 3. For this group 12% of the measurements are missing and many of the remaining measurements are carried out at shifted time points, other groups are similar in this respect. For this reason the statistical procedures to be constructed allow for measurements made at $p(i)$ time points $t_{i1}, ..., t_{i,p(i)}$ for child i, where the sets $\{t_{i1}, ..., t_{i,p(i)}\}$ may be fairly different (see the conditions of Section 3). The time points t_{ij} are regarded as given numbers. One should be certain, however, that the sets $\{t_{i1}, ..., t_{ip(i)}\}$ are not related to the vectors B_i. The unit of time is taken to be one year and time is rescaled such that for each child "$t = 0$" lies at the centre of the intended two years' period of observation.

Consider some anthropometric variable and some (sub)group of patients. Let n denote the number of children and let x_{ij} represent the j-th reference score of child i $(i = 1, ..., n; j = 1, ..., p(i))$. We consider the x_{ij} the outcomes of random variables

$$X_{ij} = B_{i1} + B_{i2}t_{ij} + ... + B_{iq}t_{ij}^{q-1} + E_{ij} \tag{1}$$

for some integer q, where the vectors $B_i = (B_{i1}, ..., B_{iq})^T$ are independent and identically distributed, and where the E_{ij} are independent of the B_i, independent of each other, and identically distributed with expectation zero.

2   Estimation

Let $\beta = (\beta_1, ..., \beta_q)^T$ and $\Lambda = (\lambda_{jk})$ denote respectively the expectation and the co-variance matrix of the vectors B_i. Define σ^2 as the variance of the disturbances E_{ij}. Assuming normality, the distribution of $X_i = (X_{i1}, ..., X_{i,p(i)})^T$ given the observation pattern $t_{i1}, ..., t_{i,p(i)}$ is determined completely. It may be expressed concisely as follows:

$$X_i \sim N(A_i\beta, A_i\Lambda A_i^T + \sigma^2 I_{p(i)}) \tag{2}$$

where A_i is the $p(i) \times q$ matrix with (j, k)-element t_{ij}^{k-1}. The covariance structure $A_i\Lambda A_i^T + \sigma^2 I_{p(i)}$ stems from RAO (1965). The case of identical distributions $(A_i = A$ for

all i: the original design) admits relatively simple estimation theory, see e.g. POORTEMA (1989).

The differences between the design matrices A_i were due to missing data and shifted measuring times. If there were no shifted measuring times the matrices A_i could be constructed by deleting rows from the original design matrix A. Then an approach like that of GILL (1988) and KLEINBAUM (1973), who deal however with other models, were feasible. Due to the shifted measuring times our approach resembles somewhat the approaches of VONESH and CARTER (1987) and HUI (1984). The estimator of β (the parameters of main interest) is a weighted mean of the individual least squares predictors/estimators of the B_i. For the weights estimators of Λ and σ^2 are needed. We apply an iteration procedure which might be called iterative generalized least squares. GOLDSTEIN (1986) applies the same method to multilevel data. ANDERSON (1973) is the first author to give some version of the method. The method requires no numerically intensive calculations and in case of convergence a solution of the maximum likelihood equations is obtained.

ANDERSON (1973) proposed an iteration procedure for estimation with respect to independent stochastic vectors $X_i \sim N(\sum_s \beta_s a_s, \sum_t \gamma_t G_t)$, where the β_s and the γ_t are unknown parameters and the a_s (G_t) are known vectors (matrices). This procedure can be generalized to independent stochastic vectors

$$X_i \sim N(\sum_s \beta_s a_{is}, \sum_t \gamma_t G_{it}) \tag{3}$$

where the known vectors a_{is} and matrices G_{it} may depend on i. Model (2) can be written in this form. Because of

$$\Lambda = \sum_j \lambda_{jj} e_j e_j^T + \sum_{j<k} \lambda_{jk}(e_j e_k^T + e_k e_j^T)$$

where e_j is the j-th unit vector of $\mathbb{R}^q$ we can write

$$A_i \Lambda A_i^T + \sigma^2 I_{p(i)} = \sum_t \gamma_t G_{it}$$

with $\gamma = (\gamma_1, ..., \gamma_m)^T = (\lambda_{11}, \lambda_{12}, \lambda_{22}, ..., \lambda_{qq}, \sigma^2)^T$ and

$$
\begin{aligned}
G_{it} &= a_{ij} a_{ij}^T && \text{if } \gamma_t = \lambda_{jj} \\
G_{it} &= a_{ij} a_{ik}^T + a_{ik} a_{ij}^T && \text{if } \gamma_t = \lambda_{jk} \ (j < k) \\
G_{it} &= I_{p(i)} && \text{if } t = m = q(q+1)/2 + 1
\end{aligned}
$$

where a_{is} is the s-th column of A_i. We thus get (3) as model equivalent to (2). The generalized iteration procedure is based on the following equations

$$\sum_k \sum_i a_{ij}^T \Gamma_i^{-1} a_{ik} \beta_k = \sum_i a_{ij}^T \Gamma_i^{-1} x_i \tag{4a}$$

$$
\begin{aligned}
&\sum_k \sum_i \operatorname{tr} \Gamma_i^{-1} G_{ij} \Gamma_i^{-1} G_{ik} \gamma_k / 2 = \\
&\sum_i \operatorname{tr} \Gamma_i^{-1} G_{ij} \Gamma_i^{-1} (x_i - \sum_s \beta_s a_{is})(x_i - \sum_s \beta_s a_{is})^T / 2
\end{aligned} \tag{4b}
$$

with $\Gamma_i = \sum_t \gamma_t G_{it}$. These equations can be derived by equating the derivatives of the log-likelihood with zero. Note that (4b) admits a simplification as $\sum_k \Gamma_i^{-1} G_{ik} \gamma_k = I_{p(i)}$, it is however convenient to base an iteration procedure on (4b). The iteration procedure is as follows. Each iteration step consists of solving the equations (4) for $\beta = (\beta_1, ..., \beta_l)^T$

and $\gamma = (\gamma_1, ..., \gamma_m)^T$ with Γ_i replaced by $\check{\Gamma}_i = \sum_t \check{\gamma}_t G_{it}$ where $\check{\gamma} = (\check{\gamma}_1, ..., \check{\gamma}_m)^T$ is the estimate of the previous iteration step. We start the iteration procedure with an initial guess $\gamma^{(0)}$. If $p(i) = p$ and $\Gamma_i = \Gamma$ for all i then the original procedure of ANDERSON (1973) is retained, this iteration procedure may be started with any positive definite guess Γ_0 for Γ.

The asymptotic properties of the modified iteration procedure, which have been established by POORTEMA (1989), resemble the properties given by ANDERSON (1973). For establishing these properties some or all of the following conditions are assumed:

(i) The numbers $p(i)$ and the elements of the a_i and the G_{it} are all smaller than some constant $\bar{c}$

(ii) The eigenvalues of the matrices $\Gamma_i = \sum_t \gamma_t G_{it}$ and $\sum_t \gamma_t^{(0)} G_{it}$ are all larger than some constant $\underline{c} > 0$.

(iii) The matrix $\sum_i A_i^T A_i$, with $A_i = [a_{i1} ... a_{i,q}]$ is invertible and $\sum_i A_i^T A_i / n \geq d_1 I_q$ for sufficiently large n for some constant d_1.

(iv) The matrix $\sum_i D_i^T D_i$, with $D_i = [\text{vec}[G_{i1}] ... \text{vec}[G_{im}]]$, is invertible and $\sum_i D_i^T D_i / n \geq d_2 I_m$ for sufficiently large n for some constant d_2.

(v) All fourth moments of the X_i are uniformly bounded.

(vi) If F_i is the distribution function of $\|X_i\|$ then there exists a distribution function F such that $\int_{|t|>c} t^2 dF_i(t) \leq \int_{|t|>c} t^2 dF(t)$ holds for arbitrary i and $c > 0$.

(vii) The distributions of the X_i are normal.

Here we used the vec-operation: for a matrix Z with rows z_i $(i = 1, ..., k)$ we have $\text{vec}[Z] = (z_1, z_2, ..., z_k)^T$. Note that conditions (i) and (vii) together imply conditions (v) and (vi).

The estimators $\hat{\beta}$ and $\hat{\gamma}$ are $n^{1/2}$-consistent $(n^{1/2}(\hat{\beta} - \beta)$ and $n^{1/2}(\hat{\gamma} - \gamma)$ are bounded in probability) after one iteration step, if the first five conditions are satisfied. If condition (vi) is added then it can be established that $\hat{\beta}$ is $AN(\beta, (\sum_i A_i \Gamma_i^{-1} A_i)^{-1})$ if at least two iteration steps are carried out. If finally condition (vii) is added then the asymptotic normality of $(\hat{\beta}^T, \hat{\gamma}^T)^T$ can be derived and the estimators are asymptotically efficient, provided (again) that at least two iteration steps are carried out. Note that one iteration step suffices to obtain consistent estimators. Note moreover that using these consistent estimators asymptotically efficient estimators are constructed, this is in line with other general theory, see e.g. DZHAPARIDZE (1983).

3 Results

With respect to starting the iteration procedure there were some difficulties. The produced estimates of Λ need not be positive definite matrices. Since the variances var (B_{i3}) are very small sometimes a negative estimate $\check{\lambda}_{33}$ is obtained. We firstly tried the initial guess '$\check{\Lambda} = 0$ and $\check{\sigma}^2 = 1$'. We secondly tried the initial guess '$\check{\lambda}_{jj} = 1$, $\check{\lambda}_{jk} = 0$, and $\check{\sigma}^2 = 1$'. Assuming that the initial guess should resemble at least a little the covariance structure studied we finally used estimates with respect to the spanwidth measurements of the entire group of Subproject B as initial guess for starting up the iteration procedure for whatever group and variable. With this initial guess (which is indeed highly arbitrary) no negative estimates for variances were found.

Beforehand it was expected that the reference scores would follow straight lines $(q = 2)$. After inspection of the data it was clear that $q = 3$ (parabolic growth curves) provides a better fit. Whether $q = 3$ fits the data satisfactorily should be tested. The model specified by $q = 4$ (cubic growth curves) provides an alternative. In view of the rather small sample sizes we did not consider more complicated models. The model $q = 3$ has been tested with the model $q = 4$ as alternative.

We present results with respect to the height measurements of the entire group $(n = 21)$ of subproject B. The analyses of other variables and other groups can be carried out analogously. After testing $q = 3$, see the appendix, we postulated $q = 3$, the estimates of the (first) three iteration steps of the estimation procedure are given by Table 1 and 2.

Table 1.   Estimates $\hat{\beta}_k$ for $q = 3$

iteration	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	−.6425	.1544	−.0333
2	−.6380	.1535	−.0417
3	−.6383	.1535	−.0412
stand.error	.2374	.0458	.0255

The comparison of growth velocities is a natural point of interest. We therefore focus on the estimation of the expectation β_2 of the coefficients B_{i2}, the growth rates at $t = 0$. A rather simple estimator of β_2 proves to be as good as the corresponding estimator of the iteration procedure. Note that the least squares estimator $\hat{B}_{i2}$ of B_{i2} is $N(\beta_2, \lambda_{22} + \sigma^2 c_{i2})$-distributed where c_{i2} is the (2,2)-element of the matrix $(A_i^T A_i)^{-1}$. If the individuals do not differ much with respect to the constants c_{i2}, an appropriate estimator of β_2 is $\sum_i \hat{B}_{i2}/n$; as standard error may serve $s/n^{1/2}$, where s is the sample standard deviation of the $\hat{B}_{i2}$. We get $.1507 \pm .0481$; the standard error is not much larger than the corresponding standard error in Table 1. The outcome of $\sum_i \hat{B}_{i2}/n$ is close to the asymptotically best estimate of the iteration procedure (the estimate $\hat{\beta}$ is as a matter of fact a weighted mean); this is in agreement with the results with respect to the other variables and groups. As the estimation of β_2 is of primary interest, the mean $\sum_i \hat{B}_{i2}/n$ is preferred by TAMMINGA (1990) who gives a thorough investigation of all data.

Table 2.   Estimates $\hat{\lambda}_{jk}$ and $\hat{\sigma}^2$ for $q = 3$

iteration	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{22}$	$\hat{\lambda}_{13}$	$\hat{\lambda}_{23}$	$\hat{\lambda}_{33}$	$\hat{\sigma}^2$
1	1.1742	−.0094	.0421	−.0082	−.0132	.0072	.0068
2	1.1814	−.0113	.0419	−.0160	−.0133	.0075	.0069
3	1.1810	−.0113	.0421	−.0157	−.0133	.0075	.0069

Using the estimates of the variances $\hat{\lambda}_{jj}$ and the covariances $\hat{\lambda}_{jk}$, correlations can be estimated as well by means of the iteration procedure. An interesting parameter is the correlation between B_{i2}, the speed of growth at $t = 0$, and B_{i1}, the level at $t = 0$; we found $-.0509 \pm .2231$. Another interesting correlation is the correlation between B_{i2}

and $B_{i1} - B_{i2} + B_{i3}$, the level at $t = -1$ (the start of the study); here we found $-.2909 \pm .2048$. The standard errors are computed using the delta method. No relation between B_{i2} and the levels B_{i1} and $B_{i1} - B_{i2} + B_{i3}$ can be established.

4 Final remark

We have been working with the actual times $t_{i1}, ..., t_{i,p(i)}$ for each individual i. Note that beforehand no interpolations or other data modifications have been applied to cope with the deviations from the original design. The general iteration procedure for estimation turns out to work well. The parameter of main interest (β_2), however, can be estimated about equally well by using a much simpler estimator: our general approach seems to be too sophisticated in this respect.

APPENDIX

For discussing the model $q = 3$ our starting point was the model $q = 4$:

$$X_{ij} = B_{i1} + B_{i2} t_{ij} + B_{i3} t_{ij}^2 + B_{i4} t_{ij}^3 + E_{ij}$$

We first tested the null hypothesis $H_0 : \lambda_{44} = 0$ against the alternative hypothesis $H_1 : \lambda_{44} \neq 0$. The test statistic has been based on the least squares estimators/predictors $\hat{B}_{i4}$ of the B_{i4}; these are independent and $\hat{B}_{i4}$ is $N(\beta_4, \lambda_{44} + \sigma^2 c_{i4})$-distributed where c_{i4} is the (4,4)-element of the matrix $(A_i^T A_i)^{-1}$. The test statistic used is

$$T = S^2 / \check{\sigma}^2$$

with

$$S^2 = \sum_i c_{i4}^{-1} (\hat{B}_{i4} - \check{\beta}_4)^2 / (n - 1),$$

$$\check{\beta}_4 = \sum_i c_{i4}^{-1} \hat{B}_{i4} / \sum_i c_{i4}^{-1},$$

$$\check{\sigma}^2 = \sum_{ij} (X_{ij} - \hat{B}_{i1} - \hat{B}_{i2} t_{ij} - \hat{B}_{i3} t_{ij}^2 - \hat{B}_{i4} t_{ij}^3)^2 / f,$$

where $f = \sum_i (p(i) - 4)$. Under H_0 $(n - 1)S^2$ has the $\sigma^2 \chi_{n-1}^2$ distribution. The statistics S^2 and $\check{\sigma}^2$ are independent and it follows that $T = S^2 / \check{\sigma}^2$ has the $F_{n-1,f}$ distribution under H_0, since $f \check{\sigma}^2$ has the $\sigma^2 \chi_f^2$ distribution. Note that the test statistic reduces to a ratio of variances if the constants c_{i4} don't differ. The hypothesis H_0 need not be rejected at level .05; the outcome of F equals 1.07 while $n - 1 = 20$ and $f = 83$.

We hence postulate $\lambda_{44} = 0$, we may set equivalently

$$B_{14} \equiv ... \equiv B_{n4} \equiv \beta_4$$

and test, in addition, the null hypothesis $H_0 : \beta_4 = 0$ against the alternative hypothesis $H_1 : \beta_4 \neq 0$. The iteration procedure for estimation was used again for re-estimating the remaining parameters; we found $\hat{\beta}_4 = .0212$ with standard error .0282. We hence need not reject the second null hypothesis and postulate the model with $q = 3$.

References

ANDERSON, T. W. (1973), Asymptotically efficient estimation of covariance matrices with linear structure, *The Annals of Statistics 1*, 135–141.

DZHAPARIDZE, K. O. (1983), On iterative procedures of asymptotic inference, *Statistica Neerlandica 37*, 181–189.

GERVER, W. J. M. (1988), Measurement of the body proportions in children, The Oosterwolde study, Thesis, University of Groningen.

GILL, R. D. (1988), Regression analysis for incomplete mixed cross-section and time-series data by a modified EM algorithm, *Sankhyā 50*, 95–102.

GOLDSTEIN, H. (1986), Multilevel mixed linear model analysis using iterative generalized least squares, *Biometrika 73*, 43–56.

HOF, M. A. VAN 'T (1977), Some statistical and methological aspects in the study of growth and development: with a special emphasis on mixed longitudinal designs, Thesis, Nijmegen Catholic University.

HUI, S. L. (1984), Curve fitting for repeated measurements made at irregular time-points, *Biometrics 40*, 691–697.

KLEINBAUM, D. G. (1973), A generalization of the growth curve model which allows missing data, *Journal of Multivariate Analysis 3*, 117–124.

POORTEMA, K. (1989), On the statistical analysis of growth, Thesis, University of Groningen.

PRAHL-ANDERSEN, B., C. J. KOWALSKI and P. HEYDENDAEL (1979), *A mixed-longitudinal interdisciplinary study of growth and development*, Academic Press, New York.

RAO, C. R. (1965), The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrika 52*, 447–458.

TAMMINGA, R. Y. J. (1990), Growth of children with cancer, Thesis, University of Groningen.

VONESH, E. F. and R. L. CARTER (1987), Efficient inference for random-coefficient growth curve models with unbalanced data, *Biometrics 43*, 617–628.