

## PC PROGRAM EXTENDING THE TWO-STAGE POLYNOMIAL GROWTH CURVE MODEL TO ALLOW MISSING DATA

AMY M. FUREY<sup>a</sup>, THOMAS R. TEN HAVE<sup>a</sup>, CHARLES J. KOWALSKI<sup>b</sup>, EMET D. SCHNEIDERMAN<sup>c</sup> and STEPHEN M. WILLIS<sup>c</sup>

<sup>a</sup>*Department of Biostatistics and* <sup>b</sup>*Department of Oral Biology, The University of Michigan, Ann Arbor, MI 48109 and* <sup>c</sup>*Department of Oral and Maxillofacial Surgery, Baylor College of Dentistry, 3302 Gaston Ave., Dallas, TX 75246 (USA)*

(Received February 17th, 1993)

(Accepted March 11th, 1993)

A stand-alone, menu-driven PC program, written in GAUSS386i, extending the analysis of one-sample longitudinal data sets satisfying the two-stage polynomial growth curve model (Ten Have et al., *Am J Hum Biol*, 3 (1991) 269–279) to allow missing data is described, illustrated and made available to interested readers. The method and the program are illustrated using data previously analyzed by the authors (Schneiderman and Kowalski, *Am J Phys Anthropol*, 67 (1985) 323–333) but with several randomly chosen data points discarded and treated as missing.

*Key words:* Longitudinal studies; Polynomial growth curves; Missing observations; PC program

### Introduction

In a recent paper [1] we provided a PC program, called 2STG, implementing Rao's [2] analysis of one-sample longitudinal data sets satisfying the two-stage polynomial growth curve model. This program can be used to determine the lowest degree polynomial adequate to fit the individual and average growth curves, to test the goodness-of-fit of the two-stage model and to obtain confidence intervals for the polynomial regression coefficients and confidence bands for the average growth curve. It has proven to be a useful data-analytic tool for situations in which no missing data are present, i.e. when each of the individuals in the sample is measured at each time point, but this is a requirement that is not always realised in practice. The purpose of the present paper is to outline the extension of this technique to accommodate individual-specific time design matrices due to Carter and Yang [3] and to provide a PC program for performing the associated computations. The technique and the program are illustrated using data first analyzed by Schneiderman and Kowalski [4], but with several randomly chosen data points discarded and treated as missing.

---

*Correspondence to:* Emet D. Schneiderman, Department of Oral and Maxillofacial Surgery, Baylor College of Dentistry, 3302 Gaston Ave., Dallas, TX 75246, USA.

### The model

As previously described [1], the two-stage polynomial growth curve model (sometimes called a random coefficients regression model) assumes that the  $T$  observations on the  $i$ th individual ( $i = 1, 2, \dots, N$ ) have the structure

$$\mathbf{x}_i \sim MVN(\mathbf{W}\tau, \mathbf{W}\Lambda\mathbf{W}' + \sigma^2\mathbf{I}) \quad (1)$$

where  $\mathbf{W}$ , the time design matrix [5], consists either of powers of the times of measurement  $t_1, t_2, \dots, t_T$ , or the values of orthogonal (where  $\mathbf{W}'\mathbf{W}$  is diagonal) or orthonormal ( $\mathbf{W}'\mathbf{W} = \mathbf{I}$ , the identity matrix) polynomials. In Eqn. (1),  $\tau$  is the  $P \times 1$  vector of regression coefficients for the average growth curve (AGC),

$$\tau = \frac{1}{N} \sum_{i=1}^N \tau_i,$$

where the coefficients specific to individual  $i$  have the distribution

$$\tau_i \sim MVN(\tau, \Lambda) \quad (2)$$

and  $\sigma^2$  measures the variability of individuals about their individual growth curves (error variance).

Swamy [6,7] generalized this model to allow differing measurement schedules for the  $N$  individuals, i.e.  $\mathbf{W}_i$  is specific to individual  $i$ .  $T$ , however, was assumed to be the same for each individual. He proposed estimators for the parameters of this generalized two-stage model and showed that they had certain desirable properties. In particular,  $\hat{\tau}$  (defined below) is efficient for large  $T$  and consistent when both  $N$  and  $T$  are large. Carter and Yang [3] noted that Swamy's results could be extended to the case where  $T$  was also individual-specific, but that one needed to require that both  $N$  and  $\min\{T_i\}$  be large, again a condition that is rarely satisfied in practice.

Swamy's estimators for the parameters of the two-stage model when both  $\mathbf{W}$  and  $T$  are allowed to be individual-specific are (all summations are from  $i = 1$  to  $N$ ):

$$\hat{\tau}_i = (\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'\mathbf{x}_i \quad (3)$$

$$\hat{\tau} = \frac{1}{N} \sum \hat{\tau}_i \quad (4)$$

$$\hat{\sigma}^2 = \frac{\Sigma(T_i - P)s_i}{\Sigma(T_i - P)} \quad (5)$$

and

$$\hat{\Lambda} = \mathbf{S}_\tau - \frac{\hat{\sigma}^2}{N} \sum (\mathbf{W}_i'\mathbf{W}_i)^{-1} \quad (6)$$

where

$$s_i = \frac{1}{T_i - P} \mathbf{x}'_i [\mathbf{I}_{T_i} - \mathbf{W}_i (\mathbf{W}'_i \mathbf{W}_i)^{-1} \mathbf{W}'_i] \mathbf{x}_i \tag{7}$$

and

$$\mathbf{S}_\tau = \frac{1}{N - 1} \sum (\hat{\tau}_i - \hat{\tau}) (\hat{\tau}_i - \hat{\tau})' \tag{8}$$

It will be noted that, in the above,  $P$  is fixed (a  $D = P - 1$  degree polynomial is fit to each individual) and that  $\mathbf{S}_\tau$  is the covariance matrix of the  $\hat{\tau}_i$ .

Carter and Yang [3] proposed a modification of  $\hat{\tau}$  that is efficient and consistent if either  $N$  or  $\min \{T_i\}$  is large, a condition which is more often satisfied in practical applications. This estimator is given by

$$\hat{\tau} = \left( \frac{1}{N} \sum \mathbf{A}_i \right)^{-1} \left( \frac{1}{N} \sum \mathbf{A}_i \hat{\tau}_i \right) \tag{9}$$

where

$$\mathbf{A}_i = [\hat{\Lambda} + \hat{\sigma}^2 (\mathbf{W}'_i \mathbf{W}_i)^{-1}]^{-1} \tag{10}$$

It can be seen that Eqn. (9) is a weighted average of the  $\hat{\tau}_i$ ; and it can be shown that the weights are in fact equal to the inverses of the variances of the  $\hat{\tau}_i$ , a long-standing statistical strategy for combining estimators with varying precisions (see e.g. Ref. 8, p. 444). To simplify some later notation, we let

$$\Omega = \left[ \frac{1}{N} \sum \mathbf{A}_i \right]^{-1} \tag{11}$$

so that

$$\hat{\tau} = \Omega \left[ \frac{1}{N} \sum \mathbf{A}_i \hat{\tau}_i \right]$$

Carter and Yang also noted that  $\hat{\Lambda}$  as given in Eqn. (6) could be non-positive definite (e.g. some of the variances might be negative) and so proposed a modification which is always positive definite and for which the large-sample properties of Swamy's estimator continue to hold. This estimator requires the computation of  $\hat{\gamma}$ , the smallest root of the determinantal equation

$$\left| \mathbf{S}_\tau - \frac{\hat{\gamma}}{N} \sum (\mathbf{W}'_i \mathbf{W}_i)^{-1} \right| = 0 \tag{12}$$

If  $\hat{\gamma} > \hat{\sigma}^2$ , no modification is required. If, however,  $\hat{\gamma} \leq \hat{\sigma}^2$  we take

$$\hat{\Lambda} = S_\tau - \frac{(\hat{\gamma} - 1/a)}{N} \sum (\mathbf{W}_i' \mathbf{W}_i)^{-1} \quad (13)$$

where

$$a = \sum T_i + \text{tr} \left[ \sum (\mathbf{W}_i' \mathbf{W}_i)^{-1} \right] \quad (14)$$

In Eqn. (14),  $\text{tr} [\cdot]$  is the trace operator, i.e. the sum of the diagonal elements of matrix within the brackets.

In summary, the  $\tau_i$  and  $\sigma^2$  are estimated in our program by Eqns. (3) and (5);  $\tau$  is estimated by Eqn. (9); and  $\Lambda$  by either Eqn. (6) or Eqn. (13) depending on whether  $\hat{\gamma} > \hat{\sigma}^2$  or  $\hat{\gamma} \leq \hat{\sigma}^2$  respectively.

### The degree of the polynomial

Carter and Yang did not explicitly address the question of the determination of  $D$ , the (common) degree of the polynomial to be fit to both the individual growth curves and to the AGC. In their illustrative example, they simply used  $D = 1$ , apparently guided by preliminary plots, etc. of the data. They did, however, indicate how hypotheses of the form  $\mathbf{H}:\mathbf{L}\tau = \tau_0$ , where  $\mathbf{L}$  is a specified matrix and  $\tau_0$  is a vector of known constants, could be tested using Hotelling's  $T^2$  statistic. We have implemented these tests in our program to aid the user in determining the degree of the polynomial to be used in the Carter-Yang procedure. We begin by asking the user to specify the degrees,  $FD$  and  $RD$ , of the so-called full and reduced models, respectively.  $FD \leq T - 1$  is a degree which the user feels should be more than adequate to model the data and  $RD$  the smallest degree which may be contemplated (often the user will choose  $RD = 1$ ). We then, using the level of significance specified by the user, perform a series of step-up goodness-of-fit tests, testing first that  $RD$  is adequate then, if necessary, that  $RD + 1$  is adequate, etc., up to degree  $FD$ , which is used if no smaller  $D$  passes the test for goodness-of-fit. These tests are similar to those described in Ref. 1 and are made using the test described above with appropriate choices of  $\mathbf{L}$  and  $\tau_0$ . For example, if  $FD = 4$  and  $RD = 1$  are specified, the first test will be based on

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and  $\tau_0' = [0, 0, 0]$ . This choice corresponds to the hypothesis that the coefficients of the quadratic, cubic and quartic terms are (simultaneously) zero, i.e. that  $\tau_3 = \tau_4 = \tau_5 = 0$ . If this hypothesis is not contradicted by the data we use  $D = 1$ . If rejected, we step-up and test  $\tau_4 = \tau_5 = 0$ , etc.

Should the user wish to bypass these tests and to specify the degree,  $D$ , to be used on other grounds, one simply takes  $FD = RD = D$ .

**Confidence intervals and bands**

We also compute confidence intervals for the elements of  $\tau$ , and confidence bands for the AGC at the planned times of measurement  $t_1, t_2, \dots, t_T$ . The user specifies the level of confidence (e.g. 0.95) to be used. Let  $\mathbf{w}$  be a  $1 \times P$  vector. Then approximate (for large values of  $N$  or  $\min\{T_i\}$ )  $(1 - \alpha) \times 100\%$  confidence intervals for the elements of  $\tau$  are given by

$$\mathbf{w}\hat{\tau} \pm \left(\frac{1}{N} \mathbf{w}\Omega\mathbf{w}'\right)^{1/2} t_{1-\alpha/2}(N-1) \tag{15}$$

by taking, in turn,

$$\mathbf{w}_1 = [1,0,0,\dots,0,0]$$

$$\mathbf{w}_2 = [0,1,0,\dots,0,0]$$

...

$$\mathbf{w}_P = [0,0,0,\dots,0,1]$$

In Eqn. (15),  $t_{1-\alpha/2}(N-1)$  denotes the  $(1 - \alpha/2) \times 100$ th percentile of the  $t$ -distribution with  $N - 1$  degrees of freedom.

To generate confidence bands for the AGC, identify  $\mathbf{w}$  with the rows of  $\mathbf{W}$  and let  $\hat{\mathbf{x}} = \mathbf{w}\hat{\tau}$ , the corresponding fitted value. Then marginal confidence bands are computed using

$$\hat{\mathbf{x}} \pm \left(\frac{1}{N} \mathbf{w}\Omega\mathbf{w}'\right)^{1/2} t_{1-\alpha/2}(N-1) \tag{16}$$

Simultaneous confidence bands are of the form

$$\hat{\mathbf{x}} \pm \left(\frac{(N-1)P}{N(N-P)} \mathbf{w}\Omega\mathbf{w}' F_{1-\alpha}(P, N-P)\right)^{1/2} \tag{17}$$

where  $F_{1-\alpha}(P, N-P)$  is the  $(1 - \alpha) \times 100$ th percentile of the  $F$ -distribution with  $P$  and  $N - P$  degrees of freedom. The former (marginal) bands are the ones given in Carter and Yang [3]; the latter were given in Ref. 1. We include both to allow comparisons with both publications.

In addition, we compute prediction intervals [3]. These are of the form

$$\hat{x}_f \pm \left[ \frac{1}{N} \mathbf{w}_f \Omega \mathbf{w}_f' + \mathbf{w}_f \Lambda \mathbf{w}_f' + \hat{\sigma}^2 \right]^{1/2} t_{1-\alpha/2}(N-1) \quad (18)$$

These intervals were not included in Ref. 1, but they can be useful in predictive/diagnostic contexts. They provide bands within which the growth curve of a new individual from the population under consideration will lie with the desired level of confidence.

### An example

The structure of the program is completely analogous to 2STG [1], except that the user is also asked to supply a code for missing data and so no detailed description is provided here. Information concerning obtaining a copy of the program, hardware requirements, etc., is provided in the Appendix. The program is invoked by issuing the single command

**gsruncy**

Our example is based on the data set consisting of mandibular ramus height measurements for  $N = 12$  young male rhesus monkeys at  $T = 5$  (equally spaced) time points, first considered in Ref. 4. Here, however, we have randomly discarded 10% of the data (6 measurements), basing the analysis on the data set

$$\mathbf{X} = \begin{bmatrix} 25.2 & 29.0 & 33.6 & 9999 & 35.8 \\ 27.3 & 32.1 & 37.0 & 41.8 & 43.5 \\ 26.3 & 9999 & 36.1 & 38.0 & 38.9 \\ 26.0 & 34.5 & 39.0 & 42.3 & 44.4 \\ 9999 & 29.5 & 34.4 & 38.3 & 37.9 \\ 28.2 & 32.5 & 36.3 & 42.3 & 43.8 \\ 25.4 & 33.4 & 38.0 & 9999 & 43.1 \\ 27.2 & 34.8 & 9999 & 44.0 & 44.0 \\ 26.0 & 34.5 & 38.0 & 43.5 & 43.8 \\ 28.5 & 33.8 & 38.0 & 39.2 & 42.0 \\ 27.0 & 9999 & 36.0 & 41.7 & 43.8 \\ 26.0 & 33.0 & 40.2 & 42.5 & 43.8 \end{bmatrix}$$

where the 9999s represent missing data. For these data with  $D = 3$ , the estimate of  $\Lambda$  was not positive definite and the Carter-Yang modification (Eqn. (13)) was required. The program CY prints both the estimator and, when necessary, the modification. The estimated error variance was 1.021. The estimated polynomial regression coefficients and their corresponding 95% confidence intervals are given in Table I.

TABLE I

95% CONFIDENCE INTERVALS FOR THE REGRESSION COEFFICIENTS WHEN MISSING DATA ARE PRESENT AND CY IS USED

| Degree | Coeff  | Std error | Confidence interval |        | Half interval |
|--------|--------|-----------|---------------------|--------|---------------|
| 0      | 20.927 | 2.224     | 16.033              | 25.822 | 4.895         |
| 1      | 5.456  | 3.065     | -1.291              | 12.202 | 6.747         |
| 2      | 0.391  | 1.089     | -2.006              | 2.788  | 2.397         |
| 3      | -0.127 | 0.116     | -0.382              | 0.128  | 0.255         |

TABLE II

95% MARGINAL CONFIDENCE BANDS FOR THE AGC WHEN MISSING DATA ARE PRESENT AND CY IS USED

| Time point | Fitted value | Std error | Confidence interval |        | Half interval |
|------------|--------------|-----------|---------------------|--------|---------------|
| 1          | 26.647       | 0.335     | 25.909              | 27.385 | 0.738         |
| 2          | 32.386       | 0.631     | 30.997              | 33.775 | 1.389         |
| 3          | 37.382       | 0.644     | 35.965              | 38.799 | 1.417         |
| 4          | 40.873       | 0.703     | 39.326              | 42.421 | 1.547         |
| 5          | 42.098       | 0.878     | 40.165              | 44.031 | 1.933         |

TABLE III

95% SIMULTANEOUS CONFIDENCE BANDS FOR AGC WHEN MISSING DATA ARE PRESENT AND CY IS USED

| Time point | Fitted value | Std error | Confidence interval |        | Half interval |
|------------|--------------|-----------|---------------------|--------|---------------|
| 1          | 26.647       | 0.786     | 25.106              | 27.385 | 1.541         |
| 2          | 32.386       | 1.480     | 29.486              | 35.285 | 2.900         |
| 3          | 37.382       | 1.510     | 34.425              | 40.399 | 2.957         |
| 4          | 40.873       | 1.649     | 37.643              | 44.104 | 3.230         |
| 5          | 42.098       | 2.060     | 38.062              | 46.133 | 4.036         |

The 95% marginal and simultaneous confidence bands for the AGC are shown in Tables II and III, respectively. The 95% prediction intervals are shown in Table IV.

The level of confidence used (e.g. 0.95) is specified by the user. Finally, the AGC and its confidence bands are plotted. Publication-quality graphs may be printed. More details concerning the output and program operation are given in Ref. 1. As

TABLE IV

95% PREDICTION INTERVALS WHEN MISSING DATA ARE PRESENT AND CY IS USED

| Time point | Fitted value | Std error | Prediction interval |        | Half interval |
|------------|--------------|-----------|---------------------|--------|---------------|
| 1          | 26.647       | 1.171     | 24.068              | 29.225 | 2.578         |
| 2          | 32.386       | 2.267     | 27.396              | 37.375 | 4.990         |
| 3          | 37.382       | 2.408     | 32.082              | 42.682 | 5.300         |
| 4          | 40.873       | 2.534     | 35.296              | 46.451 | 5.578         |
| 5          | 42.098       | 3.169     | 35.124              | 49.072 | 6.974         |

is indicated there, the screen output is automatically saved in a file, called CY.OUT, which can be edited and/or annotated using a word processor.

### Discussion

It is of interest to compare the above output from CY with that of 2STG when there are no missing data. For the complete data set, using 2STG, the estimated error variance was 1.228. The estimate of  $\Lambda$  was not positive definite and the Carter-Yang modification (Eqn. (13)) was again required. The estimated polynomial regression coefficients and 95% confidence intervals are shown in Table V; the 95% simultaneous confidence bands for the AGC in Table VI (marginal confidence bands for the AGC and prediction intervals were not included in 2STG.)

In the example considered, the presence of missing data did not change the estimates of the parameters of the two-stage model to any great extent. The widths of the confidence intervals and bands are somewhat wider, but comparable. This is despite the fact that neither  $N$  nor  $\min\{T_i\}$  is 'large'. The reader is reminded, however, that the desirable properties of the Carter-Yang procedure cited above can only be ensured when one or the other of these quantities is in fact large.

TABLE V

95% CONFIDENCE INTERVALS FOR THE REGRESSION COEFFICIENTS WHEN NO MISSING DATA ARE PRESENT AND 2STG IS USED

| Degree | Coeff  | Std error | Confidence interval |        | Half interval |
|--------|--------|-----------|---------------------|--------|---------------|
| 0      | 21.062 | 1.358     | 18.070              | 24.053 | 2.991         |
| 1      | 5.209  | 1.788     | 1.272               | 9.145  | 3.936         |
| 2      | 0.453  | 0.592     | -0.850              | 1.756  | 1.303         |
| 3      | -0.131 | 0.060     | -0.263              | 0.002  | 0.132         |



TABLE VI

95% SIMULTANEOUS CONFIDENCE BANDS FOR THE AGC WHEN NO MISSING DATA ARE PRESENT AND 2STG IS USED

| Time point | Fitted value | Std error | Confidence interval |        | Half interval |
|------------|--------------|-----------|---------------------|--------|---------------|
| 1          | 26.593       | 0.752     | 25.120              | 28.065 | 1.472         |
| 2          | 32.246       | 1.188     | 29.920              | 34.573 | 2.327         |
| 3          | 37.239       | 1.386     | 34.524              | 39.954 | 2.715         |
| 4          | 40.788       | 1.583     | 37.687              | 43.888 | 3.101         |
| 5          | 42.109       | 1.993     | 38.205              | 46.014 | 3.905         |

It will be noted that the program is structured for the analysis of data from studies designed in such a way that common times of measurement were planned, but some data were inadvertently missing. It is theoretically possible to remove the restriction of common times of measurement, i.e. to allow completely different measurement schedules for each individual. This, however, raises a number of thorny practical problems and has not been implemented at this time. The main cause for concern is a by-product of the fact, as can be seen from Eqns. (9) and (10), that a single value of  $D$  must be selected which is adequate to model each individual's growth profile. When individuals present with widely separated, non-overlapping times of measurement, extrapolation can be a problem — a polynomial of a given degree may fit an individual growth profile quite well over the range of observations for that individual, but may extrapolate poorly, especially when higher degree polynomials are involved. While this problem can arise even in the restricted context of planned times of measurement, it will generally be less severe and the program described here should be useful in many situations where incidentally missing data points are present. We suggest that use of the current version of this program be limited to data sets where the times of measurement for individuals show at least a reasonable amount of overlap and where high degree polynomials are not required to achieve an adequate fit to the data.

We should also note that Carter and Yang, while developing the modified estimator, Eqn. (13), which will always be positive semi-definite, warned that when modification is in fact necessary, "one should question the validity of the model specification". Thus this may be used as an informal test of the goodness-of-fit of the two-stage model. For a more detailed description of how one might expect the two-stage model to arise in practice, see Ref. 1.

Finally, we mention that there are other approaches to the analysis of incomplete longitudinal data sets. These, however, either depend on an iterative computational algorithm (the so-called EM algorithm) for estimating the overall regression parameters or are concerned only with the estimation of the individual regression parameters [9]. The EM algorithm [10,11] can be used in situations like the one described in this paper, but Carter and Yang noted that the EM algorithm can be slow to converge, sensitive to the starting values and may converge to local rather than global

maxima. In any event, the Carter-Yang procedure produces closed-form expressions for the estimators of the parameters of the two-stage model and should prove useful in situations in which incidentally missing data points exist.

### Acknowledgement

Supported by grant DE08730 from the National Institute of Dental Research.

### Appendix: computer implementation

A full set of PC programs for longitudinal data analysis, including this program, can be obtained on 5.25" or 3.5" diskettes (please request type) by sending \$25 to defray the cost of handling and licensing fees. These programs require an 80386 or 80486 based personal computer (PC) running the MS-DOS operating system (version 5.0 or higher is recommended, although versions as low as 3.3 will suffice). 80386 computers must also be equipped with a 80387-math coprocessor. At least 4 mb of memory is required and must be available to GAUSS386i, i.e. not in use by memory resident programs such as Windows. EGA or VGA graphic capabilities are required to display the color graphics; VGA or SVGA is suggested to optimally display the graphic results. Runtime modules are supplied with the programs so that no additional software (i.e. compiler or interpreter) is required to run these programs. One can create and edit ASCII data sets for use by these programs using the full screen editor supplied with MS-DOS version 5.0. The programs are written and compiled using GAUSS386i, version 3.0, require no additional installation or modification and are run with a single command. When requesting the programs, address inquiries to the corresponding author and make checks payable to Baylor College of Dentistry.

### References

- 1 Ten Have TR, Kowalski CJ and Schneiderman ED: PC program for analyzing one-sample longitudinal data sets which satisfy the two-stage polynomial growth curve model, *Am J Hum Biol*, 3 (1991) 269–279.
- 2 Rao CR: The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrics*, 52 (1965) 447–458.
- 3 Carter RL and Yang MCK: Large-sample inference in random coefficient regression models, *Commun Stat: Theory and Methods*, 8 (1986) 2507–2526.
- 4 Schneiderman ED and Kowalski CJ: Implementation of Rao's one-sample polynomial growth curve model using SAS, *Am J Phys Anthropol*, 67 (1985) 323–333.
- 5 Ten Have TR, Kowalski CJ and Schneiderman ED: A PC program for obtaining orthogonal polynomial regression coefficients for use in longitudinal data analysis, *Am J Hum Biol*, 4 (1992) 403–416.
- 6 Swamy PAVB: Efficient inference in a random coefficient regression model, *Econometrica*, 38 (1970) 311–323.
- 7 Swamy PAVB: *Statistical Inference in Random Coefficient Regression Models*, Springer-Verlag, Berlin, 1971.
- 8 Lehman EL: *The Theory of Point Estimation*, Wiley, New York, 1983.
- 9 Reinsel GC: Mean squared error properties of empirical Bayes estimators in a multivariate random effects general linear model, *J Am Stat Assoc*, 80 (1985) 642–650.
- 10 Dempster AP, Laird NM and Rubin DB: Maximum likelihood with incomplete data via the EM algorithm, *J Royal Stat Soc, Ser B* 39 (1981) 1–38.
- 11 Laird NM and Ware JH: Random effects models for longitudinal data, *Biometrics*, 38 (1982) 963–974.