163

STATISTICS IN MEDICINE, VOL. 13, 1643–1655 (1994)

MULTILEVEL TIME SERIES MODELS WITH APPLICATIONS TO REPEATED MEASURES DATA

HARVEY GOLDSTEIN, MICHAEL J. R. HEALY AND JON RASBASH
Institute of Education, University of London, 20 Bedford Way, London, WC1H 0AL, U.K.

SUMMARY

The analysis of repeated measures data can be conducted efficiently using a two-level random coefficients model. A standard assumption is that the within-individual (level 1) residuals are uncorrelated. In some cases, especially where measurements are made close together in time, this may not be reasonable and this additional correlation structure should also be modelled. A time series model for such data is proposed which consists of a standard multilevel model for repeated measures data augmented by an autocorrelation model for the level 1 residuals. First- and second-order autoregressive models are considered in detail, together with a seasonal component. Both discrete and continuous time are considered and it is shown how the autocorrelation parameters can themselves be structured in terms of further explanatory variables. The models are fitted to a data set consisting of repeated height measurements on children.

1. INTRODUCTION

In the typical repeated measures study we can write the common basic model for the ith measurement or reading on the jth subject as

$$y_{ij} = \sum_{k=0}^{p} \gamma_{jk} t_{ij}^{k} + \sum_{l=1}^{q} \alpha_l z_{ijl} + e_{ij} \tag{1}$$

where t refers to time or age and the z's are covariates. In the terminology of multilevel modelling[1], the successive measurements or readings constitute level 1 and are nested within the individuals, level 2. In a growth curve context, t_{ij} usually refers to chronological age. The covariates z_{lij} may be defined at either the individual or occasion level. Commonly, the lower order polynomial coefficients γ_{jk} $(k = 0, \cdots, s, s < p)$, are taken to be random at level 2 with coefficient values varying and covarying between individuals. In the standard formulation the level 1 random terms e_{ij} are assumed to be distributed independently with zero mean and constant variance.

This basic model has been studied by several authors, among the earliest being Laird and Ware;[2] Goldstein[1] (Chapter 4) extended it to the case where the level 1 variance is allowed to be a function of further variables, notably time or age. An important advantage of (1) over previous formulations (see, for example, Grizzle and Allen [3]) is that the time intervals need not be constant; the number and spacing of measurements can vary from subject to subject.

The model given by (1) is a natural formulation for growth data, since it separates explicitly between-individual and within-individual variation. From a biological point of view modelling

CCC 0277–6715/94/161643–13

Received April 1993
Revised October 1993

with randomly varying coefficients has several advantages. After fitting the model we can obtain posterior estimates of the coefficients for each individual to characterize growth in terms of its average rate of change, acceleration, etc. This might then be used diagnostically to detect those with very abnormal growth patterns. Secondly, it allows the fitting of group factors and other covariates to study both how the mean values of the growth parameters change across groups and to see how much of the group differences and the between-individual variation can be explained by other factors.

In reality the assumption of independence for the level 1 residuals is often unrealistic Equation (1) models a smooth summary curve for each individual graduating the set of measurements for that individual. Clearly, when many measurements are taken close enough together in time, and assuming growth to be a continuous process but rougher than the smoothness implied by models such as (1), then the level 1 residuals about each individual's underlying smooth curve will tend to be correlated, a positive residual tending to be followed by another positive residual etc. We therefore require models which explicitly recognize the possibility of an autocorrelation structure among the e_{ij}. The purpose of the present paper is to study such models in a multilevel framework thereby extending the existing literature.

Some recent papers have addressed this issue. Chi and Reinsel[4] and Lee[5] consider the extension of (1) when there is a fixed set of occasions with equal intervals for each subject together with a discrete time first-order autoregressive (AR(1)) model for the level 1 random variables. Diggle[6] uses a model where the covariance between level 1 residuals which are s time units apart is given by

$$\sigma^2 e^{-\alpha s^c}.$$

This model allows the spacing of occasions to vary continuously and reflects the fact that the serial correlation will decrease as s increases. Furthermore, when c is equal to one, it is the continuous time analogue of a first-order autoregressive model which is also considered by Jones and Boadi-Boateng[7]. These papers derive maximum likelihood estimators, but Diggle only considers a variance components model, that is where a simple variance term describes the level 2 variation.

Jennrich and Schluchter[8] also consider autoregressive models, but these are applied to the total covariance matrix without separately modelling the between- and within-individual variation as in (1).

We generalize (1) to consider the specification of a more general class of level 1 serial correlation models for both equal interval, discrete time and continuous time.

2. THE MULTILEVEL LINEAR MODEL

For simplicity we consider the two-level model. Extensions to three or more levels are straightforward, and in particular we can deal with autocorrelation structures at higher levels.

We can rewrite (1) in the general form

$$y_{ij} = \Sigma_k \beta_{jk} x_{ijk} + e_{ij}$$

which also allows for random coefficients of covariates. For a coefficient in this model which does vary across individuals we can write

$$\beta_{jk} = \beta_k + e_{2kj}$$

so that, writing z for x when a random coefficient is present, a general two-level model can be written as follows

$$y_{ij} = \Sigma_k x_{ijk}\beta_k + \sum_{l=1}^{p_2} z_{2lij}e_{2lj} + \sum_{m=1}^{p_1} z_{1mij}e_{1mij} \tag{2}$$

where p_2 is the number of coefficients random at level 2 and where there are p_1 coefficients random at level 1 to allow for a complex level 1 variance structure. In matrix form we have

$$Y = X\beta + E_2 + E_1$$

$$E_1 = \{e_{1m}\}, \quad E_2 = \{e_{2l}\}$$

$$E(e_{1mij}e_{1mi'j'}) = 0, \quad i \neq i', \quad \text{for all } j \text{ and } j'$$

$$E(e_{2lj}e_{2lj'}) = 0, \quad j \neq j'$$

where $e_{2l} = \{z_{2lij}e_{2lj}\}$ and e_{2lj} is the lth level 2 random variable with zero mean, $e_{1m} = \{z_{1mij}e_{1mij}\}$ and e_{1mij} is the mth level 1 random variable with zero mean, and $Z_2 = \{z_{2lij}\}$ and $Z_1 = \{z_{1mij}\}$ are the level 2 and level 1 random component design matrices. The level 2 random variables E_2 are assumed to be independent of the level 1 random variables E_1. The vector $\beta = \{\beta_k\}$ contains the fixed coefficients to be estimated, and $X = \{X_{ij}\}$ is the design matrix for the fixed coefficients. The columns of X and Z_1 and Z_2 may or may not coincide. Typically, Z_1 and Z_2 are subsets of X and the associated random variables are referred to as random coefficients corresponding to the elements of β.

In this two-level model the covariance matrix of the observations

$$V = E\{Y - X\beta)(Y - X\beta)^T\}$$

is block diagonal. The jth block of this, corresponding to observations on the jth subject, is modelled by the following structure:[9]

$$V_j = V_{1j} + Z_{2j}\Omega_2 Z_{2j}^T \tag{3}$$

where $\Omega_2 = \text{cov}(E_2)$. The matrix Z_{2j} is the subset of Z_2 for the jth block and contains the values of the explanatory variables for the variables random at level 2. The matrix V_{1j} is the contribution to V_j of the level 1 variation and is usually taken to be diagonal. Thus, in the simplest case of a variance components model with a single between-level-2 units variance σ_{e2}^2, and a single between-level-1 units variance σ_{e1}^2, we have

$$V_{1j} = \sigma_{e1}^2 I_{n_j}, \quad \Omega_2 = \sigma_{e2}^2$$

where I_{n_j} is the identity matrix of order n_j which is the number of level 1 units in the jth level 2 unit and Z_1, Z_2 are now vectors of ones. The models considered in this paper extend the model given by (3) by allowing the level 1 residuals to be autocorrelated, that is $E(e_{1mij}e_{1mi'j}) \neq 0$, and this gives rise to complex structures for the V_{1j}.

Assuming multivariate Normality, we have a choice of algorithms for obtaining maximum likelihood estimators. Here we use the Iterative Generalized Least Squares (IGLS) algorithm, one advantage of which is that it provides a conceptually straightforward procedure based upon generalized least squares regression and is easily extensible as shown below. In the absence of autocorrelation the estimates which are produced are maximum likelihood by virtue of the equivalence between maximum likelihood and IGLS estimates and between restricted maximum likelihood and restricted unbiased IGLS estimates for Normally distributed data.[9,10] The algorithm proceeds as shown in Appendix I.

3. FITTING THE DISCRETE TIME AUTOREGRESSIVE MODEL

Consider first the single level stationary AR(1) model with observations e_t given by

$$e_t = \rho e_{t-1} + v_t, \quad \text{var}(e_t) = \sigma_e^2, \quad \text{var}(v_t) = \sigma_v^2, \quad E(v_t) = 0. \tag{4}$$

We can obtain either maximum likelihood or least squares estimates for ρ, σ_e^2 and σ_v^2; for a long series these estimates are equivalent. Using the simpler, non-iterative least squares procedure we would estimate ρ by treating (4) as a simple regression through the origin.

Now consider the two-level model (1) in which the e_{ij} follow an AR(1) process. A natural generalization might be as follows. At each iteration we obtain the (predicted) level 1 residuals, $\hat{e}_{ij}$ (see for example Goldstein,[1] Appendix 3.2). Least squares estimates of the autoregressive parameters could then be obtained by a simple regression, pooling over the level 2 units. The level 1 residual estimates, however, are biased, with the bias depending on the number of level 1 units in the level 2 unit. We could work with unbiased, adjusted, residuals $\hat{e}_{ij} - \hat{e}_{.j}$ but these are negatively correlated, the correlation being $-1{\cdot}0$ if $n_j = 2$. The correlation approaches zero only as n_j becomes large. In the typical case with small n_j, therefore, this procedure will be unsuitable.

We proceed instead by extending the design matrix X^{**} (see Appendix I) to incorporate a further random parameter design vector whose coefficient is ρ. Consider the extension of (1) where the level 1 residuals e_{ij} follow an AR(1) process as in (4). The level 1 contribution to the residual covariance matrix for V_{1j} (in addition to the diagonal matrix $\sigma_e^2 I_{(n_j)}$) has (k, l) element $\rho^{|k-l|}\sigma_e^2$, and we can write, omitting the subscript for level 1,

$$V_{1j} = \sigma_e^2 I_{(n_j)} + \rho\sigma_e^2 R_{(n_j)} \tag{5}$$

where the (k, l)th element of $R_{(n_j)}$, is defined as

$$R_{(n_j)}(k, l) = \begin{cases} 0 & \text{if } k = 1 \\ \rho^{(|k-l|-1)} & \text{if } k \neq l \end{cases}$$

The vectors $\text{vech}(\sigma_e^2 R_{(n_j)})$ and $\text{vech}(I_{(n_j)})$ can be included as explanatory variable vectors in X^{**}, with the additional coefficients ρ, σ_e^2 to be estimated.

The estimation typically proceeds as follows. We first obtain a solution which excludes the autoregressive component, then using a suitable starting value for ρ calculate $R_{(n_j)}$; in many cases zero can be used. Updated estimates of ρ, σ_e^2 are obtained and after each iteration $R_{(n_j)}$ is recalculated and the procedure repeated until convergence. Care is needed over numerical precision to avoid rounding errors in the calculation of V. For prediction of future level 1 residuals we can use the estimated autocovariances, with the set of past estimated level 1 residuals as explanatory variables in a linear predictor.

The above discussion has assumed complete data, but the extension where observations at some occasions are missing is straightforward. The only modification involves noting that some subdiagonals of $R_{(n_j)}$ will contain lags of different degree.

3.1. The second-order autoregressive model

The model for a second-order autoregressive sequence in discrete time can be written as

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + v_t \tag{6}$$

so that, writing z for x when a random coefficient is present, a general two-level model can be written as follows

$$y_{ij} = \Sigma_k x_{ijk}\beta_k + \sum_{l=1}^{p_2} z_{2lij}e_{2lj} + \sum_{m=1}^{p_1} z_{1mij}e_{1mij} \tag{2}$$

where p_2 is the number of coefficients random at level 2 and where there are p_1 coefficients random at level 1 to allow for a complex level 1 variance structure. In matrix form we have

$$Y = X\beta + E_2 + E_1$$

$$E_1 = \{e_{1m}\}, \quad E_2 = \{e_{2l}\}$$

$$E(e_{1mij}e_{1mi'j'}) = 0, \quad i \neq i', \quad \text{for all } j \text{ and } j'$$

$$E(e_{2lj}e_{2lj'}) = 0, \quad j \neq j'$$

where $e_{2l} = \{z_{2lij}e_{2lj}\}$ and e_{2lj} is the lth level 2 random variable with zero mean, $e_{1m} = \{z_{1mij}e_{1mij}\}$ and e_{1mij} is the mth level 1 random variable with zero mean, and $Z_2 = \{z_{2lij}\}$ and $Z_1 = \{z_{1mij}\}$ are the level 2 and level 1 random component design matrices. The level 2 random variables E_2 are assumed to be independent of the level 1 random variables E_1. The vector $\beta = \{\beta_k\}$ contains the fixed coefficients to be estimated, and $X = \{X_{ij}\}$ is the design matrix for the fixed coefficients. The columns of X and Z_1 and Z_2 may or may not coincide. Typically, Z_1 and Z_2 are subsets of X and the associated random variables are referred to as random coefficients corresponding to the elements of β.

In this two-level model the covariance matrix of the observations

$$V = E\{Y - X\beta)(Y - X\beta)^T\}$$

is block diagonal. The jth block of this, corresponding to observations on the jth subject, is modelled by the following structure:[9]

$$V_j = V_{1j} + Z_{2j}\Omega_2 Z_{2j}^T \tag{3}$$

where $\Omega_2 = \text{cov}(E_2)$. The matrix Z_{2j} is the subset of Z_2 for the jth block and contains the values of the explanatory variables for the variables random at level 2. The matrix V_{1j} is the contribution to V_j of the level 1 variation and is usually taken to be diagonal. Thus, in the simplest case of a variance components model with a single between-level-2 units variance σ_{e2}^2, and a single between-level-1 units variance σ_{e1}^2, we have

$$V_{1j} = \sigma_{e1}^2 I_{n_j}, \quad \Omega_2 = \sigma_{e2}^2$$

where I_{n_j} is the identity matrix of order n_j which is the number of level 1 units in the jth level 2 unit and Z_1, Z_2 are now vectors of ones. The models considered in this paper extend the model given by (3) by allowing the level 1 residuals to be autocorrelated, that is $E(e_{1mij}e_{1mi'j}) \neq 0$, and this gives rise to complex structures for the V_{1j}.

Assuming multivariate Normality, we have a choice of algorithms for obtaining maximum likelihood estimators. Here we use the Iterative Generalized Least Squares (IGLS) algorithm, one advantage of which is that it provides a conceptually straightforward procedure based upon generalized least squares regression and is easily extensible as shown below. In the absence of autocorrelation the estimates which are produced are maximum likelihood by virtue of the equivalence between maximum likelihood and IGLS estimates and between restricted maximum likelihood and restricted unbiased IGLS estimates for Normally distributed data.[9,10] The algorithm proceeds as shown in Appendix I.

3. FITTING THE DISCRETE TIME AUTOREGRESSIVE MODEL

Consider first the single level stationary AR(1) model with observations e_t given by

$$e_t = \rho e_{t-1} + v_t, \quad \text{var}(e_t) = \sigma_e^2, \quad \text{var}(v_t) = \sigma_v^2, \quad E(v_t) = 0. \tag{4}$$

We can obtain either maximum likelihood or least squares estimates for ρ, σ_e^2 and σ_v^2; for a long series these estimates are equivalent. Using the simpler, non-iterative least squares procedure we would estimate ρ by treating (4) as a simple regression through the origin.

Now consider the two-level model (1) in which the e_{ij} follow an AR(1) process. A natural generalization might be as follows. At each iteration we obtain the (predicted) level 1 residuals, $\hat{e}_{ij}$ (see for example Goldstein,[1] Appendix 3.2). Least squares estimates of the autoregressive parameters could then be obtained by a simple regression, pooling over the level 2 units. The level 1 residual estimates, however, are biased, with the bias depending on the number of level 1 units in the level 2 unit. We could work with unbiased, adjusted, residuals $\hat{e}_{ij} - \hat{e}_{.j}$ but these are negatively correlated, the correlation being $-1{\cdot}0$ if $n_j = 2$. The correlation approaches zero only as n_j becomes large. In the typical case with small n_j, therefore, this procedure will be unsuitable.

We proceed instead by extending the design matrix X^{**} (see Appendix I) to incorporate a further random parameter design vector whose coefficient is ρ. Consider the extension of (1) where the level 1 residuals e_{ij} follow an AR(1) process as in (4). The level 1 contribution to the residual covariance matrix for V_{1j} (in addition to the diagonal matrix $\sigma_e^2 I_{(n_j)}$) has (k, l) element $\rho^{|k-l|}\sigma_e^2$, and we can write, omitting the subscript for level 1,

$$V_{1j} = \sigma_e^2 I_{(n_j)} + \rho\sigma_e^2 R_{(n_j)} \tag{5}$$

where the (k, l)th element of $R_{(n_j)}$, is defined as

$$R_{(n_j)}(k, l) = \begin{cases} 0 & \text{if } k = 1 \\ \rho^{(|k-l|-1)} & \text{if } k \neq l \end{cases}$$

The vectors $\text{vech}(\sigma_e^2 R_{(n_j)})$ and $\text{vech}(I_{(n_j)})$ can be included as explanatory variable vectors in X^{**}, with the additional coefficients ρ, σ_e^2 to be estimated.

The estimation typically proceeds as follows. We first obtain a solution which excludes the autoregressive component, then using a suitable starting value for ρ calculate $R_{(n_j)}$; in many cases zero can be used. Updated estimates of ρ, σ_e^2 are obtained and after each iteration $R_{(n_j)}$ is recalculated and the procedure repeated until convergence. Care is needed over numerical precision to avoid rounding errors in the calculation of V. For prediction of future level 1 residuals we can use the estimated autocovariances, with the set of past estimated level 1 residuals as explanatory variables in a linear predictor.

The above discussion has assumed complete data, but the extension where observations at some occasions are missing is straightforward. The only modification involves noting that some subdiagonals of $R_{(n_j)}$ will contain lags of different degree.

3.1. The second-order autoregressive model

The model for a second-order autoregressive sequence in discrete time can be written as

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + v_t \tag{6}$$

whence we obtain the following relationships

$$\begin{aligned}
\operatorname{cov}(e_t e_{t-1}) &= \rho_1(1-\rho_2)^{-1}\sigma_e^2 = c_1^*\sigma_e^2 \\
\operatorname{cov}(e_t e_{t-2}) &= (\rho_1 c_1^* + \rho_2)\sigma_e^2 = c_2^*\sigma_e^2 \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
\operatorname{cov}(e_t e_{t-k}) &= (\rho_1 c_{k-1}^* + \rho_2 c_{k-2}^*)\sigma_e^2 = c_2^*\sigma_e^2
\end{aligned} \tag{7}$$

We therefore extend (5) and the associated estimating procedure by defining two explanatory variables $R^{(1)}$, $R^{(2)}$ for the parameters, ρ_1, ρ_2 based upon (7). The values c_k^* are calculated recursively, based upon current values of the parameters. The extension to higher order processes follows similar lines.

To obtain starting values we can fit an AR(1) model and then add the extra explanatory variable $R^{(2)}$ to the model with a starting value for ρ_2 of zero.

4. CONTINUOUS TIME MODELS

The restriction of the models in the previous section to discrete equally spaced time points, even with possibly missing data, may be unwelcome. Even in studies which plan to collect data at fixed intervals, there is often variation in the observed intervals, and failure to recognize this may lead to inefficiency and biased estimates of the autocorrelation parameters.

We first consider a generalisation of the models of Diggle[6] and Jones and Boadi-Boateng[7] for the autocorrelation structure. Dropping the subscript for level 2, we assume that the level 1 residuals have a covariance structure given by

$$\operatorname{cov}(e_t e_{t-s}) = \sigma_e^2 f(\alpha, s) = \sigma_e^2 \exp(-g(\alpha, s)) \tag{8}$$

where $g(\alpha, s)$ is any positive increasing function of s, not necessarily linear, and α is a vector of p parameters

$$\alpha = \{\alpha_k\}, \quad k = 1, \ldots, p.$$

Estimation is described in Appendix II.

The choice for g needs to recognize that it should be positive and increasing and contain as few parameters as necessary to be flexible enough to describe real data. It will also be convenient to restrict g to a linear function of parameters to avoid computational complexities. There seems to be little substantive guidance on choice, and it is likely that different functional forms will be appropriate for different kinds of data.

An obvious first choice is the analogue of the discrete time AR(1) model

$$g(\alpha, s) = \alpha s \tag{9}$$

and the estimation is described in Appendix II.

Another possible choice for g is the polynomial $\Sigma_k \alpha_k s^k$, constrained to be positive. A difficulty with this model is that successive powers will tend to be highly correlated and this causes estimation difficulties. Alternatively, we may add an inverse polynomial term to give

$$g(\alpha, s) = \alpha_1 s + \alpha_2 s^{-1} \tag{10}$$

which is linear in the parameters and avoids the high correlations associated with the ordinary polynomial. Equation (10) can be modified easily for particular data sets, for example by choosing

s^2 and s^{-2} instead of s and s^{-1} to change the rate of serial decay. Other choices for g might involve the addition of the logarithm or non-integer powers of s.

5. MORE COMPLEX LEVEL 1 STRUCTURES

The model is readily extended to incorporate further structure at level 1, that is between measurement occasions. In some applications the autocorrelation parameter may be a function of time, with the level 1 variance remaining constant. Group differences in σ_e^2 can be incorporated by specifying suitable dummy variables. In continuous time models the parameters α_k can be functions of further explanatory variables which can be incorporated directly into the model using the formula in Appendix II. Thus we can write a more general version of (9) as

$$g(\alpha, s) = (\alpha_0 + \alpha_1 z_{1j} + \alpha_2 z_{2ij})s$$

where z_{1j} is defined at level 2 and might refer, for example, to group membership. The variable z_{2ij} is defined at level 1 and so allows the autocorrelation parameter to vary over time, for example according to season of year or age of subject. A special case of this model is where some measurements are replicated. In growth studies, for example, some subjects may be measured twice on some occasions. In this case we can extend (9) as follows

$$g(\alpha, s) = \begin{cases} \alpha_1 s & \text{if no replicate} \\ \alpha_2 & \text{if replicate} \end{cases}$$

which gives an estimate of measurement reliability $\exp(-\alpha_2)$.

Another useful extension is where the simple level 1 variance term σ_e^2 is made a function of time or age, by specifying that one or more of the polynomial coefficients vary randomly at level 1. To specify the level 1 variance as a quadratic function of time for the basic model given by (9) we write

$$g = \alpha s + \beta_1(t_1 + t_2) + \beta_2(t_1^2 + t_2^2)$$

where t_1 and t_2, respectively, refer to the two time points, and as $s \to 0$ this implies that the variance is a quadratic function of time

$$\text{var}(y_t) = \sigma_e^2 \exp\{-2(\beta_1 t + \beta_2 t^2)\}$$

$$\text{corr}(y_t y_{t+s}) = \alpha s$$

and we have the same interpretation for α as before.

We can define the level 1 variation as a function of other covariates in a similar fashion. If such a covariate is measured at level 1, for example a characteristic of a measurer or rater, then it should be a symmetric function of the two time points, t_1 and t_2. If measured at level 2, or above, it can be quite general, for example referring to group differences.

6. APPLICATION TO HEIGHT MEASUREMENTS

The data for the following examples consist of height measurements on a sample of 26 boys each measured on nine occasions between the ages of 11 and 14 years. They were obtained from a residential school in Oxfordshire.[11] The measurements were taken on the same days for all children, approximately three months apart. Although this dataset is relatively small, and the measurements are only moderately close together in time, it suffices to demonstrate the procedures we have developed. All the computations have been carried out using the ML3 software package.[12]

Table I. Height as a fourth degree polynomial on age

Parameter	Estimate (SE) (A)	Estimate (SE) (B)
Fixed		
Intercept	148·9	149·0
age	6·18 (0·35)	6·17 (0·41)
age^2	1·10 (0·35)	1·13 (0·34)
age^3	0·42 (0·16)	0·46 (0·19)
age^4	− 0·33 (0·30)	− 0·38 (0·29)

Random: level 2 covariance matrix with standard errors in brackets

	Intercept	age	age^2	age^3	Intercept	age	age^2
Intercept	61·7 (17·1)				61·6 (16·5)		
age	9·3 (3·7)	4·0 (1·2)			8·0 (2·9)	2·8 (0·7)	
age^2	1·3 (1·4)	1·0 (0·4)	0·7 (0·2)		1·4 (1·4)	0·9 (0·3)	0·6 (0·2)
age^3	− 1·7 (1·5)	− 0·9 (0·5)	− 0·1 (0·2)	0·3 (0·3)			
Random: level 1 σ_e^2	0·21 (0·02)				0·20 (0·02)		

Age is measured about an origin of 12·25 years
Height is measured in cm
Model A fits random coefficients at level 2 up to and including the cubic term, and model B fits random coefficients up to and including the quadratic term

The first model fitted ignores the presence of autocorrelation. A fourth degree polynomial is fitted to describe the mean growth and coefficients up to the cubic are assumed to vary across subjects, that is, at level 2. This is necessary in principle to allow for the fact that the maximum velocity in height growth is expected to occur in the age range of the data, at different ages for different boys. Other studies suggest that a quartic curve adequately describes the mean growth over this age range.[13] Table I shows the results of fitting this model and the simpler model in which only coefficients up to the quadratic are random. The omission of the random cubic coefficient does not appreciably alter the fixed coefficient estimates or the other random parameters substantially.

Table II shows the results of fitting a first-order autoregressive model at level 1. For this purpose the observations are assumed equally spaced in time, although the actual ages are still used to describe the polynomial structure.

For the model with coefficients up to the quadratic random the fixed part of the model remains substantially unchanged. Making the cubic coefficient random decreases the estimate of the autocorrelation between successive observations somewhat, but the likelihood ratio chi-squared test statistic for the random cubic coefficient parameters is 7·0 with 4 degrees of freedom ($P = 0{\cdot}14$). We shall retain the model with coefficients up to the quadratic random in the remaining analyses.

If we fit a second-order model, the fixed part estimates and the random parameters at level 2 for analysis B in Table II remain virtually unchanged and the estimates (standard error) for σ_e^2, ρ_1

Table II. Height as a fourth degree polynomial on age with autoregressive level 1 residuals

Parameter	Estimate (SE) (A)	Estimate (SE) (B)
Fixed		
Intercept	149·0	149·0
age	6·18 (0·41)	6·19 (0·35)
age^2	1·20 (0·36)	1·23 (0·37)
age^3	0·44 (0·19)	0·43 (0·18)
age^4	− 0·45 (0·31)	− 0·49 (0·32)

Random: level 2 covariance matrix with standard errors in brackets

	Intercept	age	age^2	age^3	Intercept	age	age^2
Intercept	61·5 (17.1)				61·5 (17.1)		
age	9·3 (3·7)	3·9 (1·2)			7·9 (3·0)	2·7 (0·8)	
age^2	1·4 (1·4)	1·0 (0·4)	0·6 (0·2)		1·5 (1·4)	0·9 (0·3)	0·6 (0·2)
age^3	− 1·8 (1·5)	− 0·8 (0·5)	− 0·1 (0·2)	0·1 (0·3)			
Random: level 1							
σ_e^2		0·24 (0·05)				0·26 (0·04)	
ρ_1		0·16 (0·15)				0·23 (0·14)	

Age is measured about an origin of 12·25 years
Height is measured in cm
Equally spaced observations are assumed for the level1 autoregressive component
Model A fits random coefficients at level 2 up to and including the cubic term, and model B fits random coefficients up to and including the quadratic term

and ρ_2 are, respectively, 0·21 (0·04), 0·06 (0·19) and − 0·16 (0·12). The correlogram for the autocorrelation is a function of the roots of the 'auxiliary' or characteristic equation for the model (see for example, Fuller,[14] Chapter 2) namely

$$m^2 - \rho_1 m - \rho_2 = 0. \tag{11}$$

In our example the roots are $0 \cdot 03 \pm 0 \cdot 40i$. The existence of complex roots implies a declining cyclical function with a period of

$$2\pi(\cos^{-1}\theta)^{-1}, \quad \theta = 0{\cdot}5\rho_1(-\rho_2)^{-0{\cdot}5};$$

here an apparent period of 1·04 years, suggesting a seasonal effect (see below). The predicted autocorrelations for periods up to a year apart are given in Table III, with coefficients up to the quadratic random.

For the second-order model the declining cyclical pattern is clear. The predicted autocorrelations are rather different for the two models, but it should be noted that the estimated standard errors for the second-order model are relatively large. In fact a wide range of values for the correlations is compatible with a period of about one year.

Table III. Predicted autocorrelations for model (B) in Table II and second-order autoregressive model

Period (yr)	Model (B)	Second-order model
0·25	0·23	0·05
0·50	0·05	− 0·15
0·75	0·01	− 0·02
1·0	0·003	0·02

It is assumed that the measurements are spaced exactly 0·25 years apart

Table IV. Height as a fourth degree polynomial on age with a seasonal component

Parameter	Estimate (SE) (A)	Estimate (SE) (B)
Fixed		
Intercept	148·9	148·9
age	6·19 (0·35)	6·19 (0·35)
age^2	2·16 (0·45)	2·15 (0·45)
age^3	0·39 (0·17)	0·39 (0·17)
age^4	− 1·54 (0·43)	− 1·54 (0·43)
cos(t)	− 0·23 (0·07)	− 0·23 (0·07)

Random: level 2 covariance matrix with standard errors in brackets

	Intercept	age	age^2	Intercept	age	age^2
Intercept	61·5 (17·1)			61·5 (17·1)		
age	7·9 (3·0)	2·7 (0·8)		7·9 (3·0)	2·7 (0·8)	
age^2	1·5 (1·4)	0·9 (0·3)	0·6 (0·2)	1·5 (1·4)	0·9 (0·3)	0·6 (0·2)
Random: level 1						
σ_e^2	0·24 (0·04)			0·23 (0·05)		
ρ_1	0·23 (0·14)			0·20 (0·22)		
ρ_2	–			− 0·03 (0·13)		

Age is measured about an origin of 12·25 years
Height is measured in cm. Model A is AR(1). Model (B) is AR(2)

6.1. Seasonal effects

The existence of a small seasonal effect in height growth is well established.[15] Assuming a one year period we can fit a seasonal component directly to the data since the times of measurement are known.

With time measured from the start of the calendar year, let α be the amplitude and γ the phase with a seasonal component given by

$$\alpha \cos(t + \gamma) = \alpha_1 \cos(t) - \alpha_2 \sin(t). \tag{12}$$

Columns A and B in Table IV show the results of fitting this model. The estimate of α_2 is very small and has been set to zero. The second-order model now has a very small estimate for ρ_2. The

Table V. Height as a fourth degree polynomial on age with an autoregressive model in continuous time

Parameter	Estimate (SE) (A)	Estimate (SE) (B)
Fixed		
Intercept	148·9	148·9
age	6·19 (0·35)	6·19 (0·35)
age^2	2·16 (0·45)	2·16 (0·45)
age^3	0·39 (0·17)	0·39 (0·18)
age^4	− 1·55 (0·43)	− 1·54 (0·46)
cos(t)	− 0.24 (0·07)	− 0·23 (0·06)

Random: level 2 covariance matrix with standard errors in brackets

	Intercept	age	age^2	Intercept	age	age^2
Intercept	61·5 (17·1)			61·4 (17·1)		
age	7·9 (3·0)	2·7 (0·8)		7·9 (3·0)	2·6 (0·8)	
age^2	1·5 (1·4)	0·9 (0·3)	0·6 (0·2)	1·5 (1·4)	0·8 (0·3)	0·5 (0·2)
Random: level 1						
σ_e^2	0·23 (0·04)			0·28 (0·11)		
$\alpha_1(s)$	6·90 (2·07)			3·61 (1·70)		
$\alpha_2(s^{-2})$	–			0·013 (0·011)		

Age is measured about an origin of 12·25 years
Height is measured in cm.
(A) fits model of equation (10); (B) fits model of equation (12)

seasonal component itself indicates that the winter is when growth is slowest, with an average height difference about an individual growth curve between summer and winter of about 0·5 cm.

The seasonal coefficient might vary further across individuals, some showing more seasonal variation than others. When this coefficient is allowed to vary randomly at level 2, however, its variance is estimated as zero. A considerably larger sample of individuals would be needed to provide an accurate estimate of this variance.

6.2. Continuous time models

The discrete time models have assumed that the children are measured exactly 0·25 years apart. This is not precisely true, however, and there is some variation with a maximum discrepancy of 0·06 years (22 days). The results of fitting the models (9), (10) together with the seasonal component are given in columns A and B of Table V.

The model (10) using s^{-1} failed to converge. This appears to be because s^{-1} does not increase fast enough as s decreases and hence g does not decrease fast enough. To overcome this a model using s^{-2} has been fitted. As with the discrete time model the second parameter (α_2) is not significant.

For comparison with the discrete time model, the serial correlation in model A for two measurements 0·25 years apart is calculated from the estimated parameters to be 0·19. This is somewhat less than the value of 0·23 obtained from model A in Table IV since the correlation is

an exponentially decreasing function of the time interval in the model of Table V where the intervals differing from 0·25 years are modelled at their actual values. The values of the other parameters are virtually unchanged.

7. DISCUSSION

The formulation of the repeated measures model as a two-level random coefficients model[2] was important in that it allowed efficient and flexible modelling in continuous time with unbalanced data. That model, however, assumed independent level 1 residuals. The present paper extends the model by showing how flexible dependency structures can be modelled, also in continuous time.

Once dependent level 1 residual terms are permitted, the choice of model for representing repeated measurement data needs careful consideration. One source of autocorrelation is simple misfit in the rest of the model, such as fitting a straight line to quadratic response data. While it is possible to model this situation using the methods of this paper, it would seem preferable to incorporate as much as possible of an individual-specific structure before introducing an autocorrelation structure into the level 1 residuals. This is important for interpreting parameter values and characterizing individuals, for example by posterior estimates of their polynomial coefficients, average growth rates etc. On the other hand, experience suggests that in circumstances where serial correlations are high because measurement occasions are closely spaced and the data are unbalanced, it is often not possible to fit a repeated measures polynomial model with high-order coefficients varying randomly, *unless* an autocorrelation structure is allowed for.

In principle we can write any time series model in terms of the autocovariances for every possible lag, at least in discrete time. With moving average models or mixed models these autocovariances involve non-linear functions of the underlying parameters and extensions to such models of the procedures described in the present paper are currently being studied.

While the discrete time autoregressive model may be useful in certain applications, in general the continuous time model has considerably more flexibility. Many data sets do not have equally spaced occasion intervals, and even where equal intervals are part of the study design, it is often difficult to achieve this. Furthermore, the level 1 variance and the autocorrelation structure itself can be modelled as functions of further variables at level 1 or level 2 so that in particular the variance and the autocorrelation can be made to depend on time or age. One difficulty is that for moderate size data sets the likelihood surface can be rather flat and this may lead to numerical convergence problems. Care is needed also in choosing starting values. In addition, research needs to be carried out to determine suitable forms for $g(\alpha)$ in a range of applications.

The two-level model can also be extended to incorporate further levels of nesting in a straightforward fashion. For example, individual students measured repeatedly may be grouped into schools. We can also consider time series structures among units at level 2 or above. For example, we may have educational examination data where successive cohorts of a given age are measured within schools, leading to a repeated measures structure at level 2 of a three level hierarchy, with schools at level 3. In such a case the autocorrelation model will apply to the level 2 structure with the standard, independence, assumption at level 1.

Multivariate response data can be incorporated by specifying a further level below that of occasion (see Goldstein,[1] Chapter 5) so that multivariate time series models can be formulated and parameters estimated by straightforward extensions of the procedures described here. In this case model (9) can be extended for the cross-variable correlations by writing

$$g = \sigma_1\sigma_2 \exp(-\alpha_{12}s).$$

In some circumstances, non-linear growth curve models are appropriate, for example where the growth curve approaches an upper asymptote. Goldstein[16] discusses such models and these can be extended readily to incorporate the time series structures described in the present paper.

Finally, the procedures described here provide a general framework for efficient estimation from collections of short time series.[17] If we assume that each short series is sampled randomly from a population of such series, then the methods of this paper can be used.

APPENDIX I: ESTIMATION IN THE MULTILEVEL LINEAR MODEL

Conditional on V the generalized least squares estimator of β is given by

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y.$$

An initial estimate of V based on an ordinary least squares analysis is used.

Conditional on $\hat{\beta}$ we obtain an updated estimate of V as follows. The estimation of the variances and covariances, which are referred to as 'random parameters', and which define the elements of V, involves calculating $A = \tilde{Y}\tilde{Y}^T$ where the residual vector $\tilde{Y} = Y - X\hat{\beta}$. The vector vech($A$) is formed by stacking the columns of the lower triangle of the symmetric matrix A under one another. This is treated as a response vector and is regressed, using generalized least squares, upon the columns of a matrix X^{**}, which is derived from X and plays the role of a design matrix for the random parameters (Goldstein,[9] Appendix 1). In the variance components model each block of X^{**} corresponding to a level 2 unit has two columns vech(I) and vech(J) corresponding to σ^2_{e1} and σ^2_{e2}, respectively, where J is a matrix of ones.

APPENDIX II: ESTIMATION FOR CONTINUOUS TIME MODELS

To obtain an updating formula for the parameters, we expand the right hand side of (8) as a Taylor series about the value $\alpha = \alpha_H$ at the current iteration, so that ignoring higher order terms we have

$$f(\alpha, s) = f_H(\alpha, s) + \sum_{k=1}^{p} (\alpha_k - \alpha_{k_H}) \left(\frac{\partial f}{\partial \alpha_k}\right)_H$$

$$= \left\{1 + \sum_{k=1}^{p} \alpha_{k_H} \left(\frac{\partial g}{\partial \alpha_k}\right)_H\right\} f_H - \sum_{k=1}^{p} \alpha_k \left(\frac{\partial g}{\partial \alpha_k}\right)_H f_H$$

where α_{k_H} is the current value of α_k.

Together with (8) this expresses the elements of each V_{1j} in the linear form

$$\sigma_e^2 T^{(0)} + \Sigma_k \alpha_k T^{(k)}$$

where we set

$$T^{(0)} = \left\{1 + \sum_{k=1}^{p} \alpha_{k_H} \left(\frac{\partial g}{\partial \alpha_k}\right)_H\right\} f_H$$

$$T^{(k)} = \sigma^2_{e_H} \left(\frac{\partial g}{\partial \alpha_k}\right)_H f_H.$$

This provides an updating procedure for estimating σ_e^2, $\alpha_1, \cdots, \alpha_p$ similar to that described in Section 3, and at convergence we obtain the maximum likelihood estimates.

When $g(\alpha, s) = \alpha s$ we can estimate α by a direct generalization of the discrete time model. Consider the first-order model given by (5). The (kl)th element of R_{n_j} now becomes

$$R_{(n_j)}(k, l) = \begin{cases} 0 & \text{if } k = l \\ \rho^{(s-1)} & \text{if } k \neq l \end{cases}$$

where s is the time interval between occasions l and k, and $\rho = \exp(-\alpha)$. Given a suitable starting value, for example using procedures similar to those of Section 3, estimation proceeds as for the discrete time model. In addition, of course, we can fit the full level 2 structure as in the discrete time case.

ACKNOWLEDGEMENTS

We are most grateful to Professor Don Burrill, Professor Eric Chi, Dr. M. Yang, Dr. Stephen du Toit and anonymous referees for their most helpful comments. This work was carried out as part of the Multilevel Models Project funded by the Economic and Social Research Council (U.K.)

REFERENCES

1. Goldstein, H. *Multilevel Models in Educational and Social Research*, Griffin, London; Oxford University Press, New York, 1987.
2. Laird, N. M., and Ware, J. H. 'Random effects models for longitudinal data', *Biometrics.*, **38**, 963–974 (1982).
3. Grizzle, J. E. and Allen, D. M. 'Analysis of growth and dose response curves'. *Biometrics*, **63**, 219–229 (1969).
4. Chi, E. M. and Reinsel, G. C. 'Models for longitudinal data with random effects and AR(1) errors', *Journal of the American Statistical Association*, **84**, 452–459 (1989).
5. Lee, J. C. 'Tests and model selection for the general growth cruve model', *Biometrics*, **47**, 147–159 (1991).
6. Diggle, P. 'An approach to the analysis of repeated measurements', *Biometrics*, **44**, 959–971 (1988).
7. Jones, R. H. and Boadi-Boateng, F. 'Unequally spaced longitudinal data with AR(1) serial correlation', *Biometrics*, **47**, 161–175 (1991).
8. Jennrich, R. I. and Schluchter, M. D. 'Unbalanced repeated measures models with structured covariance matrices', *Biometrics.* **42**, 805–820 (1986).
9. Goldstein, H. 'Multilevel mixed linear model analysis using iterative generalised least squares', *Biometrika*, **73**, 43–56 (1986).
10. Goldstein, H. 'Restricted unbiased iterative generalised least squares estimation', *Biometrika*, **76**, 622–623 (1989).
11. Harrison, G. A. and Brush, G. 'On correlations between adjacent velocities and accelerations in longitudinal growth data', *Annals of Human Biology*, **17**, 55–57 (1990).
12. Prosser, R. Rasbash, J. and Goldstein, H. *ML3: Software for Three-level Analysis. Users Guide*, Institute of Education, University of London.
13. Goldstein, H. 'Models for multilevel response variables with an application to growth curves', in Bock, R. D. (ed.) *Multilevel Analysis of Educational Data*, Academic Press, New York, 1989.
14. Fuller, W. A. *Introduction to Statistical Time Series*, Wiley, New York, 1976.
15. Tanner, J. M. *Foetus into Man*, Castlemead Publications, Ware, 1989.
16. Goldstein, H. 'Nonlinear multilevel models, with an application to discrete response data', *Biometrika*, **78**, 45–51 (1991).
17. Azzalini, A. 'Estimation and hypothesis testing for collections of autoregressive time series', *Biometrika*, **71**, 85–90 (1984).