Biometrika (1990), 77, 4, *pp.* 721-31
Printed in Great Britain

Serial correlation in unequally spaced longitudinal data

BY RICHARD H. JONES

Department of Preventive Medicine and Biometrics, School of Medicine, University of Colorado Health Sciences Center, Denver, Colorado 80262, U.S.A.

AND LYNN M. ACKERSON

Department of Biometrics, National Jewish Center for Immunology and Respiratory Medicine, Denver, Colorado 80206, U.S.A.

SUMMARY

Serial correlation in the within subject error structure in longitudinal data with unequally spaced observations is modelled using continuous time autoregressive moving averages. The models considered have both fixed and random effects in addition to serially correlated within subject errors. Two approaches are presented for calculating the exact likelihood for a model when the errors are Gaussian. The first calculates the covariance matrices for each subject for assumed values of the unknown parameters and estimates the fixed parameters by weighted least squares. The second uses a state space model and the Kalman filter to calculate the exact likelihood. Both methods involve the use of complex arithmetic. Nonlinear optimization is used to obtain maximum likelihood estimates of the parameters.

Some key words: Continuous time ARMA model; Exact likelihood; Growth curve; Kalman filter; Longitudinal data; Serial correlation; State space representation; Unequally spaced observations.

1. INTRODUCTION

In longitudinal data experiments, one or more groups of subjects are followed over time. In many experiments, the observations are equally spaced. In this case, first order autoregressions, AR(1), have long been a popular model for the within subject error structure (Potthoff & Roy, 1964; Chi & Reinsel, 1989). The problem of missing observations is not difficult to handle if the data are equally spaced (Jones, 1980, 1985; Rosner & Muñoz, 1988). Extensions to discrete time autoregressive-moving average, ARMA, processes have also been developed (Rochon & Helms, 1989).

When there is no basic sampling interval, it is necessary to use continuous time series models for serial correlation. A continuous time AR(1) model poses no real difficulties. It is similar to the discrete case where the correlation function decays exponentially as a function of the time interval (Diggle, 1988). However, for higher order continuous time ARMA processes, the computations are more difficult. It is necessary to fit linear stochastic differential equations to the error structure. This involves integration in the complex plane to calculate the covariance function using the method of residuals to invert a Fourier transform. Using a state space approach, it is necessary to evaluate matrix exponentials. Using diagonalization methods, this requires complex arithmetic since the matrices are not symmetric.

This paper first presents the methods for fitting a linear mixed model which is commonly used in longitudinal data and growth curve studies using a continuous time AR(1) structure for the within subject errors. The extensions to continuous time ARMA models are then developed. Two methods are presented for calculating the exact likelihood assuming Gaussian errors. The first is the direct evaluation of the total error structure for each subject as a function of the ARMA coefficients. The second is a state space approach using the Kalman filter to calculate the exact likelihood. The two methods are equivalent, but the two perspectives provide different insights into the problem and allow the checking of computer algorithms.

2. THE MODEL

The linear mixed model for longitudinal data considered here has been proposed by Laird & Ware (1982) and others,

$$y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i, \tag{1}$$

where y_i is an $n_i \times 1$ column vector of the response variable for subject i, X_i is an $n_i \times b$ design matrix, β is a $b \times 1$ vector of regression coefficients assumed to be fixed, and Z_i is an $n_i \times g$ design matrix for the random effects, γ_i, which are assumed to be independently distributed across subjects with distribution, $N(0, \sigma^2 B)$, where B, for between, is a general covariance matrix. The within subject errors, ε_i, are assumed to be distributed $N(0, \sigma^2 W_i)$, where W_i, for within, is the covariance matrix of a continuous time ARMA process with a scale factor, σ^2. The ε_i are also independently distributed from subject to subject and independent of the γ_i. This model is fairly general since different subjects can have different numbers of observations and the observation times can be different. This is indicated by the subscript i in the vector y_i and the matrices X_i, Z_i and W_i. It is not necessary for the matrices X_i and Z_i to be of full rank.

As is well known, because of the independence of the errors across subjects, and the normality assumptions, minus twice the log likelihood for this model is

$$l = \sum_i \left\{ n_i \log(2\pi\sigma^2) + \log|Z_iBZ_i' + W_i| + \frac{1}{\sigma^2}(y_i - X_i\beta)'(Z_iBZ_i' + W_i)^{-1}(y_i - X_i\beta) \right\}, \tag{2}$$

where the summation is across subjects. This likelihood is a function of β, σ^2, the parameters of the continuous time ARMA process that parameterize the matrices W_i and the matrix B. It is necessary to use nonlinear optimization programs to obtain maximum likelihood estimates of these parameters. However, as is also well known, β and σ^2 can be concentrated out of the likelihood function. For a given matrix B and given values of the ARMA parameters, the estimate of the fixed effects coefficient vector is

$$\hat{\beta} = \{\textstyle\sum_i X_i'(Z_iBZ_i' + W_i)^{-1}X_i\}^{-1}\{\sum_i X_i'(Z_iBZ_i' + W_i)^{-1}y_i\}, \tag{3}$$

and the estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i (y_i - X_i\hat{\beta})'(Z_iBZ_i' + W_i)^{-1}(y_i - X_i\hat{\beta}),$$

where $n = \Sigma_i n_i$ is the total number of observations from all the subjects. Substituting these estimates back into (2) gives l concentrated with respect to σ^2 and β,

$$l = n\log(2\pi\hat{\sigma}^2) + \textstyle\sum_i \log|(Z_iBZ_i' + W_i)| + n.$$

To obtain maximum likelihood estimates, this must be minimized with respect to the ARMA coefficients that parameterize the W_i matrices and the elements of the B matrix. Constants have been retained in order to compare likelihoods with some published results. To ensure that the estimated B matrix is nonnegative definite, optimization is carried out with respect to its square root. Let $B = U'U$, where U is upper triangular. Now

$$l = n \log (2\pi\hat{\sigma}^2) + \sum_i \log |(Z_i U' U Z_i' + W_i)| + n, \tag{4}$$

and the optimization is carried out with respect to the ARMA coefficients and the upper triangular elements of U.

3. Continuous time ARMA within subject errors

Doob (1953, Ch. 11, § 10, p. 542) discusses continuous time models. A continuous time ARMA(p, q) process is defined by the stochastic linear differential equation

$$\frac{d^p x(t)}{dt^p} + \alpha_{p-1} \frac{d^{p-1} x(t)}{dt^{p-1}} + \ldots + \alpha_0 x(t) = \delta_q \frac{d^q \eta(t)}{dt^q} + \delta_{q-1} \frac{d^{q-1} \eta(t)}{dt^{q-1}} + \ldots + \eta(t).$$

Here $\eta(t)$ is continuous time 'white noise' which exists only in the sense that its integral is a Wiener process with zero mean and variance σ_η^2 per unit time. For stationarity, it is necessary that $q < p$ and that the roots of

$$A(z) = \sum_{k=0}^{p} \alpha_k z^k = 0 \tag{5}$$

have negative real parts, where $\alpha_p = 1$. Equation (5) is referred to as the characteristic equation of the system. The system is 'minimum phase' or 'minimum delay' if the roots of $\Sigma\, \delta_k z^k = 0$, where the sum is over $k = 0, \ldots, q$, have nonpositive real parts, where $\delta_0 = 1$.

For a stationary continuous time ARMA process, the spectral density is

$$s(f) = \sigma_\eta^2 \frac{\left| \sum_{k=0}^{q} \delta_k (2\pi i f)^k \right|^2}{\left| \sum_{k=0}^{p} \alpha_k (2\pi i f)^k \right|^2},$$

which is referred to as a rational spectrum (Doob, 1953). In this equation, f denotes frequency in cycles per unit time. The range of f is $-\infty < f < \infty$. The covariance function of the process is

$$R(\tau) = \int_{-\infty}^{\infty} s(f) \exp (2\pi i \tau f)\, df,$$

where τ is the time lag. This integral can be evaluated by the usual method of inverting a Fourier transform by integration in the complex plane using the method of residuals (Doob, 1953). Let the roots of the characteristic equation (5) be $r_1, r_2, \ldots, r_p$, and assume that these roots are distinct. This assumption of distinct roots will also be used later in the numerical solution using the state space approach, and is not as restrictive as it may seem. When doing a numerical search, the probability of equal roots is very near zero unless an initial guess at the unknown parameters is used which produces equal roots.

In this case the program will detect the equal roots and request new starting values. The covariance function is

$$R(\tau) = \sigma_\eta^2 \sum_{k=1}^{p} \frac{(\Sigma_l \delta_l r_k^l)\{\Sigma_l \delta_l (-r_k)^l\} \exp(r_k \tau)}{-2 \operatorname{Re}(r_k) \Pi_{l \neq k} (r_l - r_k)(r_l^* + r_k)}, \tag{6}$$

where the sum Σ_l is over $l = 0, \ldots, q$, and where the product in the denominator is from $l = 1$ to p excluding $l = k$, r_k^* denotes the complex conjugate of the root, and $\operatorname{Re}(r_k)$ denotes the real part of r_k.

For an ARMA(1, 0), that is an AR(1), process, the product term in the denominator is replaced by 1 and the covariance function is $R(\tau) = -\frac{1}{2}\sigma_\eta^2 e^{r_1 \tau}/r_1$, since the single root must be real and negative. This is the well known covariance function for a continuous time AR(1) process which decays exponentially as does the discrete time AR(1) covariance function; however, a discrete time AR(1) process can have a negative autoregression coefficient causing the covariance to alternate positive and negative while decaying exponentially. This cannot happen in continuous time. The continuous time coefficient, α_0 must be positive producing a characteristic equation $z + \alpha_0 = 0$ with root $r_1 = -\alpha_0$, so that the covariance function is positive.

The ARMA(p, q) error structure can be generalized slightly by allowing the addition of observational error. Letting this observational error variance be σ_0^2, this is added to the diagonal of the within subject error covariance matrix. Observational error is necessary in the model when a continuous time ARMA process is used if there are replicate measurements for a subject at a given time. Without observational error, replicate measurements would necessarily be identical since a prediction for a time interval of zero would have zero variance.

4. THE DIRECT METHOD OF CALCULATING THE LIKELIHOOD

Jennrich & Schluchter (1986) discuss a direct method for obtaining maximum likelihood estimates in unbalanced repeated-measures models with structured covariance matrices. A data set from Potthoff & Roy (1964) is used as an example. This method can be used when the maximum number of observations for one subject is not too large.

To avoid solving for the roots of the characteristic polynomial and to constrain the ARMA model to be stationary, the autoregression coefficients can be reparameterized following Jones (1981). The characteristic polynomial (5) is written as quadratic factors with a linear factor if p is odd,

$$A(z) = (a_1 + a_2 z + z^2)(a_3 + a_4 z + z^2) \ldots \begin{cases} (a_{p-1} + a_p z + z^2) & (p \text{ even}), \\ (a_p + z) & (p \text{ odd}). \end{cases} \tag{7}$$

The roots will have negative real parts if and only if the a's are positive. A log transformation is used to constrain the a's to be positive, and the optimization is carried out with respect to the logs of the a's. From the a's the roots of the characteristic polynomial are easily calculated from the quadratic and linear factors.

For subject i, W_i is calculated using (6) and normalized by dividing by $R(0)$. If observational error is included in the model, σ_0 is used as one of the nonlinear parameters to be estimated and σ_0^2 is added to the diagonal elements of W_i. This constrains σ_0^2 to be non-negative in the estimation procedure. From the elements of U, $Z_i U'$ is calculated, and the total covariance matrix divided by σ^2 for the subject calculated as

$$Z_i B Z_i' + W_i = (Z_i U')(Z_i U')' + W_i.$$

This matrix is augmented by X_i and y_i and a Cholesky factorization carried out on the augmented matrix (Graybill, 1976, § 7.2). The augmented matrix has dimensions $n_i \times (n_i + b + 1)$, where b is the number of fixed coefficients to be estimated, the dimension of β. The Cholesky factorization replaces the augmented matrix

$$[Z_i B Z_i' + W_i \;\; X_i \;\; y_i]$$

by

$$[T_i \;\; D_i \;\; b_i],$$

where T_i is the upper triangular factorization of $Z_i B Z_i' + W_i$, and

$$D_i = (T_i')^{-1} X_i, \quad b_i = (T_i')^{-1} y_i.$$

The determinant term in the likelihood (4) is calculated as the product of the diagonal elements of T_i squared. The weighted total sum of squares, TSS, for the fixed effects regression is $b_i' b_i$ summed over subjects. The matrix $D_i' D_i$ augmented by $D_i' b_i$ is summed over subjects producing the matrix

$$[\textstyle\sum_i D_i' D_i \;\; \sum_i D_i' b_i], \tag{8}$$

which has dimension $b \times (b+1)$. Applying the Cholesky factorization to this augmented matrix produces the matrix

$$[G \;\; r]. \tag{9}$$

The sum of squares of the elements of the vector r is the weighted regression sum of squares so

$$\hat{\sigma}^2 = \frac{1}{n}\left(\text{TSS} - \sum_{k=1}^{b} r_k^2\right).$$

This completes the calculation of (4). It is not necessary to actually calculate the estimated regression coefficients during the optimization, but they can be calculated from (9) by back substitution.

5. The state space method

Kalman (1960) developed a recursive optimal estimation procedure for state space models in discrete time. Kalman & Bucy (1961) extended this work to continuous time state space models. Schweppe (1965) showed that the Kalman filter can be used to calculate likelihoods when there are unknown parameters in the model. By integrating a continuous time state space representation over time intervals, a discrete time state space representation is developed at unequally spaced observation times. Harvey & Phillips (1979) used the Kalman filter for regression models with discrete time ARMA errors by including the regression coefficients in the state vector. Jones (1986) used a modified method that does not require increasing the length of the state vector.

A state space representation of a continuous time ARMA(p, q) process is (Wiberg, 1971)

$$\frac{d}{dt}\begin{bmatrix} s(t) \\ s^{(1)}(t) \\ \vdots \\ s^{(p-1)}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & \dots & -\alpha_{p-1} \end{bmatrix} \begin{bmatrix} s(t) \\ s^{(1)}(t) \\ \vdots \\ s^{(p-1)}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \eta(t), \tag{10}$$

and the observation equation for an observation at time t_j is

$$\varepsilon(t_j) = [1 \quad \delta_1 \quad \dots \quad \delta_{p-1}] \begin{bmatrix} s(t_j) \\ s^{(1)}(t_j) \\ \vdots \\ s^{(p-1)}(t_j) \end{bmatrix} + v(t_j),$$

where $v(t_j)$ is the observational error, if present, with variance $\sigma^2\sigma_0^2$. The kth derivative of $s(t)$ with respect to t is denoted by $s^{(k)}$, and $\delta_k = 0$ if $k > q$. In matrix notation, the state space representation is

$$ds(t)/dt = Fs(t) + g\eta(t), \quad \varepsilon(t_j) = hs(t_j) + v(t_j), \tag{11}$$

where h is a row vector specifying the linear combination of the elements of the state that make up the errors. These equations express the elements of the within subject error vector ε_i in the longitudinal data model (1) in terms of a state vector $s(t)$ which is a continuous time random process and $p-1$ of its derivatives.

Duncan & Horn (1972) showed that mixed models can be put in state space form by including the random effects in the state vector. In order to put equation (1) in state space form it is necessary to augment the state vector by the vector γ_i. The complete state space representation for subject i is

$$\frac{d}{dt}\begin{bmatrix} s_i(t) \\ \gamma_i \end{bmatrix} = \begin{bmatrix} F & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} s_i(t) \\ \gamma_i \end{bmatrix} + \begin{bmatrix} g \\ 0 \end{bmatrix} \eta_i(t), \tag{12}$$

$$\xi_i(t_j) = [h \;\; z_i(t_j)] \begin{bmatrix} s_i(t_j) \\ \gamma_i \end{bmatrix} + v_i(t_j). \tag{13}$$

Here $z_i(t_j)$ is a row vector consisting of row j in Z_i of (1). Equation (12) indicates that the random effects for subject i are constant over time. The fact that the random effects vary from subject to subject is introduced into the model through the initial state covariance matrix which will be set equal to $B = U'U$. Since the elements of this prior covariance matrix will be estimated by maximum likelihood, this is an empirical Bayes method.

Jones (1981) used the state space approach to calculate the likelihood for a continuous time stationary AR(p) time series with zero mean. This is the case shown in (10)-(11) without the δ's in the h vector. The method consists of integrating the state equation over the intervals between observations to obtain a discrete state space representation corresponding to the observation points. Since the δ's appear in the observation equation, this method applies to the present situation. Integrating the state equation, with the random input removed, over a time interval from t_{j-1} to t_j gives the well-known matrix exponential solution,

$$s(t_j) = \exp\{(t_j - t_{j-1})F\}s(t_{j-1}). \tag{14}$$

There are many possible ways to evaluate matrix exponentials numerically (Moler & Van Loan, 1978). Because of the structure of our F matrix, the eigenvalues are the roots of the characteristic equation (5), r_k, and assuming that these roots are distinct, the matrix F can be rotated to diagonal form. The right eigenvector corresponding to the root r_k has lth element $C_{lk} = r_k^{l-1}$. The diagonal form of F is $F = C\Lambda C^{-1}$, where C is a square matrix, the columns of which are the right eigenvectors of F, and Λ is a diagonal matrix

with diagonal elements r_k. These matrices may contain complex elements so complex arithmetic is necessary for these calculations. Equation (14) can be written

$$s(t_j) = C \exp\{(t_j - t_{j-1})\Lambda\} C^{-1} s(t_{j-1}).$$

Defining a complex rotated state vector $\tilde{s}(t) = C^{-1}s(t)$ produces an uncoupled system where each element of the rotated state vector is predicted independently of the other elements,

$$\tilde{s}_k(t_j) = \exp\{(t_j - t_{j-1})r_k\}\tilde{s}_k(t_{j-1}),$$

or, in matrix form

$$\tilde{s}(t_j) = \Phi(t_j - t_{j-1})\tilde{s}(t_{j-1}),$$

where

$$\Phi(t_j - t_{j-1}) = \exp\{(t_j - t_{j-1})\Lambda\}$$

is the diagonal, possibly complex, rotated state transition matrix. It is not necessary to actually rotate the state vector since for each subject the initial state vector is zero, so the initial rotated state vector is zero. The rotated state vector is predicted from time point to time point and need only be rotated back to obtain the actual observations using the inverse equation $s(t) = C\tilde{s}(t)$.

In addition to predicting the state equation over an arbitrary time interval, it is necessary to calculate the contribution of the error term to this prediction. The complex prediction error due to $\eta(t)$ of the rotated state vector is (Kalman & Bucy, 1961)

$$\int_{t_{j-1}}^{t_j} \Phi(t_j - t) C^{-1} g \eta(t)\, dt.$$

Let $\kappa = C^{-1}g$, which is the last column of C^{-1}. Then the complex prediction error is (Jones, 1981)

$$\int_{t_{j-1}}^{t_j} \begin{bmatrix} \kappa_1 \exp\{(t_j - t)r_1\} \\ \kappa_2 \exp\{(t_j - t)r_2\} \\ \vdots \\ \kappa_p \exp\{(t_j - t)r_p\} \end{bmatrix} \eta(t)\, dt,$$

with a Hermitian covariance matrix that has elements

$$\sigma^2 Q_{kl}(t_j - t_{j-1}) = -\sigma^2 \frac{\kappa_k \kappa_l^*}{(r_k + r_l^*)}[1 - \exp\{(r_k + r_l^*)(t_j - t_{j-1})\}].$$

This produces the one-step prediction covariance matrix $Q(t_j - t_{j-1})$ necessary for the Kalman filter recursion. The discrete time rotated state space representation corresponding to (12) is

$$\begin{bmatrix} \tilde{s}_i(t_j) \\ \gamma_i \end{bmatrix} = \begin{bmatrix} \Phi(t_j - t_{j-1}) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{s}_i(t_{j-1}) \\ \gamma_i \end{bmatrix} + \begin{bmatrix} \zeta(t_j) \\ 0 \end{bmatrix},$$

where $\zeta(t_j)$ is a zero mean complex Gaussian random vector with covariance matrix $\sigma^2 Q(t_j - t_{j-1})$ and independent at the observation times. The corresponding observation equation (13) becomes

$$\xi_i(t_j) = [hC \;\; z_i(t_j)] \begin{bmatrix} \tilde{s}_i(t_j) \\ \gamma_i \end{bmatrix} + v_i(t_j).$$

The initial state covariance matrix for a subject is the covariance matrix of the state vector before any observations are taken on that subject. For the upper part of the state vector, this is the steady state covariance matrix of the ARMA process and its $p-1$ derivatives (Doob, 1953; Jones, 1981), and is obtained by complex integration similar to the method of obtaining the covariance function

$$\sigma^2 P_{\mu\nu}(t_1|0) = -\sigma^2 \sum_{k=1}^{p} \frac{r_k^{\mu}(-r_k)^{\nu}}{2\,\mathrm{Re}\,(r_k)\prod_{l\neq k}(r_l - r_k)(r_l^* + r_k)}.$$

The notation for the state covariance matrix, $P(t_1|0)$, indicates the covariance matrix of the state vector at the first observation time before any observations for subject i. This can be thought of as the prior covariance matrix of the state vector in that it represents what is known about the state based only on the autoregression coefficients and σ^2 before observations on a subject are taken. The initial covariance matrix of the rotated state vector is

$$\tilde{P}_c(t_1|0) = C^{-1}P(t_1|0)(C^*)^{-1},$$

where C^* denotes the complex conjugate transposed matrix. This covariance matrix may be complex and is Hermitian. The initial state covariance matrix of the complete state vector in partitioned form is

$$\tilde{P}(t_1|0) = \sigma^2\begin{bmatrix} \tilde{P}_{11}(t_1|0) & \tilde{P}_{12}(t_1|0) \\ \tilde{P}_{12}^*(t_1|0) & P_{22}(t_1|0) \end{bmatrix} = \sigma^2\begin{bmatrix} \tilde{P}_c(t_1|0) & 0 \\ 0 & U'U \end{bmatrix}. \tag{15}$$

This completes the background necessary to develop the state space approach for calculating the exact likelihood.

The Kalman recursion is carried out in such a way that the matrix and vector of the normal equations needed to estimate β, (3), the weighted total sum of squares needed to estimate σ^2, and the determinant term in the likelihood (4) are accumulated during a pass through the data. This requires that the Kalman filter be run separately on each column of the X_i matrices as well as the y_i vectors, each column being treated as a data vector. The parts of the recursion involving the covariance matrices need only be calculated once.

Since there is a state vector for each column of X_i and y_i, the state is now a $(p+g)\times(b+1)$ matrix. This matrix of transformed state vectors will be denoted

$$\tilde{S}(t) = \begin{bmatrix} \tilde{S}_c(t) \\ S_r(t) \end{bmatrix}.$$

Here, $\tilde{S}_c(t)$ denotes the first part of the state corresponding to the ARMA process and the subscript c indicates that this is complex. The subscript of the second part of the state vector, which is the part corresponding to the random effects is r, which also indicates that this part of the state is real. The subscript i denoting the subject has been suppressed. The state matrix is initialized to zero for each subject with its initial covariance matrix set to $\tilde{P}(t_1|0)$, (15).

(i) The state matrix is predicted to the next observation time,

$$\begin{bmatrix} \tilde{S}_c(t_j|t_{j-1}) \\ S_r(t_j|t_{j-1}) \end{bmatrix} = \begin{bmatrix} \Phi(t_j - t_{j-1}) & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} \tilde{S}_c(t_{j-1}|t_{j-1}) \\ S_r(t_{j-1}|t_{j-1}) \end{bmatrix}.$$

The state transition matrix in this equation is diagonal.

(ii) The covariance matrix of this prediction is

$$\tilde{P}_{11}(t_j|t_{j-1}) = \Phi(t_j - t_{j-1})\tilde{P}_{11}(t_{j-1}|t_{j-1})\Phi^*(t_j - t_{j-1}) + Q(t_j - t_{j-1}),$$

$$\tilde{P}_{12}(t_j|t_{j-1}) = \Phi(t_j - t_{j-1})\tilde{P}_{12}(t_{j-1}|t_{j-1}), \quad P_{22}(t_j|t_{j-1}) = P_{22}(t_{j-1}|t_{j-1}).$$

These first two steps are skipped at the first observation for each subject. P_{11} and P_{12} may be complex, but P_{22} will remain real.

(iii) The innovation, a real row vector of length $b+1$, at time t_j for subject i is

$$I_i(t_j) = [x_i(t_j) \;\; y_i(t_j)] - [hC \;\; z_i(t_j)]\begin{bmatrix}\tilde{S}_c(t_j|t_{j-1})\\ S_r(t_j|t_{j-1})\end{bmatrix},$$

where $x_i(t_j)$ is a row vector consisting of row j in X_i of (1), and $y_i(t_j)$ is the corresponding element of y_i.

(iv) The innovation variance is a scalar,

$$V_i(t_j) = (hC)\tilde{P}_{11}(t_j|t_{j-1})(hC)^* + 2\,\mathrm{Re}\,\{(hC)\tilde{P}_{12}(t_j|t_{j-1})z_i'(t_j)\} + z_i(t_j)P_{22}(t_j|t_{j-1})z_i'(t_j) + \sigma_0^2. \tag{16}$$

(v) The upper triangular part of a $(b+1)\times(b+1)$ matrix necessary for calculating the estimates of β and σ^2, and the determinant term are accumulated over all the observations for all the subjects,

$$M \leftarrow M + I_i(t_j)I_i'(t_j)/V_i(t_j), \quad \Delta \leftarrow \Delta + \log\{V_i(t_j)\},$$

where $\leftarrow$ indicates that the left-hand side is replaced by the right-hand side of the expression.

(vi) The Kalman gain is

$$K_i(t_j) = \frac{1}{V_i(t_j)}\begin{bmatrix}\tilde{P}_{11}(t_j|t_{j-1})(hC)^* + \tilde{P}_{12}(t_j|t_{j-1})z_i'(t_j)\\ \{(hC)\tilde{P}_{12}(t_j|t_{j-1})\}^* + P_{22}(t_j|t_{j-1})z_i'(t_j)\end{bmatrix}.$$

(vii) The updated estimate of the state is

$$\begin{bmatrix}\tilde{S}_c(t_j|t_j)\\ S_r(t_j|t_j)\end{bmatrix} = \begin{bmatrix}\tilde{S}_c(t_j|t_{j-1})\\ S_r(t_j|t_{j-1})\end{bmatrix} + K_i(t_j)I_i(t_j).$$

(viii) The updated state covariance matrix is

$$\begin{bmatrix}\tilde{P}_{11}(t_j|t_j) & \tilde{P}_{12}(t_j|t_j)\\ \tilde{P}_{12}^*(t_j|t_j) & P_{22}(t_j|t_j)\end{bmatrix} = \begin{bmatrix}\tilde{P}_{11}(t_j|t_{j-1}) & \tilde{P}_{12}(t_j|t_{j-1})\\ \tilde{P}_{12}^*(t_j|t_{j-1}) & P_{22}(t_j|t_{j-1})\end{bmatrix} - V_i(t_j)K_i(t_j)K_i^*(t_j).$$

This completes the recursion. Because of the subtraction in this equation, the recursion for the covariance matrix would usually be carried out in double precision.

The state and state covariance matrix are reinitialized for each subject, M and Δ are accumulated over all the observations for all the subjects, and l is evaluated at the end. The matrix M is the same matrix generated by the direct method in equation (8) with the total sum of squares, TSS, in the lower right-hand position, and l is calculated from this point as in the direct method.

6. AN EXAMPLE

In the example used by Jennrich & Schluchter (1986), there are four observations per subject and they are equally spaced. Eight different models for the fixed effects and error structure are considered. All of these models are special cases of the model we have developed that are designed for unequally spaced observations but also apply to equally spaced observations. The 'unstructured' or general covariance structure is obtained by letting $Z_i = I$, in this case a 4×4 identity matrix, and $W_i = 0$. Their model 5 is AR(1) with no random effects, and model 4 is called 'banded', i.e. the total 4×4 error covariance matrix for each subject is assumed to have a general Toeplitz structure, constant along each diagonal. If the observations are unequally spaced, the covariance matrix produced by a time series model would no longer be Toeplitz, but the methods presented here can be used to calculate the covariance matrix as a function of the continuous time ARMA parameters. Our approach is to calculate the exact likelihood with appropriate constraints so that covariance matrices are nonnegative definite. Maximum likelihood estimates are obtained using numerical optimization, and a numerical approximation to the information matrix is calculated at the minimum of l.

We have duplicated the results for the 'banded' model 4 using a continuous time AR(3) model for the within subject error and no random effects. The estimates of the three nonlinear parameters in the model are

$$\log a_1 = 0{\cdot}977, \quad \log a_2 = -0{\cdot}899, \quad \log a_3 = -2{\cdot}542,$$

which gives the factorized form of the characteristic equation (7)

$$A(z) = (2{\cdot}656 + 0{\cdot}407z + z^2)(0{\cdot}0787 + z),$$

which has a complex conjugate pair of roots $-0{\cdot}2035 \pm 1{\cdot}617i$, and a real root $-0{\cdot}0787$. The mean square error which estimates σ^2 is 4·668. The estimated covariance function calculated from (6) for the time spacing of two years is

$$R(0) = 4{\cdot}947, \quad R(2) = 3{\cdot}054, \quad R(4) = 3{\cdot}411, \quad R(6) = 2{\cdot}370,$$

and the correlation function is

$$\rho(0) = 1{\cdot}000, \quad \rho(2) = 0{\cdot}617, \quad \rho(4) = 0{\cdot}690, \quad \rho(6) = 0{\cdot}479.$$

7. DISCUSSION

The direct method and the state space method both have their advantages and disadvantages in different situations. When there are many subjects with a small number of observations per subject, the direct method is both faster and numerically more accurate. When there are many observations per subject, the direct method is much slower than the state space method since it constructs and stores the total covariance matrix for each subject. Both methods have been programmed and give identical results.

In the example, the advantage of our method is that the estimates are constrained to force the fitted within subject error model to be stationary. If the within subject error covariance matrix is allowed to have a general banded or Toeplitz structure, the nonlinear optimization may search over values of the parameters where the structure is not only nonstationary, but the covariance matrix may have negative roots. This can cause the optimization procedure to diverge or to converge to values of the parameters outside the parameter space.

ACKNOWLEDGEMENTS

This research was supported by a U.S. Public Health Service research grant from the National Institute of General Medical Sciences and was written while the first author was visiting the Department of Statistics, Faculty of Economics and Commerce, Australian National University.

REFERENCES

CHI, E. M. & REINSEL, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *J. Am. Statist. Assoc.* **84**, 452-9.

DIGGLE, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44**, 959-71.

DOOB, J. L. (1953). *Stochastic Processes.* New York: Wiley.

DUNCAN, D. B. & HORN, S. D. (1972). Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Am. Statist. Assoc.* **67**, 815-21.

GRAYBILL, F. A. (1976). *Theory and Application of the Linear Model.* North Scituate, Mass: Duxbury.

HARVEY, A. C. & PHILLIPS, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika* **66**, 49-58.

JENNRICH, R. I. & SCHLUCHTER, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805-20.

JONES, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22**, 389-95.

JONES, R. H. (1981). Fitting a continuous time autoregression to discrete data. In *Applied Time Series Analysis II*, Ed. D. F. Findley, pp. 651-82. New York: Academic Press.

JONES, R. H. (1985). Time series analysis with unequally spaced data. In *Handbook of Statistics*, **5**: *Time Series in the Time Domain*, Ed. E. J. Hannan, P. R. Krishnaiah and M. M. Rao, pp. 157-77. Amsterdam: North-Holland.

JONES, R. H. (1986). Time series regression with unequally spaced data. In *Essays in Time Series and Allied Processes. J. Appl. Prob.*, Special Volume **23A**, 89-98.

KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. Am. Soc. Mech. Eng. (J. Basic Eng.)* **82D**, 35-45.

KALMAN, R. E. & BUCY, R. S. (1961). New results in linear filtering and prediction theory. *Trans. Am. Soc. Mech. Eng. (J. Basic Eng)* **82D**, 95-108.

LAIRD, N. M. & WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-74.

MOLER C. & VAN LOAN, C. (1978). Nineteen dubious ways to compute the exponential of a matrix. *SIAM Rev.* **20**, 801-36.

POTTHOFF, R. F. & ROY, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313-26.

ROCHON, J. & HELMS, R. W. (1989). Maximum likelihood estimation for incomplete repeated-measures experiments under an ARMA covariance structure. *Biometrics* **45**, 207-18.

ROSNER, B. & MUÑOZ, A. (1988). Autoregressive modelling for the analysis of longitudinal data with unequally spaced examinations. *Statist. Med.* **7**, 59-71.

SCHWEPPE, F. C. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Trans. Info. Theory* **11**, 61-70.

WIBERG, D. M. (1971). *State Space and Linear Systems. Schaum's Outline Series.* New York: McGraw-Hill.

[*Received January* 1990. *Revised April* 1990]