Biometrika (1999), **86,** 3, *pp.* 677–690

Printed in Great Britain

Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation

By MOHSEN POURAHMADI

Division of Statistics, Northern Illinois University, DeKalb, Illinois 60115, U.S.A.

pourahm@math.niu.edu

Summary

We provide unconstrained parameterisation for and model a covariance using covariates. The Cholesky decomposition of the inverse of a covariance matrix is used to associate a unique unit lower triangular and a unique diagonal matrix with each covariance matrix. The entries of the lower triangular and the log of the diagonal matrix are unconstrained and have meaning as regression coefficients and prediction variances when regressing a measurement on its predecessors. An extended generalised linear model is introduced for joint modelling of the vectors of predictors for the mean and covariance subsuming the joint modelling strategy for mean and variance heterogeneity, Gabriel's antedependence models, Dempster's covariance selection models and the class of graphical models. The likelihood function and maximum likelihood estimators of the covariance and the mean parameters are studied when the observations are normally distributed. Applications to modelling nonstationary dependence structures and multivariate data are discussed and illustrated using real data. A graphical method, similar to that based on the correlogram in time series, is developed and used to identify parametric models for nonstationary covariances.

Some key words: Antedependence; Cholesky decomposition; Generalised linear model; Linear regression and autoregression; Link function; Multivariate normal; Nonstationary model; Stationary model.

1. Introduction

Modelling a covariance matrix Σ is difficult because of (a) the possibly high dimensionality of the problem and (b) the constraint that Σ must be positive definite.

Our goal is to introduce an unconstrained parameterisation for and to model a general covariance matrix in terms of covariates, as is done for the mean vector in the generalised linear models in McCullagh & Nelder (1989). Anderson's (1973) class of linear covariance models appears to be the natural starting point, but unfortunately here the linear coefficients are constrained so that a covariance matrix is positive definite. Pinheiro & Bates (1996) present five unconstrained parameterisations of a covariance matrix using Cholesky decomposition, spectral decomposition and matrix logarithmic transformation (Chiu, Leonard & Tsui, 1996; Leonard & Hsu, 1992). Their new parameters, however, do not always have simple statistical interpretation. We use the modified Cholesky decomposition of Σ^{-1}, not Σ, to propose a statistically meaningful unconstrained parameterisation of covariance and a link function, thereby removing difficulties (a) and (b). Since Σ^{-1} is the canonical covariance parameter of a multivariate normal distribution, modelling its

unconstrained parameters as a linear combination of covariates is in agreement with the approach of generalised linear models and subsumes naturally the ideas of the antedependence model of Gabriel (1962), covariance selection of Dempster (1972) and the class of graphical models in Cox & Wermuth (1996, Ch. 3), in which certain entries of Σ^{-1} or its triangular factor are set to zero.

A key result used is that (Newton, 1988, p. 359) a symmetric matrix Σ is positive definite if and only if there exists a unique unit lower triangular matrix T, with 1's as diagonal entries, and a unique diagonal matrix D with positive diagonal entries such that

$$T\Sigma T' = D. \tag{1}$$

Fortunately, T and D are easy to compute and interpret statistically: the below-diagonal entries of T are the negatives of the coefficients of $\hat{Y}_t = \mu_t + \sum_{j=1}^{t-1} \phi_{t,j}(Y_j - \mu_j)$, the linear least-squares predictor of Y_t based on its predecessors $Y_{t-1}, \ldots, Y_1$, and the diagonal entries of D are the prediction error variances $\sigma_t^2 = \mathrm{var}(Y_t - \hat{Y}_t)$, for $1 \leqslant t \leqslant n$. Since $\phi_{t,j}$ and $\log \sigma_t^2$ are unconstrained, we may model them in terms of covariates. To this end, for $t = 1, \ldots, n$ and $j = 1, \ldots, t-1$, consider the models

$$\mu_t = m(x_t, \beta), \quad \log \sigma_t^2 = v(z_t, \lambda), \quad \phi_{t,j} = d(z_{t,j}, \gamma). \tag{2}$$

where $m(.,.)$, $v(.,.)$, $d(.,.)$ are functions, x_t, z_t, $z_{t,j}$ are $p \times 1$, $q_1 \times 1$, $q_2 \times 1$ vectors of covariates, and $\beta = (\beta_1, \ldots, \beta_p)'$, $\lambda = (\lambda_1, \ldots, \lambda_{q_1})'$ and $\gamma = (\gamma_1, \ldots, \gamma_{q_2})'$ are parameters corresponding to the mean, variance and dependence, respectively. We refer to (2) as the joint mean-covariance model and note that it is composed of three submodels describing the mean, variance and dependence of a random vector. It also subsumes naturally the framework of joint modelling of mean and variance heterogeneity in Cook & Weisberg (1983) and Verbyla (1993). In fact, when $Y_1, \ldots, Y_n$ are independent, we have $\phi_{t,j} \equiv 0$ or $d \equiv 0$ and $\sigma_t^2 = \mathrm{var}(Y_t)$. For excellent reviews and references on joint modelling of the mean and variance heterogeneity see McCullagh & Nelder (1989, Ch. 10) and Chiu et al. (1996).

The outline of the paper is as follows. Section 2 introduces an unconstrained parameterisation and a generalised linear model for a covariance matrix along with examples and a discussion of the class of antedependence models and limitations of (2) in modelling structured covariances. Analysis of a real dataset illustrating details of our method for fitting (2) to data is given in § 3. Section 4 provides three distinct representations, corresponding to the three submodels in (2), of the likelihood function of a multivariate normal random vector. For simplicity we assume throughout this paper that Σ is strictly positive definite.

2. UNCONSTRAINED PARAMETERISATION

2·1. *Reparameterisation of* Σ

The idea of regression is our key tool. For $1 \leqslant t \leqslant n$, let $\hat{Y}_t$ stand for the linear least-squares predictor of Y_t based on its predecessors $Y_{t-1}, \ldots, Y_1$, and let ε_t be its prediction error with variance $\sigma_t^2 = \mathrm{var}(\varepsilon_t)$. For simplicity set $\mu = E(Y) = 0$. Thus, for $t = 1$, $\hat{Y}_1 = E(Y_1) = 0$, and for $1 < t \leqslant n$ consider the unique scalars $\phi_{t,j}$ minimising $E(Y_t - \sum_{j=1}^{t-1} c_j Y_j)^2$ with respect to the c_j's. Set $\phi_t = (\phi_{t,1}, \ldots, \phi_{t,t-1})'$, for $t = 2, \ldots, n$. Then from standard regression theory in Anderson (1984, pp. 125–38) we have

$$\hat{Y}_t = \sum_{j=1}^{t-1} \phi_{t,j} Y_j, \quad \phi_t = \Sigma_t^{-1} \sigma_t, \quad \sigma_t^2 = \sigma_{tt} - \sigma_t' \Sigma_t^{-1} \sigma_t, \tag{3}$$

where Σ_t is the $(t-1)\times(t-1)$ leading principal minor of Σ and σ_t is the column vector composed of the first $t-1$ entries of the tth column of Σ. By convention, we set all empty sums to zero, that is $\sum_{j=1}^{0} x_j = 0$. The random variables

$$\varepsilon_t = Y_t - \hat{Y}_t = Y_t - \sum_{j=1}^{t-1} \phi_{t,j} Y_j \quad (t = 1, \ldots, n) \tag{4}$$

being successive prediction errors are uncorrelated, so that, with $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$, $D = \mathrm{cov}(\varepsilon)$ is a diagonal matrix, that is $D = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Writing (4) in matrix form one obtains

$$\varepsilon = TY, \tag{5}$$

where T is a unit lower triangular matrix with $-\phi_{t,j}$ in the (t, j)th position for $2 \leqslant t \leqslant n$ and $j = 1, 2, \ldots, t-1$. From (5) and the definition of D, it follows that

$$\mathrm{cov}(\varepsilon) = T\,\mathrm{cov}(Y)T' = T\Sigma T' = D, \tag{6}$$

so that the matrix T diagonalises the covariance matrix Σ. This diagonalisation is related to the modified Cholesky decomposition of Σ and Σ^{-1} (Newton, 1988, p. 359).

Since the nonredundant entries of T and D have statistical meaning, the $\frac{1}{2}n(n+1)$ constrained and hard-to-model parameters of Σ can be traded in for the $\frac{1}{2}n(n+1)$ unconstrained and interpretable parameters $\phi_{t,j}$, $\log\sigma_t^2$, for $1 \leqslant t \leqslant n$ and $1 \leqslant j \leqslant t-1$. We refer to the new parameters $\phi_{t,j}$'s and σ_t^2's as the generalised autoregressive parameters and the innovation variances of Σ or Y.

2·2. *Generalised autoregressive parameters and regressograms*

In this section we introduce a plot similar to the correlogram in time series to be used in identifying models for generalised autoregressive parameters. For a fixed $t \geqslant 2$, following Tukey (1961) we refer to the plot of $\phi_{t,j}$ versus $j = 1, 2, \ldots, t-1$ as the tth theoretical regressogram of Σ. It would be natural to call a plot of σ_t^2 versus $t = 1, 2, \ldots, n$ the theoretical variogram of Σ, but since this term is already in use in a different context in Diggle (1988) we shall not use it here and refer to a statistical procedure employing all these plots as a regressogram-based procedure.

Heuristically and from $\hat{Y}_t$ in (3), since $\phi_{t,t-j}$ is the lag-j regression coefficient one expects it to be small for a fixed t and large j, and the sequence $\phi_{t,t-j}$, for $j = 1, 2, \ldots, t-1$, is expected to be monotone decreasing. We use empirical regressograms for assessing graphically the nature of dependence of Y_t on its predecessors and for suggesting parametric models for $\phi_{t,j}$ and σ_t^2. This is analogous to using the empirical correlogram, variogram and lorelogram to arrive at parametric models for their theoretical counterparts, as in Diggle (1988) and Heagerty & Zeger (1998).

The maximum likelihood estimators of these parameters and hence the regressograms are obtained easily using (3) along with the invariance property of and the knowledge of the maximum likelihood estimators of μ and Σ for a multivariate normal distribution in Anderson (1984, Ch. 3). More precisely, let $y_1, \ldots, y_m$ be a sample from a population with $N(\mu, \Sigma)$ distribution. Then

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} y_i, \quad S = \hat{\Sigma} = \frac{1}{m}\sum_{i=1}^{m} (y_i - \hat{\mu})(y_i - \hat{\mu})', \quad \hat{\phi}_t = S_t^{-1} s_t, \quad \hat{\sigma}_t^2 = s_{tt} - s_t' S_t^{-1} s_t, \tag{7}$$

where $S = (s_{i,j})$, S_t and s_t are the sample analogues of Σ_t and σ_t.

Conditional on $Y_1, \ldots, Y_{t-1}$ being fixed, the distribution of $\hat{\phi}_t$ is $N(\phi_t, n^{-1}\sigma_t^2 S_t^{-1})$. Hence, the significance of an individual $\phi_{t,j}$ $(t \geqslant 2)$ can be tested using the test statistic $\hat{\phi}_{t,j}/(n^{-\frac{1}{2}}\hat{\sigma}_t^2 s_t^{jj})$, which has a Student-$t$ distribution with $n-t+1$ degrees of freedom, where s_t^{jj} is the jth diagonal entry of S_t^{-1}. More generally, the significance of the whole vector ϕ_t can be tested using R_t^2, the multiple correlation coefficient between Y_t and $Y_1, \ldots, Y_{t-1}$, for $t \geqslant 3$, and (Anderson, 1984, p. 140)

$$\left(\frac{n-t+1}{t-2}\right)\left(\frac{R_t^2}{1-R_t^2}\right) \sim F_{t-2,n-t+1}.$$

2·3. *Linear mean-covariance models*

In this section we discuss the flexibility and some properties of model (2) when $m(.,.)$, $v(.,.)$ and $d(.,.)$ are linear functions of their parameters. We refer to such a model as a linear mean-covariance model.

In general, q_1 and q_2 in (2) are different and the entries of λ and γ could be quite distinct, but for simplicity and economy of notation occasionally we use the combined vector $\alpha = (\gamma', \lambda')'$ of dimension $q = q_1 + q_2$ to parameterise Σ. Similarly, a $\frac{1}{2}n(n+1) \times q$ design matrix Z of covariates is constructed by appropriately concatenating z_t's and $z_{t,j}$'s and padding them with zeros if necessary. If we define ϕ_t as in (3), the transformation $h(.)$, given by

$$h(\Sigma) = (\phi_2', \ldots, \phi_n', \log\sigma_1^2, \ldots, \log\sigma_n^2)' = Z\alpha, \tag{8}$$

is a link function (McCullagh & Nelder, 1989, p. 27) for a linear mean-covariance model. To make this linear framework useful for longitudinal data analysis, following Diggle, Liang & Zeger (1994, p. 16), we introduce the following notation for the data, parameters and covariates:

$$\begin{gathered} Y = (Y_1', \ldots, Y_m')', \quad \mu = (\mu_1', \ldots, \mu_m')', \quad \Sigma = \text{block diag}\,(\Sigma_1, \ldots, \Sigma_m), \\ X = (X_1', \ldots, X_m')', \quad \mu_i = X_i\beta, \quad Z = \text{block diag}\,(Z_1, \ldots, Z_m), \quad h(\Sigma_i) = Z_i\alpha, \end{gathered} \tag{9}$$

where now the subscript i refers to the ith subject or cluster in the study. Then

$$Y \sim N(X\beta, Z\alpha) \tag{10}$$

is a suggestive shorthand for the distribution of sample data from a population with a linear mean-covariance model. Now, (10) can be seen as an extended generalised linear model for any longitudinal/multivariate data, with the help of a slight modification of McCullagh & Nelder's (1989, p. 27) three-part specification involving two link functions (g, h), one for the mean and the other for covariance.

As a useful alternative to (8) we write $h(\Sigma)$ as a symmetric matrix Θ, where its main diagonal is the logarithm of the diagonal entries of D, its first subdiagonal, i.e. the lag-one regression coefficients, is the first subdiagonal of T and so forth. Since its entries are merely rearrangements of those of $Z\alpha$, we have

$$\Theta = \sum_{j=1}^{q} \alpha_j U_j, \tag{11}$$

where the U_j's are symmetric covariate matrices. The idea of linear covariance structure, i.e. using (11) to model functionals of a covariance, was initiated by Anderson (1973). Our model (8) or (11) is close to that in Chiu et al. (1996), but avoids some statistical and computational problems of their matrix-logarithmic covariance model where $\log\Sigma$ is modelled as in (11).

Model (8) is capable of producing nonstationary analogues of many special-structure stationary covariances available in the literature of longitudinal data analysis. Its real strength is in modelling nonstationary features where variances increase over time, and measurements equidistant in time are not equicorrelated. For additional flexibility we may rely on nonlinear functions as models for μ_t, $\log \sigma_t^2$ and $\phi_{t,j}$, in the same manner that Heagerty & Zeger (1998) use nonlinear and nonparametric functions in modelling pairwise log-odds ratios. Among special structures commonly used in longitudinal data analysis, the two-parameter compound symmetry and AR(1) models are the most popular. Others include the three-parameter damped exponential family in Muñoz et al. (1992) and the four-parameter family in Diggle (1988). Note that, for $i = 2, \ldots, n$ and $j = 1, \ldots, i-1$, the matrices T with $\phi_{i,j}$ given respectively by

$$\gamma, \quad \gamma^{i-j}, \quad \gamma^{(t_i - t_{i-j})^\theta}, \quad \gamma_j^{f(t_i,\lambda_j) - f(t_{i-j},\lambda_j)}, \quad \gamma_{i-j} \tag{12}$$

are analogous to compound symmetry, AR(1), damped exponential, structured antedependence (Zimmerman & Núñez-Antón, 1997) and banded. However, unlike the stationary case, the γ, θ and λ are unconstrained and $f(.,.)$ in (12) is a known function.

By (1), the positive definiteness of the estimated covariance matrix is guaranteed. In contrast, in Anderson's (1973) linear covariance model complicated constraints on the coefficients are needed to ensure positive definiteness, and, in the approach of Liang & Zeger (1986), the positive definiteness of the estimated covariance matrix is not guaranteed; see also Crowder (1995) and Pinheiro & Bates (1996).

2·4. *Examples*

Example 1. (a) (*Pinheiro & Bates*, 1996). For this Σ, its 6×1 vector of covariance predictors $h(\Sigma)$ is computed using

$$\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 5 \\ 1 & 5 & 14 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix} = LDL',$$

$$T = L^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}, \quad D = \mathrm{diag}(1, 4, 9), \quad h(\Sigma) = (1, 0, 1, 0, \log 4, \log 9)'.$$

From (4), it follows that $Y_1 = \varepsilon_1$, $Y_t = Y_{t-1} + \varepsilon_t$, for $t = 2, 3$.

(b) Given $h(\Sigma) = (3, -1{\cdot}5, -1, 0, -1, 2)'$ for an unknown 3×3 matrix Σ, the matrix is recovered by first constructing T and D:

$$T = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1{\cdot}5 & 1 & 1 \end{pmatrix}, \quad D = \mathrm{diag}(1, e^{-1}, e^2).$$

Then from (1) one computes Σ.

Example 2. (a) For $n = 2$ and $q = 1$ and an arbitrary covariance covariate $Z = (z_1, z_2, z_3)'$,

the generalised linear model (8) amounts to the following reparameterisation of Σ:

$$\Sigma = e^{z_2\alpha}\begin{pmatrix} 1 & z_1\alpha \\ z_1\alpha & z_1^2\alpha^2 + e^{(z_1 - z_2)\alpha} \end{pmatrix},$$

containing only one unconstrained parameter α. The parameterised correlation coefficient between Y_1 and Y_2 is given by $z_1\alpha[z_1^2\alpha^2 + \exp\{(z_3 - z_2)\alpha\}]^{-\frac{1}{2}}$, which approaches ± 1 when $z_3 - z_2$ approaches $-\infty$. In a longitudinal study with two measurements made on a subject at times $t_1 < t_2$, the choice of $z_1 = t_1$, $z_2 = -(t_2 - t_1)$ and $z_3 = -(t_2 - t_1)^2$ leads to a covariance matrix with many desirable decay properties when t_1 and t_2 grow far apart.

(b) For $n = 2$ and $q = 2$, $\alpha = (\alpha_1, \alpha_2)'$ and

$$Z = \begin{pmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \end{pmatrix}',$$

the linear model for the unconstrained entries of $h(\Sigma) = Z\alpha$ can be solved using (8) to express entries of Σ in terms of the new covariance parameters α_1, α_2 and the explanatory variables in Z as follows:

$$\sigma_{11} = e^{\alpha_1 + z_2\alpha_2}, \quad \sigma_{21} = (\alpha_1 + z_1\alpha_2)e^{\alpha_1 + z_2\alpha_2}, \quad \sigma_{22} = e^{\alpha_1 + z_3\alpha_2} + (\alpha_1 + z_1\alpha_2)^2 e^{\alpha_1 + z_2\alpha_2}.$$

(*c*) The alternative representation (11) of $h(\Sigma)$ above is

$$\Theta = \begin{pmatrix} \log\sigma_1^2 & \phi_{12} \\ \phi_{12} & \log\sigma_2^2 \end{pmatrix} = \alpha_1 J + \alpha_2 \begin{pmatrix} z_2 & z_1 \\ z_1 & z_3 \end{pmatrix} = \alpha_1 U_1 + \alpha_2 U_2,$$

where $U_1 = J$ is the 2×2 matrix of 1's and the choice for U_2 is obvious.

Next, we highlight recognisable features of regressograms for AR(1) and compound symmetry.

Example 3. The covariance matrix of an AR(1) model is given by $\Sigma = \sigma^2(\rho^{|i-j|})_{i,j=1}^n$, for $|\rho| < 1$ and $\sigma^2 > 0$. It follows from (3) that, for $t \geqslant 2$, $\phi_t = (0, \ldots, 0, \rho)'$, $\sigma_t^2 = \sigma^2$, and $\sigma_1^2 = \sigma^2(1 - \rho^2)^{-1}$, so that, assuming that $\mu = 0$, $Y_1, \ldots, Y_n$ satisfy $Y_1 = \varepsilon_1$ and $Y_t = \rho Y_{t-1} + \varepsilon_t$, for $2 \leqslant t \leqslant n$.

Here, only the lag-one generalised autoregressive parameters are nonzero and σ_1^2 is a nonlinear function of ρ. Thus, the theoretical regressograms for AR(1) and more generally for AR(p) models are simpler to recognise; they drop off to zero for lags $j > p$ and σ_t^2 is constant for $t > p$.

Example 4. The covariance matrix of a compound symmetry model is given by

$$\Sigma = \sigma^2\{(1 - \rho)I + \rho J\} \quad (-(n-1)^{-1} < \rho < 1, \sigma^2 > 0).$$

It follows from (3) that $\phi_1 = 0$, $\sigma_1^2 = \sigma^2$ and, for $t \geqslant 2$,

$$\phi_t = \rho\{1 + (t-1)\rho\}^{-1}1_{t-1}, \quad \sigma_t^2 = \sigma^2\left\{1 - \frac{(t-1)\rho^2}{1 + (t-1)\rho}\right\},$$

where 1_{t-1} is a $(t-1)$-dimensional vector of 1's. If we assume that $\mu = 0$, it follows that $Y_1, \ldots, Y_n$ satisfy

$$Y_t = \rho\{1 + (t-2)\rho\}^{-1}\sum_{j=1}^{t-1} Y_j + \varepsilon_t \quad (t = 1, \ldots, n),$$

where, unlike with the AR(1), all generalised autoregressive parameters are nonzero and

for a given t all predecessors of Y_t receive identical coefficients. Also, all generalised autoregressive parameters and innovation variances are nonlinear functions of ρ and the time. Figure 1 provides plots of these parameters versus time for a compound symmetry with $\rho = 0{\cdot}5$ and $\sigma^2 = 1$. Figure 1(a) shows a distinct compound symmetry feature where, for each t, the tth regressogram is flat. Of course, such theoretical features of regressograms for special covariance structures are crucial at the model identification stage.

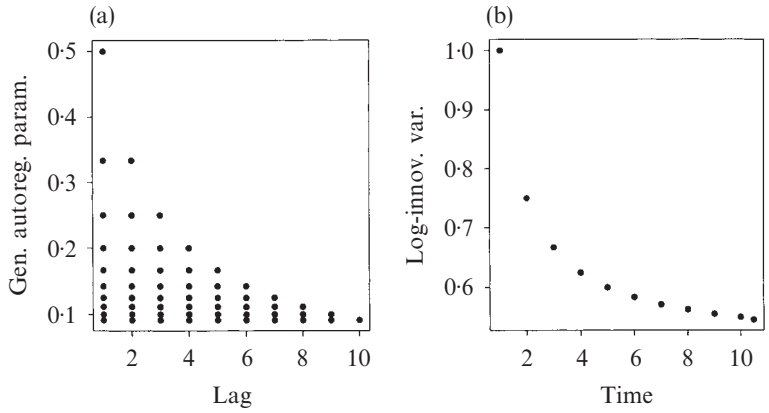

Fig. 1. Regressograms of a compound symmetry with correlation coefficient, ρ, equal to 0·5 and variance, $\sigma^2 = 1$. (a) Generalised autoregressive parameters and (b) log-innovation variances.

2·5. *Antedependence models*

In this section, as another and important example of (4) and (8), we consider the class of antedependence models of order p, denoted by AD(p) for short. The order p serves as a memory gauge, where $p = 0$ corresponds to independence and $p = n - 1$ to arbitrary multivariate dependence. The random variables $Y_1, \ldots, Y_n$, indexed by time, are said to be AD(p) if the conditional distribution of Y_t given $Y_{t-1}, \ldots, Y_1$ depends on $Y_{t-1}, \ldots, Y_{t-p}$, for all $t \geqslant p$ (Gabriel, 1962). This concept is equivalent to $Y_1, \ldots, Y_n$ having a Markovian dependence of order p (Diggle et al., 1994, p. 85).

Next, we show that AD(p) dependence of measurements is equivalent to certain generalised autoregressive parameters being zero. From (4), it follows that a normal random vector $Y = (Y_1, \ldots, Y_n)'$ with mean $\mu = (\mu_1, \ldots, \mu_n)'$ is AD(p) if and only if

$$Y_t = \mu_t + \sum_{j=1}^{p_t^*} \phi_{t,t-j}(Y_{t-j} - \mu_{t-j}) + \varepsilon_t \quad (t = 1, \ldots, n), \tag{13}$$

where $p_t^* = \min(p, t-1)$. However, (13) is equivalent to the last $n - p - 1$ subdiagonals of T or Θ or $U_1, \ldots, U_q$ being identically equal to zero. We use this simple observation to give an alternative proof of a result of Gabriel (1962, Theorem 1) characterising AD(p) in terms of certain entries of Σ^{-1} being zero; see Theorem 1(c) below.

THEOREM 1. *Let* $Y \sim N(\mu, \Sigma)$ *with* Σ *factored as in* (1), *and let* p *be a fixed integer between* 0 *and* $n - 1$. *Then the following are equivalent*:

(a) $Y_1, \ldots, Y_n$ *are* AD(p),

(b) *the last* $n - p - 1$ *subdiagonals of* T *are zero*,

(c) *the last* $n - p - 1$ *subdiagonals of* Σ^{-1} *are zero*,

(d) *the last* $n-p-1$ *subdiagonals of* $U_1, \ldots, U_q$ *in* (11) *are zero.*

The variable-order antedependence models in Macchiavelli & Arnold (1994) generalise Gabriel's constant-order antedependence models by allowing the order p to depend on times of measurements. This amounts to setting to zero certain entries of T, instead of Σ^{-1}, in a manner that is more liberal than Gabriel's, and yet less general than Dempster's (1972) method.

Although the AD(p) is more parsimonious than an arbitrary covariance, still it has too many parameters to be useful in practice. Zimmerman & Núñez-Antón (1997) seem to have been the first to reduce the number of parameters using covariates in the spirit of (2). Denoting the measurement times for an arbitrary subject by $t_1 < t_2 < \ldots < t_n$, they consider the following AD(p) model with time-dependent coefficients:

$$\begin{gathered} \phi_{ij} = \phi_j^{f(t_i,\lambda_j)-f(t_{i-j},\lambda_j)} \quad (i = p+1, \ldots, n;\ j = 1, \ldots, p), \\ \sigma_t^2 = \sigma^2 g(t, \theta) \quad (i = p+1, \ldots, n), \end{gathered} \tag{14}$$

where $\phi_1, \ldots, \phi_p$ are positive, and $f(.,.)$ and $g(.,.)$ are known functions with parameters $\lambda_1, \ldots, \lambda_p$ and θ.

2·6. *Structured covariances*

In this section we discuss limitations of (2) in modelling stationary and other structured covariances. Evidently, any structure imposed on Σ will lead to constraints on T and D so that (2) is not directly applicable. In some mildly structured cases, however, we are able to handle the constraints on T and D by choosing the covariates in Z appropriately, as in the AD(p) model of § 2·5, where certain elements of Z are set to zero. On the other hand, the stationary structure is not easily amenable to (8), because T does not have recognisable unconstrained entries, as in Theorem 1(b), say, but Examples 3 and 4 suggest that in this case one must forfeit the linearity of $\phi_{t,j}$ in (8). Also, the entries of the matrix D are order-restricted, i.e.

$$\sigma_1^2 = \sigma_{11} \geqslant \sigma_2^2 \geqslant \ldots \geqslant \sigma_n^2, \tag{15}$$

because of the constancy of the diagonal entries of a stationary covariance.

Since an $n \times n$ correlation matrix has 1's as diagonal entries, effectively it has $\frac{1}{2}n(n-1)$ distinct parameters and is structured. Hence, certain entries of T and D are either redundant or known. For example, the diagonal entries of the matrix D are monotone decreasing as in (15) with $\sigma_1^2 = \sigma_{11} = 1$, and $n-1$ of the below-diagonal entries of T are redundant. To accommodate these constraints, one may choose z_1 in (8) to be zero and, after estimating β, λ, α and the σ_t^2's, we may rearrange the remaining $z_2, \ldots, z_n$ so that (15) is satisfied. The redundancy of the below-diagonal entries of T can be resolved, for example by either judiciously setting to zero $n-1$ of its entries or following the idea of variable-order antedependence models in Macchiavelli & Arnold (1994).

3. PRELIMINARY ANALYSIS OF THE CATTLE DATA

This section relies on regressograms to identify models like (2) for a real dataset. Since modelling strategies for the mean and variance heterogeniety are well developed in Verbyla (1993) we focus mainly on modelling the dependence components $\phi_{t,j}$. The models and their parameter estimates are only preliminary and should not be regarded as final.

Table 1: *Cattle data. Sample variances* (*along the main diagonal*), *correlations* (*above the main diagonal*), *generalised autoregressive parameters* (*below the main diagonal*) *and innovation variances* (*last row*) *for the group* A *cattle*

t	1	2	3	4	5	6	7	8	9	10	11
1	106	0·82	0·76	0·66	0·64	0·59	0·52	0·53	0·52	0·48	0·48
2	1·00	155	0·91	0·84	0·80	0·74	0·63	0·67	0·60	0·58	0·55
3	0·05	0·90	165	0·93	0·88	0·85	0·75	0·77	0·71	0·70	0·68
4	−0·23	0·16	0·98	185	0·94	0·91	0·82	0·84	0·77	0·73	0·71
5	0·04	0·00	0·00	1·06	243	0·94	0·87	0·89	0·84	0·80	0·77
6	−0·02	−0·21	0·16	0·26	0·83	284	0·93	0·94	0·90	0·86	0·83
7	0·20	−0·34	−0·04	−0·03	0·15	1·00	306	0·97	0·93	0·88	0·86
8	−0·06	0·01	0·06	−0·26	0·22	0·61	0·41	341	0·97	0·94	0·92
9	0·21	−0·14	0·01	−0·22	−0·03	−0·03	0·33	0·93	389	0·96	0·96
10	−0·24	0·10	0·39	−0·26	−0·11	−0·04	−0·17	0·31	1·01	470	0·98
11	0·13	−0·23	0·09	0·23	0·01	−0·31	−0·05	−0·05	0·33	0·86	445
	106	50	29	25	27	28	37	29	16	28	9

Kenward (1987) reports an experiment in which cattle were assigned randomly to two treatment groups A and B, and their weights were recorded to study the effect of treatments on intestinal parasites. Thirty animals received treatment A and another 30 received treatment B. They were weighed $n = 11$ times over a 133-day period; the first 10 measurements on each animal were made at two-week intervals and the final measurement was made one week later. The measurement times were common across animals and were rescaled to $t = 1, 2, \ldots, 10, 10{\cdot}5$. No observation was missing so this is a balanced longitudinal dataset.

For the treatment group A with $m = 30$ animals, we assume a common mean vector μ and an 11×11 covariance matrix Σ. A profile plot of the data reveals that the weights have an upward trend and their variances tend to increase over time, which suggests nonstationary covariance structure. This is confirmed by the upper diagonal entries in Table 1, which are the sample correlations. Furthermore, correlations within the subdiagonals are not constant and increase over time, giving a second indication that a stationary covariance is not appropriate for the data. Table 1 gives the sample correlations, generalised autoregressive parameters and the innovation variances, and the latter two are plotted in Fig. 2(a), (c). These plots reveal that both the sample generalised autoregressive parameters and the logarithms of the innovation variances are cubic functions of the lag. That is, for $t = 1, 2, \ldots, 10, 11$

$$\begin{aligned} \log \hat{\sigma}_t^2 &= \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \varepsilon_{t,v}, \\ \hat{\phi}_{t,j} &= \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \varepsilon_{t,j,d} \quad (j = 1, 2, \ldots, t-1). \end{aligned} \tag{16}$$

The variance parameters $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)'$ and dependence parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)'$ can be estimated using the maximum likelihood method developed in § 4·1, but for demonstration and simplicity we use the least squares method here; the estimates are given in Table 2.

The differences in the magnitudes of sample σ_t^2's and $\phi_{t,j}$'s both in this case and in general are the main reason for separate parameterisation of innovation variances and generalised autoregressive parameters in (16) and (2). The fitted variances, correlations, innovation variances and generalised autoregressive parameters using (16) are given in

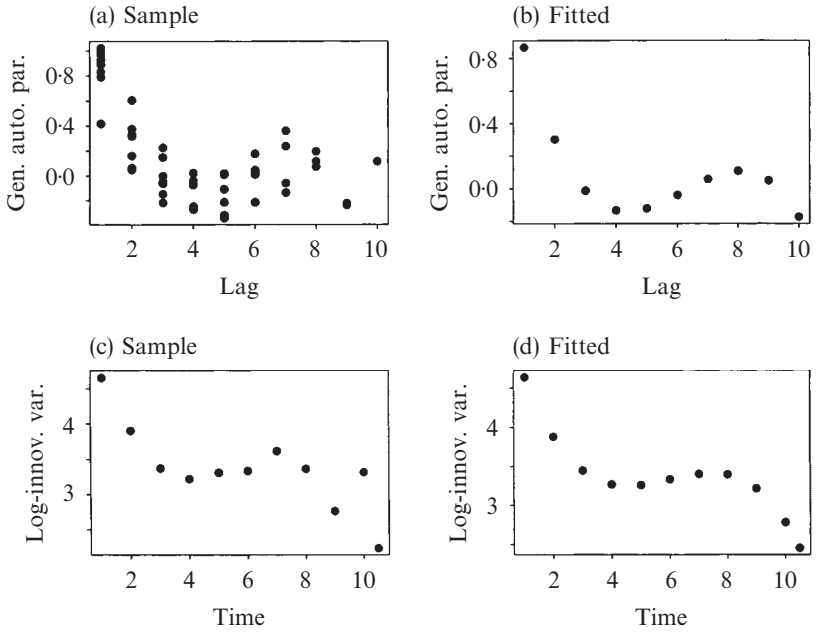

Fig. 2. Sample and fitted regressograms for the cattle data. (a) Sample generalised autoregressive parameters, (b) fitted generalised autoregressive parameters, (c) sample log-innovation variances and (d) fitted log-innovation variances. The fitted values in (b) and (d) are from the fitted cubic polynomials in (16).

Table 2: *Cattle data. Least squares estimates of the parameters of the two cubic polynomials in* (16)

Parameters	$j=1$	$j=2$	$j=3$	$j=4$
λ_j	3·37	−1·47	0·24	−0·93
γ_j	0·18	−1·71	1·64	−1·11

Table 3. Comparison with their sample values in Table 1 shows a surprisingly good agreement and reveals the potential power of regressograms in suggesting parsimonious models for Σ. Note that, with $n = 11$, the unstructured covariance has 66 parameters, but Kenward (1987) and Macchiavelli & Arnold (1994) used 30 and 26 parameters, respectively, in their constant and variable-order antedependence models. Zimmerman & Núñez-Antón (1997), using antedependence models with time as covariate, see (14), were able to reduce the number of required parameters to as low as 6. Our preliminary linear model (16) with only 8 parameters achieved results comparable to those of Zimmerman & Núñez-Antón (1997), in the sense that the entries of the fitted covariance matrices are 'close' to each other. To compare the fits more rigorously, it is standard to rely on penalised likelihood criteria such AIC and BIC. In our context of covariance model selection the BIC is defined as

$$\text{BIC} = -\frac{2}{m}L + p\frac{\log m}{m}, \tag{17}$$

where m is the sample size, L is the maximised loglikelihood for a covariance model and

Table 3: *Cattle data. Fitted innovation variances* (*along the main diagonal*), *correlations* (*above the main diagonal*), *and generalised autoregressive parameters* (*below the main diagonal*) *for the group* A *cattle.*

t	1	2	3	4	5	6	7	8	9	10	11
1	99	0·75	0·76	0·74	0·69	0·63	0·57	0·52	0·50	0·50	0·44
2	0·87	59	0·89	0·89	0·86	0·80	0·74	0·67	0·63	0·62	0·59
3	0·30	0·87	35	0·95	0·93	0·89	0·82	0·76	0·71	0·68	0·65
4	−0·01	0·30	0·87	24	0·96	0·93	0·87	0·81	0·76	0·73	0·69
5	−0·13	−0·01	0·30	0·87	22	0·96	0·92	0·86	0·80	0·77	0·73
6	−0·12	−0·13	−0·01	0·30	0·87	25	0·95	0·91	0·86	0·82	0·78
7	−0·04	−0·12	−0·13	−0·01	0·30	0·87	30	0·95	0·91	0·88	0·85
8	0·06	−0·04	−0·12	−0·13	−0·01	0·30	0·87	33	0·95	0·94	0·91
9	0·11	0·06	−0·04	−0·12	−0·13	−0·01	0·30	0·87	31	0·97	0·96
10	0·05	0·11	0·06	−0·04	−0·12	−0·13	−0·01	0·30	0·87	19	0·99
11	−0·17	0·05	0·11	0·06	−0·04	−0·12	−0·13	−0·01	0·30	0·87	9

p is the number of covariance parameters. A smaller value of BIC is associated with a better fitting model. The values of L and BIC for the covariance induced by model (16) are $-1051{\cdot}81$ and 71·03, respectively. These compare quite favourably with the corresponding values $-1054{\cdot}20$ and 70·96 of the AD(2) covariance model chosen by Zimmerman & Núñez-Antón (1997).

Note that the fitted generalised autoregressive parameters in Table 3 are constant along the subdiagonals; equivalently the fitted T is a band matrix. This is an artifact of model (16) and the chosen covariates whereby, for instance, the fitted value for $\hat{\phi}_{t,1}$ is $\hat{\gamma}_1 + \hat{\gamma}_2 + \hat{\gamma}_3 + \hat{\gamma}_4$ and does not depend on t. Only the first two subdiagonals have sizeable fitted values, 0·87 and 0·30, relative to others, similar to the sample-based pattern present in Table 1.

Next, the sample generalised autoregressive parameters are tested for significance using the F-tests described at the end of § 2·2. The observed values of the test statistic for testing $H_0: \phi_t = 0$ for $t = 3, \ldots, 11$ are 41·81, 25·49, 18·33, 13·59, 7·19, 7·16, 10·08, 4·00 and 5·14, and the corresponding critical values for significance level 0·05 are 5·12, 4·46, 4·35, 4·53, 5·05, 6·16, 8·89, 19·37 and 240·50, respectively. These tests indicate that, for $t = 3, \ldots, 9$, at least some entries of ϕ_t are significantly different from zero.

4. The likelihood function and model fitting

4·1. *The likelihood function*

The factorisation (1) facilitates considerably the computation of the multivariate normal likelihood function, which has three distinct representations corresponding to the three sets of parameters or submodels in (2) when the observations $Y_i \sim N(\mu, \Sigma)$, for $i = 1, 2, \ldots, m$, are independent. Moreover, we shall see that for the generalised linear model (8) the loglikelihood is a quadratic function of the dependence parameters γ.

Thanks to (1), expressions involving the determinant and the inverse of Σ can be handled easily. Also, since the action of the matrix T in (5) is to map any vector to its vector of prediction errors, if we define $r_i = y_i - \mu_i = (r_{i,t})_{t=1}^{n}$, we obtain $Tr_i = r_i - \hat{r}_i$, for $i = 1, 2, \ldots, m$, where $\hat{r}_{i,t}$ as in (3) is the best linear predictor of $r_{i,t}$ based on its predecessors $r_{i,j}$, for $1 \leqslant j \leqslant t-1$. In the following, we also need $r(t) = (r_{i,t})_{i=1}^{m}$, which is the vector of centred observations made on the tth occasion on all m subjects.

By (1), the quadratic form Q in the exponent of the likelihood function can be written

$$Q = \sum_{i=1}^{m} (y_i - \mu_i)'\Sigma^{-1}(y_i - \mu_i) = \sum_{i=1}^{m} r_i' T' D^{-1} T r_i$$
$$= \sum_{i=1}^{m} (r_i - \hat{r}_i)' D^{-1} (r_i - \hat{r}_i) = \sum_{i=1}^{m} \sum_{t=1}^{n} \frac{(r_{i,t} - \hat{r}_{i,t})^2}{\sigma_t^2} = \sum_{t=1}^{n} \frac{\text{RSS}_t}{\sigma_t^2},$$

where

$$\text{RSS}_t = \sum_{i=1}^{m} (r_{i,t} - \hat{r}_{i,t})^2 \tag{18}$$

is the residual sum of squares from the analysis of covariance of $r(t)$ with $r(t-1), \ldots, r(1)$ as covariates (Kenward, 1987). From (18) and if we assume a linear mean-covariance model, it is evident that RSS_t and hence Q are quadratic functions of the correlation parameters γ:

$$\text{RSS}_t = \sum_{i=1}^{m} \left(r_{i,t} - \sum_{j=1}^{t-1} \phi_{t,j} r_{i,j} \right)^2 = \sum_{i=1}^{m} \left\{ r_{i,t} - \left(\sum_{j=1}^{t-1} z_{t,j}' r_{i,j} \right) \gamma \right\}^2 = \sum_{i=1}^{m} \{ r_{i,t} - z'(i,t)\gamma \}^2,$$
$$Q = \sum_{i=1}^{m} \sum_{t=1}^{n} \sigma_t^{-2} \{ r_{i,t} - z'(i,t)\gamma \}^2 = \sum_{i=1}^{m} \{ r_i - Z(i)\gamma \}' D^{-1} \{ r_i - Z(i)\gamma \}, \tag{19}$$

where

$$z(i,t) = \sum_{j=1}^{t-1} z_{t,j} r_{i,j}, \quad Z(i) = (z(i,1), \ldots, z(i,n))', \tag{20}$$

are respectively $q_2 \times 1$ and $n \times q_2$ matrices.

The loglikelihood $L(\beta, \lambda, \gamma;\ Y)$, up to the additive constant $mn \log 2\pi$, satisfies

$$-2L(\beta, \lambda, \gamma;\ Y) = m \log |\Sigma| + \sum_{i=1}^{m} (y_i - X_i\beta)'\Sigma^{-1}(y_i - X_i\beta)$$
$$= m \sum_{t=1}^{n} \log \sigma_t^2 + \sum_{t=1}^{n} \frac{\text{RSS}_t}{\sigma_t^2}$$
$$= m \sum_{t=1}^{n} \log \sigma_t^2 + \sum_{i=1}^{m} \{ r_i - Z(i)\gamma \}' D^{-1} \{ r_i - Z(i)\gamma \}. \tag{21}$$

If we use the above representations of the loglikelihood in (21) the score vector and the Fisher expected information can be computed, and a three-stage estimation procedure can be developed by viewing (21) as involving three sub-models for the mean, variance and correlation (Smyth, 1989; Verbyla, 1993). For given (λ, γ) or Σ, the first equation in (21) defines the mean model with y_i as its response; for given β and γ, the second identity is viewed as the variance model with RSS_t as response; and, for given β and λ, the third identity can be viewed as the correlation model with r_i as response.

For unbalanced data, the nature of the computations involved in maximising the likelihood function is similar to those in (21). Let us define

$$Q(\alpha) = \sum_{i=1}^{m} \{ Y_i - X_i\hat{\beta}(\alpha) \}' \Sigma_i^{-1} \{ Y_i - X_i\hat{\beta}(\alpha) \} = \sum_{i=1}^{m} Q_i(\alpha),$$

where $\Sigma_i = \Sigma_i(t_i;\ \alpha)$ is the $n_i \times n_i$ covariance matrix of Y_i, X_i is the $n_i \times p$ matrix of covariates

for the ith subject, and, for a given α, $\hat{\beta} = \hat{\beta}(\alpha)$ stands for the generalised least squares estimator of β. Then the likelihood function of the unbalanced data satisfies

$$-2L(\hat{\beta}, \alpha) = \sum_{i=1}^{m} \{\log|\Sigma_i| + Q_i(\alpha)\}.$$

For each α, evaluation of $L(\hat{\beta}, \alpha)$ involves at most m determinants and inverses of Σ_i. Since q, the dimension of α, is usually small, one can use either the Nelder–Mead simplex algorithm or a quasi-Newton algorithm (Diggle, 1988), neither of which requires partial derivatives of $L(\hat{\beta}, \alpha)$. However, these algorithms do not provide the ingredients necessary to compute standard errors of $\hat{\alpha}$. When this is needed, one may use the Newton–Raphson algorithm using the first two partial derivatives of $L(\hat{\beta}, \alpha)$. The details of the exact Fisher scoring and Newton–Raphson methods are rather lengthy and are hence deferred to a follow-up paper.

4·2. *Model fitting*

In fitting model (8) to data a slight modification of Diggle's (1988) three-stage approach can be adopted. Since RSS_t in (18) is the residual sum of squares from the analysis of covariance of $r(t)$ with $r(t-1), \ldots, r(1)$ as covariates, using (21) one could use the technique of Kenward (1987) to express the likelihood-ratio tests in terms of individual components from analyses of covariance even though Y is not assumed to be antedependent.

To implement our procedure, we must identify potential covariance covariates. The experience of the last two decades indicates that the matrix Z_i in (9) usually consists of subject-specific covariates, with times t_{ij} and time-separations $z_{i,jk} = |t_{ij} - t_{ik}|$ often playing the most prominent roles. They fall into two distinct categories depending on whether there is a desire to fit stationary or nonstationary models. The stationary description is more common in longitudinal data analysis (Diggle, 1988) even when the data do not necessarily support it. For nonstationary dependence t_{ij} is used often as covariate: this is evident in (16); in the mixed model approach of Laird & Ware (1982), where the covariance is expressed in terms of the subject-specific covariates including t_{ij}; and in the structured antedependence models of Zimmerman & Núñez-Antón (1997), where $\phi_{i,j}$ is modelled as in (12). In addition, any existing special structure covariance such as the compound symmetry, AR(1), can serve as the building blocks and serve as covariate matrix in (11). The set-up of (11) and (10) allows us to combine, compare and test for a particular covariance structure when faced with several alternative special structure covariances $\Sigma_1, \ldots, \Sigma_q$ (Chiu et al., 1996). This is particularly attractive since, because of (2) or (10), we are able to fit nested covariance models (Diggle et al., 1994, Ch. 5).

5. Future work

Much more work is needed to bring this methodology to the current level of generalised linear model theory for mean modelling. Our follow-up paper is concerned with the problems of maximum likelihood estimation of β and α in (8) using Fisher scoring and iterative Newton–Raphson methods. For nonnormal data, β and α would be estimated using the idea of generalised estimating equations in Liang & Zeger (1986) and applying (1)–(2) to the working correlation matrices. Since the mean-covariance model (2) extends the mean-variance model in Verbyla (1993), his procedures on maximum likelihood and restricted maximum likelihood and diagnostics would be extended to our more general set-up.

ACKNOWLEDGEMENT

I would like to thank Professor Nan Laird, for pointing out the importance of non-stationary covariances in longitudinal data analysis, Professors Bala Hosmane and Sanjib Basu of Northern Illinois University for their interest and many discussions, and the editor and a referee for their concrete suggestions that improved the presentation of the paper considerably. The work was supported by grants from the National Security Agency and the National Science Foundation.

REFERENCES

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1**, 135–41.

ANDERSON, T. W. (1984). *An Introduction to Multivariate Analysis*, 2nd ed. New York: John Wiley.

CHIU, T. Y. M., LEONARD, T. & TSUI, K. W. (1996). The matrix-logarithm covariance model. *J. Am. Statist. Assoc.* **91**, 198–210.

COOK, R. D. & WEISBERG, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70**, 1–10.

COX, D. R. & WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation.* London: Chapman and Hall.

CROWDER, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika* **82**, 407–10.

DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28**, 157–75.

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

DIGGLE, P. J., LIANG, K. Y. & ZEGER, S. L. (1994). *Analysis of Longitudinal Data.* Oxford: Oxford University Press.

GABRIEL, K. R. (1962). Ante-dependence analysis of an ordered set of variables. *Ann. Math. Statist.* **33**, 201–12.

HEAGERTY, P. J. & ZEGER, S. L. (1998). Lorelogram: A regression approach to exploring dependence in longitudinal categorical response. *J. Am. Statist. Assoc.* **93**, 150–62.

KENWARD, M. G. (1987). A method for comparing profiles of repeated measurements. *Appl. Statist.* **36**, 296–308.

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

LEONARD, T. & HSU, J. S. J. (1992). Bayesian inference for a covariance matrix. *Ann. Statist.* **20**, 1669–96.

LIANG, K. Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

MACCHIAVELLI, R. E. & ARNOLD, S. F. (1994). Variable order antedependence models. *Commun. Statist.* A **23**, 2683–99.

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

MUÑOZ, A., CAREY, V., SCHOUTEN, J. P., SEGAL, M. & ROSNER, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* **48**, 733–42.

NEWTON, H. J. (1988). *TIMESLAB: A Time Series Analysis Laboratory.* Pacific Grove, CA: Wadsworth & Brooks/Cole.

PINHEIRO, J. D. & BATES, D. M. (1996). Unconstrained parameterizations for variance-covariance matrices. *Statist. Comp.* **6**, 289–96.

SMYTH, G. K. (1989). Generalized linear models with varying dispersion. *J. R. Statist. Soc.* B **51**, 47–60.

TUKEY, J. W. (1961). Curves as parameters, and touch estimation. In *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, **1**, Ed. J. Neyman, pp. 681–94. Berkeley: University of California Press.

VERBYLA, A. P. (1993). Modelling variance heterogeneity: Residual maximum likelihood and diagnostics. *J. R. Statist. Soc.* B **55**, 493–508.

ZIMMERMAN, D. L. & NÚÑEZ-ANTÓN, V. (1997). Structured antedependence models for longitudinal data. In *Modeling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions*, Springer Lecture Notes in Statistics, No. 122, Ed. T. G. Gregoire et al., pp. 63–76. New York: Springer-Verlag.

[*Received April* 1998. *Revised December* 1998]