

Computational Aspects of Likelihood-Based Estimation of First-Order Antedependence Models

by

Dale L. Zimmerman
Department of Statistics and Actuarial Science, University of Iowa
Iowa City, Iowa 52242, U.S.A.

Vicente Núñez-Antón
Departamento de Econometría y Estadística, Universidad del País Vasco - Euskal Herriko Unibertsitatea,
Avenida del Lehendakari Aguirre 83, 48015 Bilbao - Vizcaya, Spain

Hammou El-Barmi
Department of Mathematics, University of Texas
Austin, Texas 78712, U.S.A.

SUMMARY

Computational aspects of the problem of estimating the parameters of first-order antedependence models for the covariance structure of longitudinal data are considered. These models may be useful when serial correlation exists among measurements within subjects but is nonstationary. For those situations in which measurement times are common across subjects, the maximum likelihood estimators of unstructured first-order antedependent model parameters can be given explicitly. When measurement times are not identical across subjects or when the first-order antedependence model is more structured, however, the likelihood-based estimation of model parameters must be accomplished numerically and would thus appear to require extensive computations. In this article we show how this computational burden can be substantially reduced. The usefulness of the results is illustrated by an analysis of longitudinal data from a 100-km race.

Key words: Antedependence models, Longitudinal data, Patterned covariance matrices, Repeated measurements.

1. INTRODUCTION

Several approaches to the analysis of longitudinal data are based on various cases of the general linear model

$$\mathbf{Y}_k = \mathbf{X}_k\boldsymbol{\beta} + \mathbf{e}_k, \quad k = 1, 2, \dots, n \quad (1)$$

where \mathbf{Y}_k is the $p_k \times 1$ vector of responses for subject k ; \mathbf{X}_k is a $p_k \times q$ design matrix of rank q for subject k ; the \mathbf{e}_k 's are independent random vectors whose distribution is multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\Sigma_k = \Sigma_k(\boldsymbol{\theta})$; $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are vectors of unknown parameters; p_k is the number of occasions on which the k th subject is observed (the occasions need not be equally spaced or identical across subjects); and n is the number of subjects. Methods of analysis based on model (1) differ with respect to the model assumed for the covariance structure, i.e., how Σ_k is assumed to depend on $\boldsymbol{\theta}$. Three types of models for the within-subject covariance structure have received the most attention: multivariate, random effects, and stationary autoregressive models. Ware (1985) reviews these main types and describes their relative advantages and disadvantages; these models or minor variations of them have also been considered by Jennrich and Schluchter (1986), Lee (1988), Diggle (1988), Schluchter (1988), Jones (1990), Jones and Boadi-Boateng (1991), and Muñoz, Carey, Schouten, Segal, and Rosner (1992).

A fourth group of models for the covariance structure of (1) that has received relatively less attention are the antedependence models, first introduced by Gabriel (1962) but also considered by Byrne and Arnold (1983), Kenward (1987), Jones (1993), Macchiavelli and Arnold (1994), and Diggle, Liang, and Zeger (1994). A set of random variables W_1, W_2, \dots, W_m whose joint distribution is multivariate normal is said to be s th-order antedependent if W_i and W_{i+k+1} are independent given $W_{i+1}, W_{i+2}, \dots, W_{i+k}$ for all i and for all $k \geq s$. Antedependence models, like stationary autoregressive models, allow for serial correlation but they are more general in the sense that they do not stipulate that all variances are equal nor that correlations between all pairs of measurements (on the same subject) equidistant in time are equal.

Byrne and Arnold (1983) considered parameter estimation for an unstructured first-order antedependence model for the case in which measurement times are common across subjects. In this case an explicit expression exists for the maximum likelihood estimator of the covariance matrix. When measurement times are not

identical across subjects or when one wishes to fit a more structured first-order antedependence model, however, the maximum likelihood estimators cannot be given explicitly and so must be obtained by numerically maximizing the likelihood function. If the number of measurement times is large, the computations required to accomplish this appear initially to be formidable. The purpose of this article is to describe how this computational burden can be substantially reduced.

2. First-order Antedependence Models

2.1. Unstructured Model

Suppose that a first-order antedependence model, hereafter called an AD(1) model, is adopted for the elements of each error vector $\mathbf{e}_1, \dots, \mathbf{e}_n$ in (1). For an arbitrary subject k consider the covariance matrix $\Sigma_{i_k}(\theta)$ of \mathbf{e}_{i_k} , which we assume to be positive definite, and for economy of notation let us omit the subscript indexing the subject and suppress the dependence on θ . Let $0 < t_1 < t_2 < \dots < t_p$ denote the time points at which observations are taken on this subject, let Y_i denote the response at time t_i , and $\sigma_i^2 = \text{var}(Y_i)$ and $\rho_{ij} = (\sigma_i \sigma_j)^{-1} \text{cov}(Y_i, Y_j)$. The definition of first-order antedependence is equivalent to the condition (see e.g., Feller, 1966, sect. III.8)

$$\rho_{ij} = \rho_{im} \rho_{mj} \quad \text{for all } i < m < j,$$

which in turn is easily shown to be equivalent to the condition

$$\rho_{ij} = \prod_{l=i}^{j-1} \rho_{l,l+1} \quad \text{for all } i + 1 < j.$$

So for an AD(1) model we can write Σ as follows, putting $\rho_i = \rho_{i,i+1}$:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_1 & \sigma_1 \sigma_3 \rho_1 \rho_2 & \sigma_1 \sigma_4 \rho_1 \rho_2 \rho_3 & \cdots & \sigma_1 \sigma_p \prod_{i=1}^{p-1} \rho_i \\ & \sigma_2^2 & \sigma_2 \sigma_3 \rho_2 & \sigma_2 \sigma_4 \rho_2 \rho_3 & \cdots & \sigma_2 \sigma_p \prod_{i=2}^{p-1} \rho_i \\ & & \sigma_3^2 & \sigma_3 \sigma_4 \rho_3 & \cdots & \vdots \\ & & & \ddots & \cdots & \vdots \\ \text{symm} & & & & \sigma_{p-1}^2 & \sigma_{p-1} \sigma_p \rho_{p-1} \\ & & & & & \sigma_p^2 \end{bmatrix} \quad (2)$$

Observe that an unstructured, positive definite AD(1) model imposes no restrictions on the variances (except

that they be positive); all that is required of the correlations for positive definiteness are that they satisfy the restrictions $-1 < \rho_i < 1$ for $i = 1, \dots, p-1$. Moreover, the model specifies that for each fixed measurement time t_i , the magnitude of the correlation between the measurement at time t_i and measurements at subsequent times is a monotone decreasing function of the temporal spacing but that this function need not be identical for all i . Such monotonicity or near-monotonicity is a commonly observed feature of the correlation structure of many actual longitudinal data sets. Also note that the covariance matrix of an unstructured AD(1) model is completely determined by the $2p - 1$ elements on its main diagonal and first super-diagonal, or equivalently, by the $2p - 1$ parameters $\sigma_1^2, \dots, \sigma_p^2, \rho_1, \dots, \rho_{p-1}$. Thus the unstructured AD(1) model has fewer parameters than the multivariate model (which has $p(p+1)/2$ parameters), but it has more parameters than a stationary first-order autoregressive model (which has only two parameters).

2.2. Structured Models

Although there are fewer parameters in the unstructured AD(1) model than in the multivariate model, they may still be too numerous to be estimated efficiently from the available data. In such cases a more parsimonious, structured version of the AD(1) model may be more useful. One step towards parsimony is taken by assuming that

$$\rho_i = \rho^{f(t_{i+1}, \lambda) - f(t_i, \lambda)}, \quad \sigma_i = \sigma^2 g(t_i; \psi), \quad (3)$$

where $0 < \rho < 1$, $\sigma^2 > 0$, and λ and ψ are vectors of relatively few parameters. In this family of AD(1) models the correlations are restricted to be positive, a feature usually exhibited by actual longitudinal data. Particular choices of the functions $f(\cdot)$ and $g(\cdot)$ may, of course, impose additional restrictions on the correlations and variances.

Potentially useful choices for $f(\cdot)$ and $g(\cdot)$ might include simple linear models (linear in the elements of λ and ψ) and power models. Indeed, the special case of (3) in which λ is a scalar,

$$f(t, \lambda) = \begin{cases} (t^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log t & \text{if } \lambda = 0 \end{cases} \quad (4)$$

and $g(t; \psi) \equiv 1$, with parameter space $\{(\rho, \lambda, \sigma^2): 0 < \rho < 1, -\infty < \lambda < \infty, \sigma^2 > 0\}$, was used by Núñez-Antón and Woodworth (1994) in a study of the efficacy of cochlear implants. This model, in which the

Box-Cox family of power transformations is applied to the time scale, was adopted to account for the authors' belief that subjects "learned" over time, with the result that responses equidistant in time became more highly correlated as the study progressed. This kind of nonstationary correlational behavior corresponds to $\lambda < 1$ in (4). On the other hand, a decrease in correlations between equidistant responses as the study progresses would correspond to $\lambda > 1$. We note that the model specified by putting $\lambda = 1$ in (4) and taking $g(t; \psi) \equiv 1$ is the continuous-time analogue of a stationary first-order autoregressive process.

There are many other possibilities, of course, for $f(\cdot)$ and $g(\cdot)$. For instance, a location shift parameter could be incorporated into (4) to remove the possible arbitrariness of starting the time scale at $t = 0$ and to avoid the boundary problem at $t = 0$ when $\lambda \leq 0$. This suggestion has been made previously for the Box-Cox transformation, in the different context of regression analysis, by a number of authors (e.g. Atkinson, 1985). As for $g(\cdot)$, a family of power transformations might be useful when the variance of the response is expected to increase (or decrease) monotonically over time, as is typical of growth data, for example.

3. Computational Aspects of Likelihood-based Estimation

Under model (1) the log-likelihood function, apart from an additive constant, is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = -\frac{1}{2} \sum_{k=1}^n \log |\Sigma_k(\boldsymbol{\theta})| - \frac{1}{2} \sum_{k=1}^n (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta})' \Sigma_k^{-1}(\boldsymbol{\theta}) (\mathbf{Y}_k - \mathbf{X}_k \boldsymbol{\beta}).$$

Maximum likelihood (ML) estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are values $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ that maximize L , or equivalently, $\hat{\boldsymbol{\theta}}$ is any value of $\boldsymbol{\theta}$ that maximizes

$$\begin{aligned} L^*(\boldsymbol{\theta}; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = & -\frac{1}{2} \sum_{k=1}^n \log |\Sigma_k(\boldsymbol{\theta})| - \frac{1}{2} \sum_{k=1}^n \mathbf{Y}_k' \Sigma_k^{-1}(\boldsymbol{\theta}) \mathbf{Y}_k \\ & + \frac{1}{2} \left[\sum_{k=1}^n \mathbf{X}_k' \Sigma_k^{-1}(\boldsymbol{\theta}) \mathbf{Y}_k \right]' \left[\sum_{k=1}^n \mathbf{X}_k' \Sigma_k^{-1}(\boldsymbol{\theta}) \mathbf{X}_k \right]^{-1} \left[\sum_{k=1}^n \mathbf{X}_k' \Sigma_k^{-1}(\boldsymbol{\theta}) \mathbf{Y}_k \right], \end{aligned}$$

and $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\beta}} = \left[\sum_{k=1}^n \mathbf{X}_k' \Sigma_k^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{X}_k \right]^{-1} \left[\sum_{k=1}^n \mathbf{X}_k' \Sigma_k^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{Y}_k \right].$$

Several authors (Diggle, 1988; Muñoz et al., 1992; Núñez-Antón and Woodworth, 1994) have recommended the use of restricted maximum likelihood (REML) estimation over ML estimation for its better bias properties.

A REML estimate of θ is any value $\tilde{\theta}$ that maximizes

$$L^{**}(\theta; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = L^*(\theta; \mathbf{Y}_1, \dots, \mathbf{Y}_n) - \frac{1}{2} \log \left| \sum_{k=1}^n \mathbf{X}'_k \Sigma_k^{-1}(\theta) \mathbf{X}_k \right|$$

and the corresponding estimate of β is given by

$$\tilde{\beta} = \left[\sum_{k=1}^n \mathbf{X}'_k \Sigma_k^{-1}(\tilde{\theta}) \mathbf{X}_k \right]^{-1} \left[\sum_{k=1}^n \mathbf{X}'_k \Sigma_k^{-1}(\tilde{\theta}) \mathbf{Y}_k \right].$$

If $\Sigma_k(\theta)$ is given by an unstructured AD(1) model for all $k = 1, \dots, n$, and if measurement times are identical (but not necessarily equally spaced) across subjects, then $\Sigma_1 = \Sigma_2 = \dots = \Sigma_k \equiv \Sigma$ say, and provided $n \geq q + 2$, an explicit expression exists for the maximum likelihood estimator of Σ (Byrne and Arnold, 1983). On the other hand, if measurement times are not identical across subjects, or one wishes to fit a more structured AD(1) model such as one of those described in the previous section, explicit expressions for the ML (or REML) estimators generally do not exist and estimates of θ must be obtained by maximizing L^* (or L^{**}) numerically (e.g. by grid search or gradient algorithms).

The numerical maximization of L^* or L^{**} would appear to be a formidable computational problem, particularly when at least some subjects are measured on a "large" number of occasions, for then the determinants and inverses of some "large" covariance matrices must be obtained. Of course, substantial savings can result from computing the determinant and inverse only once for each distinct pattern of measurement times that occurred in the study. Even greater savings can be had, however, by using explicit formulae for the determinant and inverse of a positive definite first-order antedependent matrix, which can be given as special cases of results given by Barrett (1979) for matrices satisfying a slightly more general condition which he called "the triangle property." A determinantal formula equivalent to Barrett's has appeared in the literature on antedependence models (viz., Theorem 1 of Byrne and Arnold, 1983), but due perhaps to the fact that Barrett did not refer to, and indeed may not have been aware of, antedependence models *per se*, an explicit inversion formula has not appeared or been referred to in the antedependence model literature. It is well-known (e.g. Gabriel, 1962), however, that the inverse of a nonsingular AD(1) covariance matrix is tridiagonal.

Barrett's formulas for $|\Sigma|$ and $\Sigma^{-1} = \sigma^{ij}(\)$, they specialize to a $p \times p$ positive definite, unstructured first-order antedependent matrix Σ of the form (2), are as follows :

$$|\Sigma| = \left(\prod_{i=1}^p \sigma_i^2 \right) \left(\prod_{i=1}^{p-1} (1 - \rho_i^2) \right) \quad (5)$$

$$\sigma^{ij} = \begin{cases} [\sigma_1^2(1 - \rho_1^2)]^{-1} & \text{for } i = j = 1 \\ [\sigma_p^2(1 - \rho_{p-1}^2)]^{-1} & \text{for } i = j = p \\ (1 - \rho_{i-1}^2 \rho_i^2) [\sigma_i^2(1 - \rho_{i-1}^2)(1 - \rho_i^2)]^{-1} & \text{for } i = j \neq 1, p \\ -\rho_i [\sigma_i \sigma_j (1 - \rho_i^2)]^{-1} & \text{for } |i - j| = 1 \\ 0 & \text{for } |i - j| > 1 \end{cases} \quad (6)$$

Thus, the determinant and inverse of Σ can be obtained with $O(p)$ computations, in contrast to the $O(p^3)$ computations required for performing these operations for an arbitrary positive definite matrix. Expressions for the determinant and inverse of a more structured first-order antedependent covariance matrix such as that prescribed by (3) and (4) can, of course, be obtained from (5) and (6) by substituting for $\{\sigma_i\}$ and $\{\rho_i\}$.

A small computational study was conducted to investigate how much savings could actually be achieved by using (5) and (6). The study consisted of computing the inverse and determinant of $N \times N$ AD(1) covariance matrices by each of two methods: Method 1, which uses (5) and (6); and Method 2, which uses the IMSL subroutines DLINDS (for the inverse) and DLFTDS and DLFDDS (for the determinant). The IMSL subroutines exploit only the property that the matrix is real, symmetric, and positive definite. The study was performed on a CONVEX supercomputer in FORTRAN using vectorization and optimization options; all computations were done in double precision. Table 1 gives the CPU times required by each method for a range of values of N . These results indicate that as N increases from 10 (corresponding to a rather small number of measurement times) to 5000 [which might be regarded as unrealistically large; however, see Wegman (1994)], the relative reduction in computation achieved by Method 1 increases from 57% to 97%.

4. Example

In order to illustrate further the usefulness of the results of the previous section, we present in this section an analysis of data from a 100-km race held in 1984 in the United Kingdom. The data, which were kindly provided by Ian Jolliffe of the University of Kent, consist of the "split" times for each of 80 competitors in each 10-km section of the race. In addition to the split times, the data set contains the ages of all but four of the competitors; to simplify the presentation these four observations are set aside and the analysis is based on the

remaining 76 competitors. Some exploratory and graphical analyses of these data are reported by Everitt (1994a, 1994b). Here we shall go beyond exploratory analyses and actually fit models to the data. Our goal is to find a parsimonious model that describes how the performance of a competitor on each 10-km section is related to the section number ($t = 1, 2, \dots, 10$), and to the competitor's age.

Table 2 gives the sample means, variances, and correlations among the 10 split times; correlations of split times with age are also given. The sample means indicate that the split times tend to increase over the first 80 km of the race, but then level off and perhaps even decrease slightly over the last 20 km. This suggests that a model whose mean structure is a linear function of t may be inadequate. As for the variances and correlations, we observe that: (a) the variances tend to increase as the race progresses; (b) the correlations among split times are positive and quite large; (c) the correlations between age and each split time are weak; (d) the correlation between the split time for a fixed 10-km section and split times for successive sections tends to decrease monotonically; (e) the correlations between split times for adjacent 10-km sections are smaller in the later sections of the race than in the earlier sections. The first and last of these points indicate that models for the covariance structure that allow for nonstationarity in both the variances and correlations should be entertained.

Taking into account the features just noted, we fit a particular AD(1) model of form (1), which we call the full model, to the data. The mean structure of this model consisted of a cubic function of t and a linear function of age, i.e.,

$$E(Y_{kt}) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 x_k, \quad k = 1, \dots, 76 \text{ and } t = 1, \dots, 10,$$

where Y_{kt} is the split time for competitor k on section t and x_k is the age of competitor k . The model's covariance structure was that given by (3), with $f(\cdot)$ given by (4) and

$$g(t; \psi) = 1 + \psi_1 t + \psi_2 t^2.$$

Thus, $L^{**}(\cdot)$ is a function of the 5×1 parameter vector $\theta = (\rho, \lambda, \sigma^2, \psi_1, \psi_2)'$. Note, in contrast, that the restricted log-likelihood function associated with an unstructured AD(1) model would be a function of 19 parameters.

Following Diggle (1988), the maximum of $L^{**}(\cdot)$ was sought using the Nelder-Mead simplex algorithm (Nelder and Mead, 1965), with a step size of .01 for each variable. A change in $L^{**}(\cdot)$ of 1.0×10^{-10} on

successive iterations was taken as the convergence criterion, and we took $\tilde{\theta}_0 = (\tilde{\rho}_0, 1, \tilde{\sigma}_0^2, 0, 0)'$ as a starting value, where $\tilde{\rho}_0$ and $\tilde{\sigma}_0^2$ were obtained from a preliminary fit of a stationary autoregressive model of order one. After 229 iterations, the convergence criterion was satisfied, yielding the estimate $\tilde{\theta} = (\tilde{\rho} = 0.929, \tilde{\sigma}^2 = 1.600, \tilde{\psi}_1 = 0.590, \tilde{\psi}_2 = 0.0450)'$. To verify that $\tilde{\theta}$ corresponds to the global maximum and is thus the unique REML estimate, a grid search was performed over a $15 \times 16 \times 16 \times 21$ grid of values for $(\rho, \lambda, \psi_1, \psi_2)$; the grid was four-dimensional rather than five-dimensional since σ^2 can be “concentrated out” of $L^{**}(\cdot)$ (see Diggle, 1988). In this way $\tilde{\theta}$ was confirmed to be the REML estimate of θ , and the corresponding estimate of β was $\tilde{\beta} = (43.428, 1.354, 0.253, -0.0169, 0.0688)'$. Observe that $\tilde{\lambda} > 1$, which corresponds to the aforementioned decrease, as the race progresses, in the correlation between adjacent split times.

Various submodels were fitted and compared to the full model by a series of stepdown tests. For a submodel corresponding to a hypothesis about parameters of the mean structure, the test consists of computing twice the difference of the maximized value of the log-likelihood function for the full model and the value of the log-likelihood function for the submodel evaluated at the full-model estimates, then comparing this to a chosen percentile of the central chi-square distribution with degrees of freedom equal to the difference in the numbers of parameters in the two models. We based these tests on the ordinary, rather than the restricted, log-likelihood so that they would be equivalent to Wald tests. For submodels corresponding to hypotheses about covariance parameters, the method of Harville and Callanan (1990) was used; see also Everitt (1995, p. 132). This method consists of determining whether a particular hypothesized value of θ lies in the set of θ -values for which $2[L^{**}(\tilde{\theta}) - L^{**}(\theta)]$ exceeds the upper α percentage point of the chi-square distribution with degrees of freedom equal to the difference in the number of covariance parameters in the full model and the submodel. These tests indicated that: (a) age is not an important explanatory variable ($P = 0.73$); (b) the linear term in the mean structure is a statistically significant regressor, but somewhat surprisingly (in view of the aforementioned apparent speed-up over the last 20 km of the race) the quadratic and cubic terms are not statistically significant at traditional levels ($P = 0.26$ and $P = 0.31$, respectively); (c) inclusion of the quadratic term in $g(\cdot)$ does not significantly improve the fit ($P = 0.21$) but the linear term does; (d) λ is significantly different from 1.0 ($P < 0.005$), i.e., the correlations are nonstationary. In view of these results, we refitted the submodel of the full model in which $\beta_2 = \beta_3 = \beta_4 = \psi_2 = 0$. Estimates of the parameters were as follows: $\tilde{\beta}_0 = 45.57, \tilde{\beta}_1$

$= 2.285, \tilde{\rho} = 0.948, \tilde{\lambda} = 1.789, \tilde{\sigma}^2 = 15.16, \tilde{\psi}_1 = 1.018$. Estimates of the standard errors of prediction at each 10-km interval could then be computed using the variance estimate

$$\text{estvar} \begin{bmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{bmatrix} = \left[\sum_{k=1}^{76} \mathbf{X}'_k \Sigma_k^{-1}(\tilde{\theta}) \mathbf{X}_k \right]^{-1}.$$

To carry out the analysis just described required that a 10×10 AD(1) covariance matrix be inverted, and have its determinant computed, tens of thousands of times. Even more such computations would have been required had there been any missing data, for in that case the same covariance matrix would not apply to each subject. Clearly, the aforementioned 57% reduction in computation when $N = 10$, afforded by utilizing (5) and (6), is very welcome in this situation and would be even more so if N was larger.

In the analysis presented here we have restricted our attention to highly structured AD(1) models for the covariance structure. For these data the measurement times are identical across subjects, so alternatively we could have used the unstructured AD(1) approach of Byrne and Arnold (1983). We prefer our approach, however, as it models the interesting features of the covariance structure that were readily apparent in Table 2. It is possible, of course, that higher-order antedependence models may fit these data better than any AD(1) model, even a completely unstructured one. Indeed, an application of Gabriel's (1962) method for determining the order of antedependence suggests that an AD(5) model fits these data significantly better than antedependence models of order four or less and not significantly worse than antedependence models of order six or higher. However, an unstructured AD(5) model for these data has 45 parameters in its covariance structure, so it is neither as parsimonious nor as interpretable as the structured AD(1) model we have fitted. An unstructured variable-order AD model (Macchiavelli and Arnold, 1994), which would probably have considerably fewer than 45 covariance parameters, could be fit to these data but such a model would not account for some of the clear structural features (e.g., the increasing variances) of the covariance matrix. The development and fitting of structured, higher-order antedependence models (of either constant or variable order) is a problem requiring further research.

5. Concluding Remarks

In this note we have shown how the computational burden of likelihood-based estimation of first-order antedependent model parameters can be substantially reduced. It is our hope that these results will be of use to practitioners who might wish to fit AD(1) models (particularly highly structured models) to their longitudinal data but presently are stymied by the apparently large computing requirements.

We close by describing how the results herein can be used in conjunction with a slightly more general model for the covariance structure of longitudinal data. One feature of the structured AD(1) model given by (3) that may sometimes be unrealistic, especially when measurement involves subsampling of material from each subject, is that the correlation between observations on the same subject tends to one as their temporal spacing becomes small. Consequently it may be desirable to consider a model for the covariance structure in which independent measurement errors, each with variance τ_k^2 , are added to the AD(1) model. That is, we may wish to consider a model in which $\text{var}(\mathbf{Y}_k) = \Sigma_k(\theta) + \tau_k^2 \mathbf{I} \equiv \Gamma_k$ where $\tau_k^2 > 0$, $\Sigma_k(\theta)$ is AD(1), and Γ_k is positive definite for all $k = 1, \dots, n$. Matrix Γ_k is not first-order antedependent, so (5) and (6) cannot be used directly to obtain $|\Gamma_k|$ and Γ_k^{-1} . However, using well-known formulas we have that

$$|\Sigma_k(\theta) + \tau_k^2 \mathbf{I}| = \tau_k^{2p_k} |\Sigma_k(\theta)| |\Sigma_k^{-1}(\theta) + \tau_k^{-2} \mathbf{I}|$$

and

$$[\Sigma_k(\theta) + \tau_k^2 \mathbf{I}]^{-1} = \tau_k^{-2} \mathbf{I} - \tau_k^{-4} [\Sigma_k^{-1}(\theta) + \tau_k^{-2} \mathbf{I}]^{-1},$$

where the matrix $\Sigma_k^{-1}(\theta) + \tau_k^{-2} \mathbf{I}$ is symmetric and tridiagonal. Thus, the determinant and inverse of Γ_k can be obtained efficiently by combining (5) and (6) with available algorithms for computing the determinant and inverse of symmetric tridiagonal matrices, e.g., the algorithms of Fiedler (1986, Theorem 8.5) and Ikebe (1979).

ACKNOWLEDGEMENTS

Zimmerman's work was partially supported by the National Science Foundation under Grant DMS-9206966. Núñez-Antón's work was supported by the University of the Basque Country under Research Grant UPV 038.321-HA 021/94.

REFERENCES

Atkinson, A. C. (1985). *Plots, Transformations and Regression. An Introduction to Graphical*

- Methods of Diagnostic Regression Analysis*. New York: Oxford University Press.
- Barrett, W. W. (1979). A theorem on inverses of tridiagonal matrices. *Linear Algebra and its Applications* 27, 211-217.
- Byrne, P. J., and Arnold, S. F. (1983). Inference about multivariate means for a nonstationary autoregressive model. *Journal of the American Statistical Association* 78, 850-855.
- Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics* 44, 959-971.
- Diggle, P. J., Liang, K.-Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. New York: Oxford University Press.
- Everitt, B. S. (1994a). Exploring multivariate data graphically: a brief review with examples. *Journal of Applied Statistics* 21, 63-94.
- Everitt, B. S. (1994b). *A Handbook of Statistical Analyses Using S-Plus*. London: Chapman and Hall.
- Everitt, B. S. (1995). The analysis of repeated measures: a practical review with examples. *The Statistician* 44, 113-135.
- Feller, W. (1966). *An Introduction to Probability Theory and its Applications Volume II*. Chichester: Wiley.
- Fiedler, M. (1986). *Special Matrices and their Applications in Numerical Mathematics*. Dordrecht: Martinus Nijhoff.
- Gabriel, K. R. (1962). Ante-dependence analysis of an ordered set of variables. *Annals of Mathematical Statistics* 33, 201-212.
- Harville, D. A., and Callanan, T. (1990). Computational aspects of likelihood-based inference for variance components. In *Advances in Statistical Methods for Genetic Improvement of Livestock*, D. Gianola and K. Hammonds, eds., pp. 136-176, Heidelberg: Springer-Verlag.
- Ikebe, Y. (1979). On inverses of Hessenberg matrices. *Linear Algebra and its Applications* 24, 93-97.
- Jennrich, R. L., and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 42, 805-820.
- Jones, R. H. (1990). Serial correlation or random subject effects? *Communications in Statistics - Simulation and Computation* 19, 1105-1123.

- Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach*. London: Chapman and Hall.
- Jones, R. H., and Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics* 47, 161-175.
- Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Applied Statistics* 36, 296-308.
- Lee, J. C. (1988). Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association* 83, 432-440.
- Macchiavelli, R. E. and Arnold, S. F. (1994). Variable order ante-dependence models. *Communications in Statistics — Theory and Methods* 23, 2683-2699.
- Muñoz, A., Carey, V., Schouten, J. P., Segal, M. and Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 48, 733-742.
- Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal* 7, 308-313.
- Núñez-Antón, V. and Woodworth, G. G. (1994). Analysis of longitudinal data with unequally spaced observations and time-dependent correlated errors. *Biometrics* 50, 445-456.
- Schluchter, M. D. (1988). Analysis of incomplete multivariate data using linear models with structured covariance matrices. *Statistics in Medicine* 7, 317-324.
- Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician* 39, 95-101.
- Wegman, E. J. (1994). Huge data sets and the frontiers of computational feasibility. Technical Report No. 110, Center for Computational Statistics, George Mason University.

Table 1. CPU times, in seconds, for computing the inverse and determinant of an $N \times N$ AD(1) covariance matrix by each of two methods. The methods are described in Section 3.

N	Method 1	Method 2
10	0.00176	0.00411
50	0.00763	0.02837
100	0.02574	0.10930
500	0.60614	4.90808
1000	2.81839	31.17996
5000	114.86062	4050.86823

Table 2. Summary statistics for the 100-km race data: (a) sample means (in minutes) and variances; (b) sample correlations.

(a)	Section (t)	Mean	Variance
	1	47.80	26.89
	2	50.92	34.78
	3	49.63	49.01
	4	53.27	58.89
	5	54.69	91.41
	6	60.12	149.90
	7	62.38	107.85
	8	69.33	152.22
	9	68.70	144.99
	10	67.36	167.21

(b)

$t = 1$	1.0											
$t = 2$.95	1.0										
$t = 3$.84	.89	1.0									
$t = 4$.78	.82	.92	1.0								
$t = 5$.60	.63	.75	.88	1.0							
$t = 6$.60	.62	.72	.84	.94	1.0						
$t = 7$.52	.54	.60	.69	.75	.84	1.0					
$t = 8$.45	.48	.61	.69	.78	.84	.78	1.0				
$t = 9$.51	.51	.56	.65	.73	.77	.69	.75	1.0			
$t = 10$.38	.40	.44	.49	.52	.64	.72	.65	.77	1.0		
age	.16	.13	.01	.05	-.02	-.04	-.04	-.21	-.13	-.12	1.0	