SORT 31 (2) July-December 2007, 181-200

Bayesian joint modelling of the mean and covariance structures for normal longitudinal data

Edilberto Cepeda-Cuervo[1] and Vicente Núñez-Antón[2]

[1] *Universidad Nacional de Colombia,* [2] *Universidad del País Vasco*

Abstract

We consider the joint modelling of the mean and covariance structures for the general antedependence model, estimating their parameters and the innovation variances in a longitudinal data context. We propose a new and computationally efficient classic estimation method based on the Fisher scoring algorithm to obtain the maximum likelihood estimates of the parameters. In addition, we also propose a new and innovative Bayesian methodology based on the Gibbs sampling, properly adapted for longitudinal data analysis, a methodology that considers linear mean structures and unrestricted covariance structures for normal longitudinal data. We illustrate the proposed methodology and study its strengths and weaknesses by analyzing two examples, the race and the cattle data sets.

MSC: 62F15; 62J05; 62F10; 62P10

Keywords: Antedependence models; Bayes estimation; Fisher scoring; Gibbs sampling

1 Introduction

Continuous longitudinal data consist of repeated measurements on the same subject over time. These measurements are typically correlated and there have been several proposals in the literature to handle stationary or nonstationary correlations and variances, as well as balanced or unbalanced longitudinal data (see, e.g., Laird and Ware, 1982; Diggle *et al.*, 1994 or Zimmerman and Núñez-Antón, 2001).

Address for correspondence: Vicente Núñez-Antón. Departamento de Econometría y Estadística (E.A.III). Facultad de Ciencias Económicas y Empresariales. Universidad del País Vasco-Euskal Herriko Unibertsitatea. Avenida Lehendakari Aguirre 83. E-48015 Bilbao, Spain. Telephone: +34 94 6013749; fax: +34 94 6013754. E-mail: vicente.nunezanton@ehu.es.

Received: September 2007

Accepted: October 2007

In the context of the parametric multivariate regression model for longitudinal data and under normality, the response variable for each of the m experimental units under study, each having n observations over time, is denoted by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in})'$, $i = 1, \ldots, m$. In this way, the $nm \times 1$ response vector $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_m)'$ contains the responses for all subjects under study, and it is assumed that the $\mathbf{Y}_i$'s are independently normally distributed as $N(\boldsymbol{\mu}, \Sigma_i = \sigma^2 \mathbf{I}_n)$, with $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{I}_n$ being the identity matrix of order n. Here, X is the $n \times p$ design matrix containing the set of explanatory variables, and $\boldsymbol{\beta}$ is the set of mean parameters, so that $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$. This model can be formally written as

$$\mathbf{Y}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i, \text{ with } \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \Sigma_i = \sigma^2 \mathbf{I}_n), \tag{1}$$

As is well known, this model assumes that the errors are independently and normally distributed with mean zero and unknown constant variance σ^2. Under the model above, we have that $E(\mathbf{Y}) = \boldsymbol{\mu}^* = (\boldsymbol{\mu}, \ldots, \boldsymbol{\mu})'$ and that $\Sigma_{\mathbf{Y}}$ is a block-diagonal matrix with diagonal matrix elements given by $\Sigma_i = \sigma^2 \mathbf{I}_n$, $i = 1, \ldots, m$.

If ϵ_{ij} and ϵ_{ik}, $j \neq k$, $i = 1, \ldots, m$, are not independent, then $Var(\boldsymbol{\epsilon}_i) = \Sigma_i$ is no longer a diagonal matrix and it would be necessary to model and estimate the off-diagonal elements of the covariance matrix. This modelling approach usually requires to impose some constraints on the elements of Σ_i to guarantee its positive definiteness. For example, in stationary Gaussian processes, such as the ones used in Geostatistics, the covariance between two observations is explicitly determined by their correlation function. More specifically, it is modelled as a function of the (Euclidean) distance between these two observations. Moreover, and given that some of the properties of this function are imposed by its spatial structure, only correlation functions belonging to the families where these requirements hold can be considered (see, e.g., Diggle and Verbyla, 1998, or Stein, 1999).

Longitudinal data typically consist of several measurements taken over time in each of the experimental units in the sample. It falls into the framework of correlated observations on the same subject and/or experimental unit, and it requires the specification and estimation of both the mean and the covariance structures. Most of the parametric approaches have concentrated on normal linear models (see, e.g., Pourahmadi, 1999 and 2000, or Zimmerman and Núñez-Antón, 2001). A central idea to be able to efficiently estimate the covariance matrix was first introduced by Macchiavelli and Arnold (1994) and Macchiavelli and Moser (1997) and it was based on its Cholesky decomposition. This approach has been used for several joint modelling proposals for the mean and covariance structures in the context of longitudinal data (see, e.g., Pourahmadi, 1999 and 2000, Pan and MacKenzie, 2003, 2006 and 2007 or Pan and Ye, 2006). Zimmerman and Núñez-Antón (1997) and Zimmerman, Núñez-Antón and El Barmi (1998) also proposed a joint modelling of the mean and covariance structures, and Núñez-Antón and Zimmerman (2000) addressed the possibility of having random coefficients and other alternative nonstationary models in a joint modelling proposal for

the mean and covariance structures in the context of longitudinal data. In addition, there have been only a few proposals within the Bayesian framework (see, e.g., Cepeda, 2001, Cepeda and Gamerman, 2000 and 2004, Daniels and Pourahmadi, 2002, or Pourahmadi and Daniels, 2002) and all of them proposed specific and restricted parametric structures for the mean, the innovation variances and the autoregressive parameters in the model. Cepeda and Gamerman (2000) proposed a Bayesian methodology for modeling mean and variance heterogeneity, using normal prior distributions for both the mean and variance parameters in the regression model. Cepeda (2001), also using normal prior distributions, extended this methodology to allow for a joint modelling of the mean and covariance structures. These latter results are included in Cepeda and Gamerman (2004). Independently, Daniels and Pourahmadi (2002) and Pourahmadi and Daniels (2002), also proposed the use of normal prior distributions for both the mean and covariance parameters, but they did not include any explicit algorithm to fit the joint mean and covariance model and use a data set (Pourahmadi and Daniels, 2002) and simulations (Daniels and Pourahmadi, 2002) to illustrate their proposals. Moreover, their proposals focused on modelling the covariance structure and did not include simulations or applications where there was a joint modelling approach proposal for the mean and covariance structure. In this paper, we consider the general antedependence model (Gabriel, 1962, Macchiavelli and Arnold, 1994, or Zimmerman and Núñez-Antón, 1997), and propose a joint modelling approach for the mean and covariance structures, estimating the mean and autoregressive parameters, and the innovation variances in the longitudinal data context. This general model does not impose any specific or restricted parametric structure on the innovation variances and autoregressive parameters, as was the case in previous proposal. We initially consider a new and computationally efficient classical estimation algorithm based on the Fisher scoring algorithm to obtain the estimators of the parameters. This proposal is very convenient and appealing in many cases, especially in the ones where the number of observational units in the study is large, such as in the examples used here to illustrate our proposals (i.e., the race data and the cattle data). In these cases there is a better agreement between sample regressograms and fitted autoregressive parameters and innovation variances, resulting in a better estimation of the parameters in the mean structure. In addition, we also propose a new and innovative Bayesian methodology based on the Gibbs Sampling (Geman and Geman, 1984), properly adapted for longitudinal data analysis, a methodology that considers linear mean structures and unrestricted covariance structures for normal longitudinal data. This methodology allows the researcher to be able to incorporate relevant prior information in the data analysis, as well as to obtain the parameter estimates when the number of observational units in the study is small. In this specific case, we can also estimate credibility intervals for the parameters of interest in the model.

We illustrate the proposed methodology and study its relative strengths and weaknesses when compared to other proposed methods by analyzing two examples,

the race and the cattle data sets. Moreover and for the cases where no prior information is available to be implemented in our Bayesian methodology proposal, we also include the methodology for the possibility of using noninformative priors. The comparison of the results obtained with the classic methodology proposal and with the noninformative priors Bayesian proposal allows us to be able to evaluate their efficiency. As will be seen in the examples presented here, the estimates obtained under these two alternative proposals are very similar.

The paper is organized as follows. Section 2 introduces the general model used in the context of longitudinal data analysis. In Section 3 we introduce the proposed classic methodology for this type of data. Section 4 includes the proposed Bayesian methodology, which is finally applied to the race and cattle data sets in Section 5. Section 6 includes some general conclusions.

2 The General Model

As we have already indicated, one of the main issues in the modelling approach we propose requires $Var(\boldsymbol{\epsilon}_i) = \Sigma_i$, $i = 1, \ldots, m$, to be nonnegative definite so that its inverse can be efficiently calculated and, in addition, it should also be allowed to have a general form so that it is not too restrictive. Pourahmadi (1999) proposed a general approach where all of these conditions hold. Note that, since observations on different subjects are assumed to be independent and, thus, only within-subject covariance structures need to be considered, we suppress the subscript i (identifying the subjects) when describing these structures. More specifically and following the general model settings presented in Cepeda and Gamerman (2004), let us consider the general antedependence model (Gabriel, 1962 or Zimmerman and Núñez-Antón, 1997), where for a given individual having n observations we have that:

$$Y_{ij} - \mu_j = \sum_{k=1}^{j-1} \phi_{jk}(Y_{ik} - \mu_k) + \nu_j, \qquad \nu_j \sim N(0, \sigma_j^2), \qquad (2)$$
$$i = 1, \ldots, m, \qquad j = 1, \ldots, n,$$

and that $E(Y_{ij}) = \mu_j$, where μ_j is assumed to be a linear function of the vector of parameters $\boldsymbol{\beta}$. In addition, the ν_j's are assumed to be mutually independent and, by convention, we set all empty sums to zero, that is $\sum_{k=1}^{0} z_k = 0$. In this way, (2) can be rewritten in matrix form as

$$\boldsymbol{\nu} = \mathbf{T}(\mathbf{Y}_i - \boldsymbol{\mu}), \boldsymbol{\nu} \sim N(\mathbf{0}, \mathbf{D}), \text{ and } \mathbf{D} = diag(\sigma_1^2, \ldots, \sigma_n^2), \qquad (3)$$

where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$, $\mathbf{T} = \{t_{ij}\}_{i=1,\ldots,n}^{j=1,\ldots,n}$, with

$$t_{ij} = \begin{cases} 1 & \text{if } j = i \\ -\phi_{ij} & \text{if } j < i \\ 0 & \text{otherwise} \end{cases}$$

and

$$Var(\boldsymbol{\nu}) = \mathbf{D} = \mathbf{T}\, Var(\mathbf{Y}_i - \boldsymbol{\mu})\, \mathbf{T}' = \mathbf{T}\Sigma_i\mathbf{T}' = \mathbf{T}\Sigma\mathbf{T}' \tag{4}$$

As a direct result of equations (3) and (4), Σ can be indirectly calculated by computing $\mathbf{D}$ and $\mathbf{T}$. In addition, we should point out that the triangular decomposition in equation (4) is unique. Moreover, given that Σ is a symmetric matrix if and only if there exists a unique lower triangular matrix $\mathbf{T}$, with ones in the diagonal, and a unique diagonal matrix $\mathbf{D}$ with positive diagonal entries such that $\mathbf{T}\Sigma\mathbf{T}' = \mathbf{D}$, we also have that Σ is positive definite (Pourahmadi, 1999). Therefore, from (3), we have that

$$\tilde{\mathbf{Y}}_i = (\mathbf{I}_n - \mathbf{T})\tilde{\mathbf{Y}}_i + \boldsymbol{\nu} = \Phi\tilde{\mathbf{Y}}_i + \boldsymbol{\nu},\ i = 1, \ldots, m, \tag{5}$$

where $\tilde{\mathbf{Y}}_i = (\mathbf{Y}_i - \boldsymbol{\mu})$, and the k-th row of the $n \times n$ matrix $\Phi = (\mathbf{I}_n - \mathbf{T})$ contains $[n - (k-1)]$ zeroes and the $(k-1)$ components of the autoregressive parameter vector $\boldsymbol{\phi}_k = (\phi_{k1}, \ldots, \phi_{k,k-1})'$, $k = 2, \ldots, n$.

3 Classic Methodology

Under the assumption that $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in})' \sim$ *i.i.d.* $N(\boldsymbol{\mu}, \Sigma)$, $i = 1, \ldots, m$, where $\boldsymbol{\mu}$ is assumed to depend linearly on $\boldsymbol{\beta}$ (i.e., $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{X}$ being the $n \times p$ design matrix), and $\Sigma^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{T}$, the likelihood function is given by

$$L(\boldsymbol{\beta}, \Phi, \mathbf{D}|\mathbf{Y}) \propto |\mathbf{D}|^{-m/2} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu}^*)'\Sigma_{\mathbf{Y}}^{-1}(\mathbf{Y} - \boldsymbol{\mu}^*)\right\},$$

where $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_m)' \sim N(\boldsymbol{\mu}^*, \Sigma_{\mathbf{Y}})$ and $|\Sigma| = |\mathbf{T}'|\,|\mathbf{D}|\,|\mathbf{T}| = |\mathbf{D}|$. Note that in the equation above, $\boldsymbol{\mu}^* = (E(\mathbf{Y}_1), \ldots, E(\mathbf{Y}_m))' = (\boldsymbol{\mu}, \ldots, \boldsymbol{\mu})'$ and $\Sigma_{\mathbf{Y}}$ is a block diagonal matrix with diagonal matrix elements given by Σ, so that $\Sigma_{\mathbf{Y}}^{-1}$ is a block diagonal matrix with diagonal matrix elements given by $\Sigma^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{T}$.

Therefore, the log-likelihood function can be written as $\ell(\boldsymbol{\beta}, \Phi, \mathbf{D}|\mathbf{Y}) = \log L(\boldsymbol{\beta}, \Phi, \mathbf{D}|\mathbf{Y}) \propto -m \log|\mathbf{D}| - (\mathbf{Y} - \boldsymbol{\mu}^*)'\Sigma_{\mathbf{Y}}^{-1}(\mathbf{Y} - \boldsymbol{\mu}^*)$, so that the components of the corresponding score function are given by

$$\begin{aligned}\frac{\partial \ell}{\partial \boldsymbol{\beta}} &= -\mathbf{X}^{*\prime}\Sigma_{\mathbf{Y}}^{-1}(\mathbf{Y}-\mathbf{X}^{*}\boldsymbol{\beta}) \\ \frac{\partial \ell}{\partial \phi_{ij}} &= -\frac{1}{2}(\mathbf{Y}-\mathbf{X}^{*}\boldsymbol{\beta})'\left(\frac{\partial \Sigma_{\mathbf{Y}}^{-1}}{\partial \phi_{ij}}\right)(\mathbf{Y}-\mathbf{X}^{*}\boldsymbol{\beta}) \\ & \quad (i=1,\ldots,n,\ j=1,\ldots,i-1),\end{aligned}$$

where $\mathbf{X}^{*} = (X',\ldots,X')'$ is the $nm \times p$ design matrix. Thus, we have that

$$\begin{aligned}I_{\boldsymbol{\beta},\phi} &= E\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \phi}\right) = 0 \\ I_{\boldsymbol{\beta},\sigma^2} &= E\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \sigma^2}\right) = 0\end{aligned}$$

Now, given that the log-likelihood function can be written as $\ell = \log L \propto -m\log|\mathbf{D}| - \frac{1}{\sigma_1^2}\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^{*} - \cdots - \frac{1}{\sigma_n^2}(\mathbf{Y}_n^{*}-\tilde{\boldsymbol{\mu}}_n)(\mathbf{Y}_n^{*}-\tilde{\boldsymbol{\mu}}_n)$, where $\mathbf{Y}_1^{*}$ is the m-dimensional vector with i-th component given by $(Y_{i1}-\mu_1)$, $\mathbf{Y}_k^{*}$ $(k=2,\ldots,n)$ is the m-dimensional vector with i-th component given by $(Y_{ik}-\mu_k)$, and $\tilde{\boldsymbol{\mu}}_k$, $k=2,\ldots,n$, is the m-dimensional vector with i-th component given by $\phi_{k1}(Y_{i1}-\mu_1)+\cdots+\phi_{k,k-1}(Y_{i,k-1}-\mu_{k-1})$, we can write

$$\begin{aligned}\frac{\partial \ell}{\partial \mu_1} &= \frac{1}{\sigma_1^2}\sum_{i=1}^{m}(Y_{i1}-\mu_1) \\ \frac{\partial \ell}{\partial \sigma_1^2} &= -\frac{m}{2\sigma_1^2}+\frac{1}{2\sigma_1^4}\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^{*}\end{aligned}$$

Therefore, the maximum likelihood estimators of μ_1 and σ_1^2 are given by $\hat{\mu}_1 = \frac{1}{m}\sum_{i=1}^{m} Y_{i1}$ and $\hat{\sigma}_1^2 = \frac{1}{m}\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^{*}$. Now, for $\boldsymbol{\phi}_k = (\phi_{k1},\ldots,\phi_{k,k-1})$ and σ_k^2, $k=2,\ldots,n$, and if we let $\tilde{\mathbf{X}}_k$ be an $m\times(k-1)$ matrix with columns given by $\mathbf{Y}_1^{*},\ldots,\mathbf{Y}_{k-1}^{*}$, we have that

$$\begin{aligned}\frac{\partial \ell}{\partial \boldsymbol{\phi}_k} &= \frac{1}{\sigma_1^2}\tilde{\mathbf{X}}_k'(\mathbf{Y}_k^{*}-\tilde{\mathbf{X}}_k\boldsymbol{\phi}_k) \\ \frac{\partial \ell}{\partial \sigma_k^2} &= -\frac{m}{2\sigma_k^2}+\frac{1}{2\sigma_k^4}(\mathbf{Y}_k^{*}-\tilde{\boldsymbol{\mu}}_k)'(\mathbf{Y}_k^{*}-\tilde{\boldsymbol{\mu}}_k)\end{aligned}$$

Thus, the maximum likelihood estimators of $\boldsymbol{\phi}_k$ and σ_k^2 $(k=2,\ldots,n)$ are given by

$$\begin{aligned}\hat{\boldsymbol{\phi}}_k &= (\tilde{\mathbf{X}}_k'\tilde{\mathbf{X}}_k)^{-1}(\tilde{\mathbf{X}}_k'\mathbf{Y}_k^{*}) \\ \hat{\sigma}_k^2 &= \frac{1}{m}(\mathbf{Y}_k^{*}-\tilde{\boldsymbol{\mu}}_k)'(\mathbf{Y}_k^{*}-\tilde{\boldsymbol{\mu}}_k)\end{aligned} \tag{6}$$

The steps of the algorithm used to obtain the maximum likelihood estimators of both the mean and variance parameters follow:

1. Set some arbitrary initial values for $\boldsymbol{\phi}_k$ and positive initial values for σ_k^2, $k = 1, \ldots, n$.
2. Compute Σ and update $\boldsymbol{\beta}$ by using $\dfrac{\partial \ell(\boldsymbol{\beta}, \Phi, \mathbf{D}|\mathbf{Y})}{\partial \boldsymbol{\beta}} = 0$.
3. Update $\boldsymbol{\phi}_k$ and σ_k^2, by solving the equations $\dfrac{\partial \ell(\boldsymbol{\beta}, \Phi, \mathbf{D}|\mathbf{Y})}{\partial \boldsymbol{\phi}_k} = 0$ and $\dfrac{\partial \ell(\boldsymbol{\beta}, \Phi, \mathbf{D}|\mathbf{Y})}{\partial \sigma_k^2} = 0$, $k = 1, \ldots, n$.
4. Repeat steps 2 and 3 until convergence.

4 Bayesian Methodology

As in the classic approach, we also assume that $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in})'$ *i.i.d.* $\sim N(\boldsymbol{\mu}, \Sigma)$, $i = 1, \ldots, m$, where $\boldsymbol{\mu}$ is assumed to depend linearly on $\boldsymbol{\beta}$, and $\Sigma^{-1} = \mathbf{T}' \mathbf{D}^{-1} \mathbf{T}$. Therefore, the likelihood function is given by

$$L(\boldsymbol{\beta}, \Phi, \mathbf{D}|\mathbf{Y}) \propto |\mathbf{D}|^{-m/2} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu}^*)' \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \boldsymbol{\mu}^*)\right\},$$

where $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_m)'$, $|\Sigma| = |\mathbf{T}'|\,|\mathbf{D}|\,|\mathbf{T}| = |\mathbf{D}|$ and $\Phi = (\mathbf{I}_n - \mathbf{T})$.

If we now let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \Phi, \mathbf{D})'$, under the Bayesian approach and in order to obtain the posterior distribution for the parameters, we need to assume a prior distribution $P(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$. Without loss of generality and for simplicity, we assume independent prior distributions such that $\boldsymbol{\beta} \sim N(\mathbf{b}_0, \mathbf{B})$, $\boldsymbol{\phi}_k \sim N(\mathbf{l}_{0k}, \Sigma_{\boldsymbol{\phi}_k})$, $\psi_1 = 1/\sigma_1^2 \sim G(\alpha_1, \lambda_1)$ and $\psi_k = 1/\sigma_k^2 \sim G(\alpha_k, \lambda_k)$ $(k = 2, \ldots, n)$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$, $\Psi = (\psi_1, \ldots, \psi_n)'$, and $G(r, s)$, represents the gamma distribution with parameters $r > 0$ and $s > 0$. As usual, another possibility for the prior distribution for $\boldsymbol{\theta}$ could be to assume a noninformative prior distribution such as, for example, assume Jeffreys prior distributions.

From Bayes' theorem, the posterior conditional distribution for $\boldsymbol{\beta}$, $\pi_{\boldsymbol{\beta}} = \pi(\boldsymbol{\beta}|\Phi, \Psi, \mathbf{Y})$, is given by

$$\begin{aligned} \pi(\boldsymbol{\beta}|\Phi, \Psi, \mathbf{Y}) &\propto \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}^*\boldsymbol{\beta})' \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \mathbf{X}^*\boldsymbol{\beta}) - \frac{1}{2}(\boldsymbol{\beta} - \mathbf{b}_0)' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}_0)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{b}^*)' \mathbf{B}^{*-1} (\boldsymbol{\beta} - \mathbf{b}^*)\right\}, \end{aligned} \tag{7}$$

where $\mathbf{b}^* = \mathbf{B}^*(\mathbf{B}^{-1}\mathbf{b}_0 + \mathbf{X}^{*\prime}\Sigma_{\mathbf{Y}}^{-1}\mathbf{Y})$ and $\mathbf{B}^* = (\mathbf{B}^{-1} + \mathbf{X}^{*\prime}\Sigma_{\mathbf{Y}}^{-1}\mathbf{X}^*)^{-1}$. Therefore, we have that $\pi_{\boldsymbol{\beta}} = \pi(\boldsymbol{\beta}|\Phi, \Psi, \mathbf{Y}) \sim N(\mathbf{b}^*, \mathbf{B}^*)$ and, thus, it would be possible to sample $\boldsymbol{\beta}$ directly from $\pi_{\boldsymbol{\beta}}$. That is, values of $\boldsymbol{\beta}$ can be proposed directly from $\pi_{\boldsymbol{\beta}}$ and accepted with probability one. This is the basic description and motivation for the Gibbs sampler (Geman and Geman, 1984).

Now, given that $\Sigma^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{T}$, we have that, for $i = 1, \ldots, m$,

$$\tilde{\mathbf{Y}}_i'\mathbf{T}'\mathbf{D}^{-1}\mathbf{T}\tilde{\mathbf{Y}}_i \quad = \quad [(\mathbf{I}_n - \Phi)\tilde{\mathbf{Y}}_i]'\mathbf{D}^{-1}[(\mathbf{I}_n - \Phi)\tilde{\mathbf{Y}}_i]$$

Therefore, by taking into account the independence between individuals and using the equation above, the quadratic form appearing in the log-likelihood function, $Q(\mathbf{Y}) = (\mathbf{Y} - \boldsymbol{\mu}^*)'\Sigma_{\mathbf{Y}}^{-1}(\mathbf{Y} - \boldsymbol{\mu}^*)$, can be rewritten as

$$Q(\mathbf{Y}) \quad = \quad \frac{1}{\sigma_1^2}\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^* + \cdots + \frac{1}{\sigma_n^2}(\mathbf{Y}_n^* - \tilde{\boldsymbol{\mu}}_n)'(\mathbf{Y}_n^* - \tilde{\boldsymbol{\mu}}_n),$$

where $\tilde{\boldsymbol{\mu}}_k$, $k = 2, \ldots, n$ is the m-dimensional vector with i-th component given by $\phi_{k1}(Y_{i1} - \mu_1) + \cdots + \phi_{k,k-1}(Y_{i,k-1} - \mu_{k-1})$, $\mathbf{Y}_1^* = (Y_{11} - \mu_1, Y_{21} - \mu_1, \ldots, Y_{m1} - \mu_1)'$ is the vector of centered observations for all m individuals at the first time point; $\mathbf{Y}_2^* = (Y_{12} - \mu_2, Y_{22} - \mu_2, \ldots, Y_{m2} - \mu_2)'$, is the vector of centered observations for all m individuals at the second time point; and $\mathbf{Y}_n^* = (Y_{1n} - \mu_n, Y_{2n} - \mu_n, \ldots, Y_{mn} - \mu_n)'$ is the vector of centered observations for all m individuals at the n-th time point.

In this way, the maximum likelihood function can be written as:

$$L(\boldsymbol{\beta}, \Phi, \mathbf{D}) \propto |\mathbf{D}|^{-m/2} \exp\left\{-\frac{1}{2\sigma_1^2}\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^* - \cdots - \frac{1}{2\sigma_n^2}(\mathbf{Y}_n^* - \tilde{\boldsymbol{\mu}}_n)'(\mathbf{Y}_n^* - \tilde{\boldsymbol{\mu}}_n)\right\} \tag{8}$$

Thus, if we assume independent normal prior for the $\boldsymbol{\phi}_k$'s, we can obtain, from the application of Bayes' theorem, that the posterior full conditional distribution for $\boldsymbol{\phi}_k$ is given by

$$\pi(\boldsymbol{\phi}_k|\boldsymbol{\beta}, \mathbf{D}, \Phi_{-k}) \propto \sigma_k^{-1} \exp\Big\{-\frac{1}{2\sigma_k^2}(\mathbf{Y}_k^* - \tilde{\boldsymbol{\mu}}_k)'(\mathbf{Y}_k^* - \tilde{\boldsymbol{\mu}}_k) - \frac{1}{2}(\boldsymbol{\phi}_k - \mathbf{l}_{0k})'\Sigma_{\boldsymbol{\phi}_k}^{-1}(\boldsymbol{\phi}_k - \mathbf{l}_{0k})\Big\}, \tag{9}$$

where Φ_{-k} represent the parameters in Φ excluding the corresponding ones for $\boldsymbol{\phi}_k$. Therefore, the conditional posterior distribution is given by

$$\tilde{\pi}(\boldsymbol{\phi}_k|\boldsymbol{\alpha}, \boldsymbol{\lambda}) \quad \propto \quad \exp\left\{-\frac{1}{2}(\boldsymbol{\phi}_k - \mathbf{l}_k^*)'\Sigma_k^{*-1}(\boldsymbol{\phi}_k - \mathbf{l}_k^*)\right\},$$

where $\mathbf{l}_k^* = \Sigma_k^*\left(\Sigma_{\boldsymbol{\phi}_k}\mathbf{l}_{0k} + \frac{1}{\sigma_k^2}\tilde{\mathbf{X}}_k'\tilde{\mathbf{Y}}_k\right)$ and $\Sigma_k^* = \left(\Sigma_{\boldsymbol{\phi}_k}^{-1} + \frac{1}{\sigma_k^2}\tilde{\mathbf{X}}_k'\tilde{\mathbf{X}}_k\right)^{-1}$. From the above, we have that

$$\pi_{\boldsymbol{\phi}_k} = \pi(\boldsymbol{\phi}_k|\boldsymbol{\beta}, \mathbf{D}, \Phi_{-k}, \mathbf{Y}) \sim N(\mathbf{l}_k^*, \Sigma_k^*) \tag{10}$$

So it is possible to sample $\boldsymbol{\phi}_k$ directly from $\pi_{\boldsymbol{\phi}_k}$. Values of $\boldsymbol{\phi}_k$ can be proposed directly from $\pi_{\boldsymbol{\phi}_k}$ and accepted with probability 1. This is the basic description and motivation for the Gibbs sampler (Geman and Geman, 1984).

Finally, to be able to sample σ_k^2, $k = 1, 2, 3, \ldots, n$, as we have already seen before, we propose the use of gamma priors for the ψ_k's and, thus, from the straight application

of Bayes' theorem, we obtain gamma posterior distributions for the ψ_k's, so that the sampling procedure can be easily handled by using the Gibbs sampler.

More concretely, let $\tilde{\mathbf{Y}}_k^* = (\mathbf{Y}_k^* - \tilde{\boldsymbol{\mu}}_k)$ be a random sample of size m from a $N(\mathbf{0}, \sigma_k^2)$ distribution ($k = 1, \ldots, n$), with $\psi_k = 1/\sigma_k^2$. That is, $\tilde{\mathbf{Y}}_k^*$ represents a sample of m individuals at time t_k, $k = 1, \ldots, n$. Now, given that the gamma family is closed under sampling, we can assume a gamma prior distribution so that $\psi_k \sim G(\alpha_k = n_{0k}/2, n_0\sigma_{0k}^2)$, where n_{0k} is a natural number and $\sigma_{0k}^2 > 0$. Therefore, the posterior distribution of ψ_k can be directly obtained by using Bayes' theorem, so that

$$\pi(\psi_k|\boldsymbol{\beta}, \mathbf{D}, \tilde{\mathbf{Y}}_k^*) \propto \psi_k^{[(n_{0k}+m)/2]-1} \exp\{-(n_{0k}\sigma_{0k}^2 + ms_{0k}^2)\psi_k/2\} \tag{11}$$

This expression corresponds to the kernel of the gamma distribution. Thus, we have that

$$\pi(\psi_k|\boldsymbol{\beta}^{(c)}, \boldsymbol{\phi}^{(c)}) = G\Big(\frac{n_{0k}+m}{2}, \frac{n_{0k}\sigma_{0k}^2 + ms_{0k}^2}{2}\Big),$$

where $s_{0k}^2 = \frac{1}{m}\Sigma_{i=1}^m \tilde{\mathbf{Y}}_k^{*\prime}\tilde{\mathbf{Y}}_k^*$, $k = 1, \ldots, n$.

If we decide to assume constant or noninformative priors, the sampling procedure for each of the parameters involved in the estimation process is described below.

Given $\mathbf{D}$, Φ and a constant prior distribution for $\boldsymbol{\beta}$, we can sample $\boldsymbol{\beta}$ from

$$\pi(\boldsymbol{\beta}|\Phi, \Psi, \mathbf{Y}) \propto \left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{b}^*)'\mathbf{B}^{*-1}(\boldsymbol{\beta} - \mathbf{b}^*)\right\}, \tag{12}$$

where $\mathbf{b}^* = \mathbf{B}^*(\mathbf{X}^{*\prime}\Sigma_{\mathbf{Y}}^{-1}\mathbf{Y})$ and $\mathbf{B}^* = (\mathbf{X}^{*\prime}\Sigma_{\mathbf{Y}}^{-1}\mathbf{X}^*)^{-1}$.

Given $\boldsymbol{\beta}$, $\mathbf{D}$ and a constant prior distribution for the $\boldsymbol{\phi}_k$'s, and letting Φ_{-k} represent the parameters in Φ excluding the corresponding ones for $\boldsymbol{\phi}_k$, we can sample the $\boldsymbol{\phi}_k$'s ($k = 1, 2, \ldots, n$) from (9) by following the procedure below:

1. Sample $\boldsymbol{\phi}_1$ from $\pi(\boldsymbol{\phi}_1|\boldsymbol{\beta}, \mathbf{D}, \Phi_{-1}) \propto \sigma_1^{-1} \exp\left\{-\frac{1}{2\sigma_1^2}\mathbf{Y}_1^{*\prime}\mathbf{Y}_1^*)\right\}$.
2. Sample $\boldsymbol{\phi}_2$ from $\pi(\boldsymbol{\phi}_2|\boldsymbol{\beta}, \mathbf{D}, \Phi_{-2}) \propto \sigma_2^{-1} \exp\left\{-\frac{1}{2\sigma_2^2}(\mathbf{Y}_2^* - \tilde{\boldsymbol{\mu}}_2)'(\mathbf{Y}_2^* - \tilde{\boldsymbol{\mu}}_2)\right\}$.
3. And so on, up to sample $\boldsymbol{\phi}_n$ from

$$\pi(\boldsymbol{\phi}_n|\boldsymbol{\beta}, \mathbf{D}, \Phi_{-n}) \propto \sigma_n^{-1} \exp\left\{-\frac{1}{2\sigma_n^2}(\mathbf{Y}_n^* - \tilde{\boldsymbol{\mu}}_n)'(\mathbf{Y}_n^* - \tilde{\boldsymbol{\mu}}_n)\right\}.$$

Finally, given $\mathbf{D}, \boldsymbol{\beta}$ and a constant prior for the ψ_k's, we can sample the ψ_k's from

$$\pi(\psi_k|\boldsymbol{\beta}^{(c)}, \boldsymbol{\phi}^{(c)}) = G\Big(\frac{m}{2}, \frac{ms_{0k}^2}{2}\Big),$$

where s_{0k}^2 is defined as before.

5 Examples

In order to illustrate and motivate the methods proposed in this paper, we present the analysis of two data sets, referred to here as the race data and the cattle data. Previous approaches for analyzing these data sets have been reported elsewhere (Zimmerman and Núñez-Antón, 1997; Zimmerman *et al.*, 1998; Pourahmadi, 1999, 2000 and 2002; Zimmerman, 2000; Núñez-Antón and Zimmerman, 2000; Wu and Pourahmadi, 2003; Pan and MacKenzie, 2003, 2006 and 2007).

The race data consist of the "split" times for each of $m = 80$ competitors in each 10-km section of a 100-km race held in 1984 in the United Kingdom. The data include, in addition to the split times, the ages of all but four of the competitors. Measurement times are evenly spaced and common to all subjects in the study and no observations were missing. A previous analysis (Zimmerman *et al.*, 1998) showed the age variable to be non-significant, so we will ignore it here. Our analysis centres on the study of how competitor performance on each 10-km section is related to the section number ($t = 1, 2, \ldots, 10$) (i.e., $n = 10$), as well as on the adequate modelling of the variances and the correlations between successive measurements being made on any given subject.

The cattle data (Kenward, 1987) come from a designed experiment in which cattle receiving two treatments, say A and B, for intestinal parasites were weighed $n = 11$ times over a 133-day period. Thirty animals received treatment A and thirty received treatment B (i.e., $m = 60$). The first 10 measurements on each animal were made at two-week intervals and the final measurement was made after a one-week interval. Measurement times were common across animals and rescaled to $t = 1, \ldots, 9, 10, 10.5$, and no observations were missing. We wish to study how cattle growth is affected by the different treatments and, in addition, we concentrate on how the mean changes with time, as well as on the adequate modelling of the variances and the correlations between successive measurements being made on any given experimental unit.

5.1 Analysis of the Race Data

Figure 1 shows the profile plot for the race data. The profile plot indicates that the mean split time tends to increase over the first 80 km of the race but then levels off (perhaps reflecting the "kick" that well-conditioned runners generally show near the end of a race). Figure 1 also shows that variances tend to increase over the course of the race, and that the behaviour of many runners in the later sections of the race is more erratic, in the sense that consecutive same-runner split times fluctuate more. The increase of the mean does not seem to be linear with time and, thus, a quadratic or cubic model in time may be more appropriate to model the overall mean growth in this data set. Based on this and, as in Zimmerman *et al.* (1998), we use a cubic in time model for the overall mean weight. That is, the overall mean weight, as a function of the section number t,

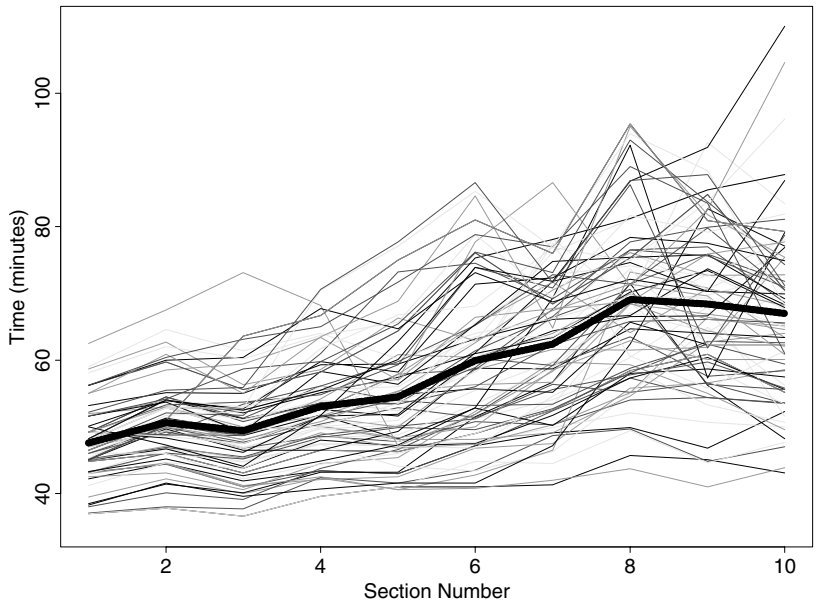

Figure 1: *Profile plot for the race data. The thicker line indicates the mean profile for all individuals in the study.*

will be given by:

$$\mu_j = \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + \beta_3 t_j^3, \quad j = 1, \ldots, 10,$$

where

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_{10})' = (\mu_{\{t_1=1\}}, \ldots, \mu_{\{t_{10}=10\}})'$$

and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$.

We also need to estimate the variance-covariance matriz Σ. In order to do so, and as a direct result from equations (3) and (4), both the elements of the diagonal matrix $\mathbf{D}$ and the ϕ_{ij}'s in the lower triangular matrix $\mathbf{T}$ need to be estimated. In order to have a not too restrictive and a more general procedure, we consider noninformative prior distributions in all mean and covariance parameters. As it is well known and as we will see later on in this section, with the use of noninformative prior information, the parameter estimates (both for the mean and variance-covariance matrix) obtained with the proposed Bayesian methodology are very similar to the ones obtained with the proposed classic methodology.

The estimates for the mean parameters obtained with the proposed Bayesian methodology, together with their standard deviations in parenthesis, are

$$\boldsymbol{\beta} = (46.320(0.711), 1.267(0.483), 0.556(0.167), -0.047(0.014))',$$

Table 1: *Estimated means, variances and correlations for the race data set with the Bayesian methodology. In the table, Section refers to the section number.*

Section	1	2	3	4	5	6	7	8	9	10
Correlations										
	1.0									
	.94	1.0								
	.75	.72	1.0							
	.72	.70	.92	1.0						
	.49	.44	.76	.87	1.0					
	.57	.54	.72	.84	.91	1.0				
	.47	.43	.60	.69	.75	.84	1.0			
	.49	.53	.54	.64	.66	.80	.72	1.0		
	.53	.50	.55	.65	.69	.76	.68	.75	1.0	
	.37	.35	.47	.51	.54	.65	.72	.62	.77	1.0
Means	48.1	50.7	53.9	57.3	60.7	63.8	66.3	68.0	68.5	67.6
Variances	27.0	39.7	48.5	56.2	94.9	142.1	105.6	154.7	144.2	165.6

and the ones obtained with the proposed classic approach are

$$\boldsymbol{\beta} = (46.247, 1.258, 0.564, -0.048)'.$$

Given that, as expected, the estimates obtained with the proposed classic methodology are very similar to the ones obtained with the Bayesian methodology, we do not include their estimated standard deviations. Note that all coefficients in the linear model are different from zero at a 95% credibility level and that the signs of the coefficients are consistent with the behaviour seen in Figure 1. In addition, our estimates are also similar to the ones obtained by Zimmerman *et al.* (1998) with the difference that our variance-covariance structure is more general than theirs because they have assumed a structured antedependent model of order one and we have a general unstructured antedependent model. As a way of comparing the estimated mean time as a function of the section number, as in Figure 1, Table 1 shows these estimated values using the estimates obtained from the Bayesian approach. As can be seen, this behaviour is consistent with the one observed in Figure 1.

The estimated innovation variances obtained with the proposed Bayesian methodology under noninformative priors with their corresponding estimated standard deviations in parenthesis, as well as those obtained with the proposed classic methodology are

Table 2: *Estimated autoregressive parameters obtained with the proposed Bayesian methodology for the race data set. Estimated standard deviations, in parenthesis, are also included. As in the* **T** *matrix (see Section 2), there are ones on the main diagonal.*

1									
1.137	1								
(0.028)									
0.773	0.215	1							
(0.216)	(0.158)								
−0.010	0.105	0.924	1						
(0.076)	(0.091)	(0.043)							
0.293	−0.701	−0.179	1.554	1					
(0.079)	(0.067)	(0.043)	(0.038)						
−0.053	0.517	−0.370	0.100	1.117	1				
(0.103)	(0.100)	(0.084)	(0.063)	(0.028)					
0.200	−0.308	0,221	−0.078	−0.199	0.858	1			
(0.102)	(0.144)	(0.086)	(0.060)	(0.055)	(0.033)				
−0.897	1.028	−0.150	−0.093	−0.197	0.849	0.240	1		
(0.159)	(0.138)	(0.117)	(0.073)	(0.067)	(0.055)	(0.031)			
0.749	−0.261	−0.211	−0.058	0.298	0.156	0.092	0.397	1	
(0.091)	(0.155)	(0.110)	(0.066)	(0.061)	(0.039)	(0.028)	(0.029)		
−0.610	0.073	0.612	−0.283	−0.608	0.308	0.576	−0.118	0.726	1
(0.143)	(0.193)	(0.145)	(0.082)	(0.072)	(0.050)	(0.031)	(0.037)	(0.022)	

$$\begin{pmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \\ \hat{\sigma}_3^2 \\ \hat{\sigma}_4^2 \\ \hat{\sigma}_5^2 \\ \hat{\sigma}_6^2 \\ \hat{\sigma}_7^2 \\ \hat{\sigma}_8^2 \\ \hat{\sigma}_9^2 \\ \hat{\sigma}_{10}^2 \end{pmatrix}_{\text{Bayes}} = \begin{pmatrix} 27.653(4.489) \\ 4.811(1.162) \\ 21.818(4.736) \\ 9.085(1.493) \\ 16.496(2.812) \\ 19.616(3.319) \\ 31.837(5.182) \\ 47.552(7.980) \\ 50.997(8.399) \\ 49.414(8.123) \end{pmatrix}, \quad \begin{pmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \\ \hat{\sigma}_3^2 \\ \hat{\sigma}_4^2 \\ \hat{\sigma}_5^2 \\ \hat{\sigma}_6^2 \\ \hat{\sigma}_7^2 \\ \hat{\sigma}_8^2 \\ \hat{\sigma}_9^2 \\ \hat{\sigma}_{10}^2 \end{pmatrix}_{\text{Classic}} = \begin{pmatrix} 27.980 \\ 4.544 \\ 21.167 \\ 8.803 \\ 16.463 \\ 19.009 \\ 30.946 \\ 46.477 \\ 49.537 \\ 47.683 \end{pmatrix}$$

The estimated values of all innovation variances are somehow consistent with the increasing behaviour of variances seen in Table 1 and Figure 1. However, we must be cautious about this issue because these are not the response variances. We will come back to this matter when reporting the estimated variances and correlations for the race data (see Table 1). Table 2 shows the estimated autoregressive parameters obtained with the Bayesian approach. Moreover, the standard deviations obtained with the proposed

Bayesian approach show that some of the autoregressive parameters are not different from zero at a 95% credibility level. For brevity and because estimates obtained for the autoregressive parameters with the classic approach are similar to those obtained with the Bayesian approach, we do not include them. However, as estimates obtained with the use of the two proposed approaches are quite similar, we can see that some of the autoregressive parameters are not statistically significant at the same level. Finally, Table 1 shows the estimated variances and correlations for the race data obtained with the Bayesian approach. The behaviour of these estimated values is consistent with the one observed in the corresponding sample variances and correlations (not included here for brevity).

5.2 Analysis of the Cattle Data

For brevity, and as in Zimmerman and Núñez-Antón (1997), we report here the results for group A only. Figure 2 shows the profile plot for cattle in group A of the cattle data. The profile plot indicates that the means and variances of the responses are increasing over the course of the experiment, with the more rapid growth occurring in the first few weeks of the study. The increase of the mean does not seem to be linear with time and, thus, a quadratic or cubic model in time may be more appropriate to model the overall mean growth in this cattle group. Based on the analysis above and, as in Pourahmadi

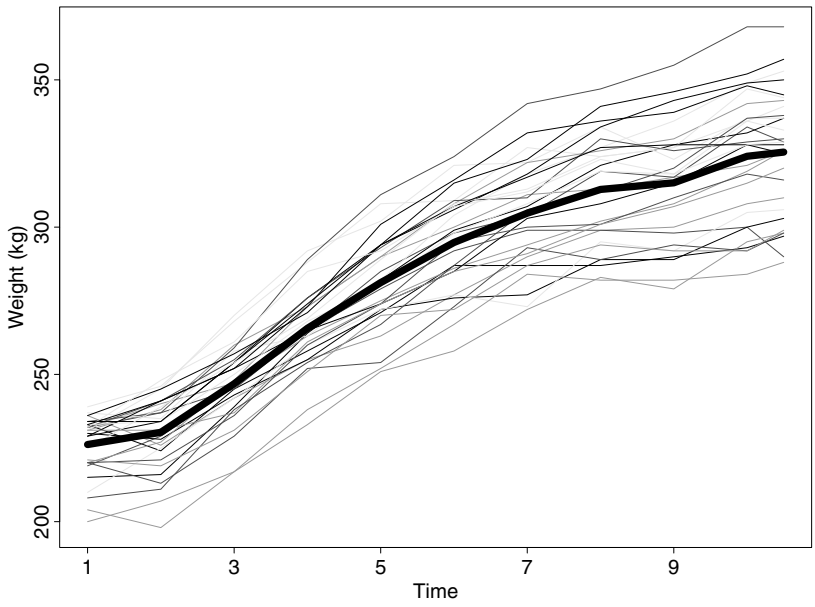

Figure 2: *Profile plot for the cattle data, treatment A. The thicker line indicates the mean profile for all individuals in this group.*

(1999) or Cepeda (2001), we use a cubic in time model for the overall mean weight. That is, the overall mean weight, as a function of the section number t, will be given by:

$$\mu_j = \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + \beta_3 t_j^3, \quad j = 1, \ldots, 11,$$

where

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_{11})' = (\mu_{\{t_1=1\}}, \ldots, \mu_{\{t_{10}=10\}}, \mu_{\{t_{11}=10.5\}})'$$

and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$. As in the analysis of the race data, we estimate the variance-covariance matriz $\boldsymbol{\Sigma}$ from equations (3) and (4), by estimating the elements of the diagonal matrix $\mathbf{D}$ and the ϕ_{ij}'s in the lower triangular matrix $\mathbf{T}$. In this case, we also consider noninformative prior distributions in all mean and covariance parameters. Given the similarity of the estimates obtained from the Bayesian and classic approaches when considering this type of priors, we only report the ones from the Bayesian methodology. The estimates for the mean parameters obtained with the proposed Bayesian methodology, together with their standard deviations in parenthesis, are

$$\boldsymbol{\beta} = (227.575(3.041), 10.003(2.565), 1.154(0.541), -0.117(0.034))'.$$

Note that all coefficients in the linear model are different from zero at a 95% credibility level, and that the signs of the coefficients are consistent with the behaviour seen in Figure 2. In addition, our estimates are also similar to the ones obtained by previous authors that have analyzed this data, with the difference that our variance-

Table 3: *Estimated means, variances and correlations for the group A cattle data set with the Bayesian methodology.*

Time	1	2	3	4	5	6	7	8	9	10	10.5
Correlations											
	1.0										
	.74	1.0									
	.74	.89	1.0								
	.62	.68	.87	1.0							
	.58	.59	.79	.94	1.0						
	.53	.53	.75	.91	.95	1.0					
	.49	.50	.70	.83	.87	.93	1.0				
	.51	.58	.75	.83	.88	.92	.93	1.0			
	.51	.62	.73	.73	.76	.82	.89	.83	1.0		
	.46	.52	.68	.73	.78	.85	.88	.94	.94	1.0	
	.44	.44	.63	.70	.75	.81	.86	.92	.92	.98	1.0
Means	238.62	251.3	264.8	278.6	291.8	303.9	314.0	321.6	325.8	326.0	324.4
Variances	109.5	228.4	194.6	202.5	269.7	317.5	335.1	372.2	452.3	510.5	454.8

covariance structure is more general than theirs. As a way of comparing the estimated cattle weight as a function of time, as in Figure 2, Table 3 shows these estimated values using the estimates obtained from the Bayesian approach. As can be seen, this behaviour is consistent with the behaviour observed in Figure 2.

The estimated innovation variances obtained with the proposed Bayesian methodology under noninformative priors with their corresponding estimated standard deviations in parenthesis are

$$\begin{pmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \\ \hat{\sigma}_3^2 \\ \hat{\sigma}_4^2 \\ \hat{\sigma}_5^2 \\ \hat{\sigma}_6^2 \\ \hat{\sigma}_7^2 \\ \hat{\sigma}_8^2 \\ \hat{\sigma}_9^2 \\ \hat{\sigma}_{10}^2 \\ \hat{\sigma}_{11}^2 \end{pmatrix}_{\text{Bayes}} = \begin{pmatrix} 109.524(30.490) \\ 102.496(42.180) \\ 38.595(15.072) \\ 42.015(13.012) \\ 29.769(8.330) \\ 29.811(8.357) \\ 45.612(12.784) \\ 32.955(9.345) \\ 23.515(6.791) \\ 37.974(11.438) \\ 10.805(3.024) \end{pmatrix}$$

The estimated values of all innovation variances are somehow consistent with the increasing behaviour of variances seen in Table 3 and Figure 2. However, we must be cautious about this issue because these are not the response variances. We will come back to this matter when reporting the estimated variances and correlations for the group A cattle data (see Table 3). If we wish to compare these results with those reported in Pourahmadi (1999) or Cepeda (2001), we could estimate the log-innovation variances. We have done so and estimates are very similar to theirs. Thus, this can be used as a way to indicate that log-innovation variances could be modelled as cubic polynomials (see Pourahmadi, 1999 or Cepeda, 2001). We do not consider it necessary to include these estimated values here.

Table 4 shows the estimated autoregressive parameters with the Bayesian approach. The standard deviations obtained with the proposed Bayesian approach show that the autoregressive parameters are different from zero at a 95% credibility level and they are all consistent with the results presented in Pourahmadi (1999). Finally, Table 3 shows the estimated variances and correlations for the group A cattle data obtained with the Bayesian approach. The behaviour of these estimated values is consistent with the one observed in the corresponding sample variances and correlations (not included here for brevity).

Table 4: *Estimated autoregressive parameters obtained with the proposed Bayesian methodology for the group A cattle data set. Estimated standard deviations, in parenthesis, are also included. As in the* **T** *matrix (see Section 2), there are ones on the main diagonal.*

1										
1.072	1									
(0.172)										
0.236	0.698	1								
(0.110)	(0.097)									
0.062	−0.416	1.251	1							
(0.118)	(0.150)	(0.181)								
0.104	−0.133	−0.015	1.146	1						
(0.050)	(0.071)	(0.059)	(0.068)							
−0.014	−0.200	0.166	0.295	0.789	1					
(0.059)	(0.087)	(0.060)	(0.075)	(0.039)						
−0.027	0.017	0.116	−0.278	0.050	1.044	1				
(0.085)	(0.092)	(0.069)	(0.099)	(0.056)	(0.069)					
−0.182	0.261	0.100	−0.386	0.150	0.538	0.516	1			
(0.069)	(0.110)	(0.059)	(0.087)	(0.060)	(0.053)	(0.063)				
−0.011	0.208	0.166	−0.362	−0.216	−0.204	0.416	1.053	1		
(0.077)	(0.058)	(0.072)	(0.062)	(0.055)	(0.067)	(0.037)	(0.067)			
0.052	−0.306	0.226	−0.147	0.007	0.112	−0.122	0.562	0.617	1	
(0.099)	(0.103)	(0.077)	(0.072)	(0.058)	(0.065)	(0.042)	(0.083)	(0.072)		
0.192	−0.330	−0.011	0.269	0.047	−0.292	−0.050	−0.015	0.213	0.907	1
(0.042)	(0.051)	(0.077)	(0.058)	(0.048)	(0.052)	(0.037)	(0.057)	(0.073)	(0.048)	

6 Conclusions

We have proposed a joint modelling approach for the mean and covariance structures in the context of normal longitudinal data. In the proposals presented here, the mean is modelled in a linear form and the covariance structure is left unrestricted in the sense that no specific parametric model is imposed on it, except for the fact that its modelling makes use of the general antedependence model specification, which has shown to be most useful in practice for nonstationary situations such as the ones present in the data sets analyzed here (see, e.g., Kenward, 1987, Pourahmadi, 1999 and 2000, Núñez-Antón and Zimmerman, 2000 or Zimmerman and Núñez-Antón, 2001). The proposals include both a classic and a Bayesian approach, allowing for the possibility of having noninformative priors for the latter. The behaviour of the proposed methodology is evaluated by analyzing two data sets and it has proved to be consistent and reasonable when compared to previous and less general proposals.

Extensions allowing for nonlinear mean structures are being considered at the moment but are beyond the scope of this paper.

Acknowledgements

Cepeda's work was supported by a grant from Universidad Nacional de Colombia. Núñez-Antón's work was supported by Ministerio Español de Educación y Ciencia, FEDER, Universidad del País Vasco (UPV/EHU) and Departamento of Educación del Gobierno Vasco (UPV/EHU Econometrics Research Group) under research grants MTM2004-00341, MTM2007-60112 and IT-334-07. The authors thank Dr. Dani Gamerman for helpful comments and suggestions which led to substantial improvement in the presentation of the material in this paper.

References

Cepeda, E.C. (2001). *Variability Modeling in Generalized Linear Models*. Unpublished Ph.D. Thesis. Mathematics Institute, Universidade Federal do Rio de Janeiro.

Cepeda, E.C. and Gamerman, D. (2000). Bayesian modeling of variance heterogeneity in normal regression models. *Brazilian Journal of Probability and Statistics (REBRAPE)*, 14, 207-221.

Cepeda, E.C. and Gamerman, D. (2004). Bayesian modeling of joint regressions for the mean and covariance matrix. *Biometrical Journal*, 4, 430-440.

Daniels, M.J. and Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, 89, 553-566.

Diggle, P.J. and Verbyla, A. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics*, 54, 401-415.

Diggle, P.J., Liang, K.-Y. and 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. *Annals of Mathematical Statistics*, 33, 201-212.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Kenward, M.C. (1987). A method for comparing profiles of repeated measurements. *Applied Statistics*, 36, 296-308.

Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics*, 38, 963-974.

Macchiavelli, R.E. and Arnold, S.F. (1994). Variable order antedependence models. *Communications in Statistics. Theory and Methods*, 23, 2683-2699.

Macchiavelli, R.E. and Moser, E.B. (1997). Analysis of repeated mesurements with ante-dependence covariance models. *Biometrical Journal*, 39, 339-350.

Núñez-Antón, V. and Zimmerman, D.L. (2000). Modelling nonstationary longitudinal data. *Biometrics*, 56, 699-705.

Pan, J.X. and MacKenzie, G. (2003). On modelling mean-covariance structures in longitudinal studies.

Biometrika, 90, 239-244.

Pan, J.X. and MacKenzie, G. (2006). Regression models for covariance structures in longitudinal studies. *Statistical Modelling*, 6, 43-57.

Pan, J.X. and MacKenzie, G. (2007). Modelling conditional covariance in the linear mixed model. *Statistical Modelling*, 7, 49-71.

Pan, J.X. and Ye, H. (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika*, 93, 927-941.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86, 667-690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, 87, 425-435.

Pourahmadi, M. (2002). Graphical diagnostics for modeling unstructured covariance matrices. *International Statistical Review*, 70, 395-417.

Pourahmadi, M. and Daniels, M.J. (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics*, 58, 225-231.

Stein, M.L. (1999). *Interpolation of Spatial Data. Some Theory for Kriging*. New York: Springer.

Wu, W.B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 90, 831-844.

Zimmerman, D.L. (2000). Viewing the correlation structure of longitudinal data through a PRISM. *The American Statistician*, 54, 310-318.

Zimmerman, D.L. and Núñez-Antón, V. (1997). Structured antedependence models for longitudinal data. In *Modelling Longitudinal and Spatially Correlated Data. Methods, Applications, and Future Directions*, T.G. Gregoire, D.R. Brillinger, P.J. Diggle, E. Russek-Cohen, W.G. Warren, and R. Wolfinger (eds.), 63-76. Lecture Notes in Statistics No. 122. New York: Springer-Verlag.

Zimmerman, D.L. and Núñez-Antón, V. (2001). Parametric modelling of growth curve data: An overview (with comments). *Test*, 10, 1-73.

Zimmerman, D.L., Núñez-Antón, V. and El Barmi, H. (1998). Computational aspects of likelihood-based estimation of first-order antedependence models. *Journal of Statistical Computation and Simulation*, 60, 67-84.