SORT 44 (1) January-June 2020, 171-200 DOI: 10.2436/20.8080.02.99

Bayesian structured antedependence model proposals for longitudinal data

Edwin Castillo-Carreno[1], Edilberto Cepeda-Cuervo[1]
and Vicente Núñez-Antón[2]

Abstract

An important problem in Statistics is the study of longitudinal data taking into account the effect of other explanatory variables, such as treatments and time and, simultaneously, the incorporation into the model of the time dependence between observations on the same individual. The latter is specially relevant in the case of nonstationary correlations, and nonconstant variances for the different time point at which measurements are taken. Antedependence models constitute a well known commonly used set of models that can accommodate this behaviour. These covariance models can include too many parameters and estimation can be a complicated optimization problem requiring the use of complex algorithms and programming. In this paper, a new Bayesian approach to analyse longitudinal data within the context of antedependence models is proposed. This innovative approach takes into account the possibility of having nonstationary correlations and variances, and proposes a robust and computationally efficient estimation method for this type of data. We consider the joint modelling of the mean and covariance structures for the general antedependence model, estimating their parameters in a longitudinal data context. Our Bayesian approach is based on a generalization of the Gibbs sampling and Metropolis-Hastings by blocks algorithm, properly adapted to the antedependence models longitudinal data settings. Finally, we illustrate the proposed methodology by analysing several examples where antedependence models have been shown to be useful: the small mice, the speech recognition and the race data sets.

MSC: 62F15, 62J05, 62P10.

Keywords: Antedependence models, Bayesian methods, Gibbs sampling, Mean-covariance modelling, Nonstationary correlation.

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 propos-

[1] Department of Statistics, Faculty of Sciences, Universidad Nacional de Colombia, Carrera 30, 45-03, Building 476, Bogotá D.C. (Colombia), ecastillo@unal.edu.co, ecepedac@unal.edu.co.

[2] (Corresponding author) Department of Applied Economics III (Econometrics and Statistics), Faculty of Economics and Business, University of the Basque Country UPV/EHU. Avda. Lehendakari Aguirre 83, 48015 Bilbao. (Spain). vicente.nunezanton@ehu.eus.

Received: July 2019
Accepted: May 2020

als in the literature to handle stationary or nonstationary correlations and variances, as well as balanced or unbalanced longitudinal data (Diggle et al., 2002; Weiss, 2005; Verbeke and Molenberghs, 2000; Fitzmaurice et al., 2009). A general fixed effects regression model for longitudinal data can be defined by assuming that the response variable $\mathbf{Y}_i$ can be explained with the model given by:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i,\ i = 1,\ldots,m, \tag{1}$$

where $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{in_i})^\top$ is the $n_i \times 1$ vector of responses for subject i, X_i is the $n_i \times q$ design matrix of rank q, which includes the covariates for the i-th subject; $\boldsymbol{\epsilon}_i$ is the vector of errors, assumed to follow a multivariate normal distribution with mean $\mathbf{0}$, and a given variance-covariance matrix so that $\text{Var}(\mathbf{Y}_i) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \sigma^2 V_{0i}$, whereas $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_k)^\top$ and $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_q)^\top$ are k and q-dimensional vectors of unknown parameters for the variance-covariance and mean model, respectively. Here, n_i represents the number of observations available for the i-th subject. If the number of observations available for each subject is the same (i.e., $n_i = n$, $\forall i$), we have a balanced data set. However, observations are, in general, not equally spaced. In addition, m represents the number of individuals in the study, and $N = \sum_{i=1}^{m} n_i$ represents the total number of observations. Fitting for the mean and covariance structure can be carried out by using maximum likelihood estimation methods with numerical maximization, such as the Newton Raphson or the EM algorithms (Ware, 1985). The model's assumptions include independence of responses from different subjects, multivariate normality of responses, and either no missing data or, at worst, ignorably missing responses (Laird, 1988).

The approach of fitting a regression model for longitudinal data by means of specifying the variance-covariance structure includes the possibility of having several different structures, which can be stationary or nonstationary in terms of correlation between observations along time, and homogeneous or heterogeneous in terms of variance as a function of the time at which observations are taken. Among the most commonly used covariance structures featuring homogeneous variances and stationary correlations are the compound symmetry (CS), autoregressive structures of order p (AR(p)), autoregressive with moving average structures of order p and q (ARMA(p,q)) models (Weiss, 2005). Models for nonstationary correlations and heterogeneous variances include the heterogeneous versions of the previous models, which are not always the best-fitting models for this type of settings. Therefore, mode general models, such as the integrated autoregressive with moving average model (ARIMA) or generalizations of the autoregressive models, such as the unstructured and structured versions of the antedependence models of order s (AD(s) or SAD(s)) need to be implemented (Núñez-Antón and Zimmerman, 2001; Zimmerman and Núñez-Antón, 2010). Estimation in any of these variance-covariance parametric models is commonly carried out by restricted maximum likelihood methods, together with the use of recursive algorithms. Some computational software packages such as, for example, SAS© or SPSS©, include the possibility of specifying some particular variance-covariance parametric choices to estimate

this type of models. Estimation in higher order AD or SAD models usually requires the use of specific numerical algorithms that need to be directly programmed in specific computing languages or available software statistical packages.

Bayesian estimation proposals for longitudinal data settings where no specific variance-covariance structure is considered in the model specification were previously developed (Brown, Kenward and Bassett, 2001), where a Wishart prior distribution was assumed for the covariance structure. In addition, the joint estimation for the mean structure and some simple covariance structures under the assumption of prior normal distribution for the mean parameters, and inverse gamma distributions for the variance in the proposed model were previously proposed (Hui and Berger, 1983). Moreover, the advantages of this proposal for fitting a growth curve model in post-menopause female bone calcium loss were also illustrated. The first proposal introduced Bayesian longitudinal models by taking into account regression structures in both the mean and the variance-covariance matrix of normal observations (Cepeda-Cuervo, 2001). This approach was based on the modelling proposal that used the Cholesky's matrix decomposition (Macchiavelli and Arnold, 1994; Pourahmadi, 1999). More specifically, by assuming normal prior distributions for the mean and variance regression structures parameters, a Bayesian methodology was introduced to fit the proposed models building the kernel transition functions from observational working variables (Cepeda-Cuervo, 2001). Results and some of the extensions of this work have also been presented by several authors (Cepeda and Gamerman, 2004; Cepeda-Cuervo and Núñez-Antón, 2007; Cepeda-Cuervo and Núñez-Antón, 2009; Cepeda-Cuervo, 2011), where, in addition, observational units are allowed to be correlated. These proposals included a detailed description of the optimization algorithms, as well as simulation and case studies that allowed for the comparison of the Bayesian and classic proposals for the analysis of this type of data. A Bayesian version of first-order multivariate antedependence model has also been developed (Jiang et al., 2015). Finally, it is interesting to briefly mention Bayesian AD models within the framework of Bayesian hierarchical mixed linear models, where, in general, authors have assumed that the errors are independent and identically distributed (i.i.d.), and also conjugate prior distributions for the parameters in the proposed models (Congdon, 2020; Gelman et al., 2014a; Gill, 2014). More specifically, by following the proposals in Congdon, 2020, some possible extensions of Bayesian hierarchical mixed linear models can be considered, so that allowing for autocorrelated errors is possible. That is, there exists the possibility of assuming that the covariance matrix of the random errors, or that of the random effects, follows an AD model (Fahrmeir, Kneib and Lang, 2013). Our proposals would allow researchers to develop these models and to extend non-Bayesian previous methods for hierarchical mixed linear models with AD structures, such as, for example, the ones in Jaffrézic and Pletcher (2000), Jaffrézic et al. (2002) and Yang and Tempelman (2012), to a Bayesian context.

In this paper we propose a Bayesian method for the joint estimation of the mean and covariance parameters in the regression longitudinal models settings under the normality assumption, and also allowing for the specification of several different variance-

covariance structures. Our proposals start by considering variance-covariance models with stationary correlations and homogeneous variances, as is the case in the CS, AR(1) and ARMA(1,1) models, so that they are then generalized to consider nonstationary correlations and heterogeneous variances, such as is the case in the structured antedependence model of order one, or SAD(1) model. That is, we extend the previous proposal (Cepeda-Cuervo, 2001) to consider parametric more parsimonious variance-covariance models that have been shown to be more useful in longitudinal data settings than those of the unstructured AD model previously considered therein. For each one of the variance-covariance structures considered here, we provide a detailed description of the estimation algorithm constructed for each specific case, including the Gibbs sampling and the Metropolis-Hasting by blocks algorithm used under each of the assumed covariance structures. In order to be able to assess the behaviour of the estimation proposed algorithms, for the specific cases of CS, AR(1) and ARMA(1,1) variance-covariance structures, a real data set analysis for the *Small Mice* balanced data set (Izenman and Williams, 1989; Weiss, 2005) is carried out. As for the SAD(1) variance-covariance structures and given that, as previously mentioned (Zimmerman and Núñez-Antón, 2010), the proposed variance-covariance model depends on the specific data sets and on their underlying structure, we compare two specific structured models based on the analysis of the *Speech Recognition* data set (Tyler et al., 1988; Núñez-Antón and Woodworth, 1994; Zimmerman, Núñez-Antón and El Barmi, 1998), and also on the analysis of the *100-Km Race* data set, kindly provided by Ian Jollife of the University of Kent (Zimmerman et al., 1998).

The paper is organized as follows. In Section 2, we introduce and describe the basic characteristics of the variance-covariance models we consider. In Section 3 we include the Bayesian longitudinal model proposals, as well as the posterior distributions and kernel transition functions for each of the models considered, which include the proposed algorithms and prior distribution assumptions required for each of them. In Section 4 we introduce and describe the data sets to be analysed, as well as the main objectives of the data set analyses. In Section 5, we analyse the different data sets under the Bayesian proposals included, describe the results and compare them with those obtained with previous classic approaches. We also present a sensitivity analysis for the estimates obtained under our proposals. Finally, in Section 6, we provide some general conclusions and final practical recommendations.

2 Some covariance structures

As already mentioned by several authors (Weiss, 2005; Núñez-Antón and Zimmerman, 2001), some of the clear advantages of parametric modelling approaches for the variance-covariance matrix in longitudinal data settings are the following: (a) they help to optimize the obtention of estimates for the parameters in the mean structure; (b) they allow to obtain the most appropriate estimates for the standard errors for the estimators

of the parameters included in the mean structure (i.e., $\boldsymbol{\beta}$); (c) in most cases, they provide a feasible and effective solution when estimating models in data sets with missing data or when times at which measurements are taken are not the same for all of the individuals in the study; and (d) estimates are still valid even for the cases where the number of observations on each individual is relatively large when compared to the number of individuals in the study.

Specific variance-covariance structures to be introduced in this paper consider that all of the variances and covariances within a given individual are functions of a vector of parameters with a small or moderate number of elements, which, as in equation (1), will be denoted by $\boldsymbol{\theta}$. That is, the covariance model $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ defines a family of possible variance-covariance matrices depending on the k-order vector of parameters $\boldsymbol{\theta}$. Parameter estimation for the covariance structure is usually carried out by maximum likelihood or restricted maximum likelihood methods (Diggle et al., 2002). One of the main challenges and problems when modelling covariance structures within longitudinal data settings is to be able to select the so-called "best-fitting" or "most appropriate" covariance structure for the specific data set under study. Most researchers agree that, in order to do so, a combination of graphical methods, exploratory descriptive analysis, as well as profile plots tools provides the necessary and required information to be able to narrow down the possible covariance structure choices to the ones that can be considered as optimal choices for the data set under study (Verbeke and Molenberghs, 2000; Fitzmaurice et al., 2009). Our suggestion for proposing or considering "reasonable" covariance structures for a specific data set, which we have followed in Section 5, can be summarized in the following items (Zimmerman and Núñez-Antón, 2010):

- Compute the means for the different time points and build a profile plot, using the `matplot` function in R, for the observations in your data set. The behaviour of the means along time will provide the user clear ideas about the type of mean function that needs to be used for the mean model. In addition, the profile plot also provides information about the possible behaviour of the variance for the different time points in the data set. Compute the variances for the different time points, as well as the correlation matrix for the corresponding data set, so that stationarity in variance and correlation can be better assessed.
- Build the corresponding ordinary scatterplot graph - OSM, using the `splom` function in the **lattice** package in R, semivariogram, or PRISM (Zimmerman, 2000), to better assess the correlation structure in your data set. These graphs are built for a saturated mean model, which considers a mean parameter for each time point. Inspection of these graphs will provide the user with a clear idea about the different covariance models that could be considered for the data set under study. The user is now able to propose a set of suitable covariance structures for the data set under study, and the best fitting covariance model will be selected on the basis of some specific goodness-of-fit criteria.

- Finally, the user can test or assess for possible mean model reductions by fitting alternative mean models and selecting the best fitting one on the basis of specific tests or goodness-of-fit criteria. Later in this section, specially for antedependence models, we describe different formulations for the covariance models considered here. For easiness of comprehension and understanding of these formulations, we recommend the use of variance-covariance formulations such as the ones in equations (4), (7), (8) and (9), which are the ones we use in the applications in Section 5, as well as in the Bayesian proposals in Section 4.

We now introduce the different variance-covariance structures, $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$, that we include in our methodological Bayesian proposals. It is worth mentioning that, besides the covariance structures introduced here, there are additional structures that interested readers may wish to read about (Weiss, 2005; Núñez-Antón and Zimmerman, 2001).

The simplest variance-covariance structure, besides the obvious independence structure which is not of real interest within these settings, corresponds to the so-called compound symmetry (CS), equicovariance or equicorrelation model, which is defined by assuming that homogeneous or constant variances in time and equal correlations between different measurements on the same subject. That is, $\text{Var}(Y_{ij}) = \sigma^2,\ j = 1,\ldots,n_i$, and $\text{Corr}(Y_{ij},Y_{il}) = \rho,\ j \neq l$. There is a heterogenous version of the CS model, CSH, where variances are allowed to change over time (Núñez-Antón and Zimmerman, 2001).

The first order autoregressive structure, AR(1), includes two covariance parameters, σ^2 and ρ, with $\text{Var}(Y_{ij}) = \sigma^2,\ j = 1,\ldots,n_i$, and ρ is the correlation parameter such that $\text{Corr}(Y_{ij},Y_{il}) = \rho^{|t_{ij}-t_{il}|},\ j \neq l$. This type of serial correlation differs from the CS model correlation because in the autoregressive model of order one, the correlation decreases as a power function of time. As can be easily seen, the AR(1) model assumes homogeneous variances and stationary correlations. That is, variances are constant over time and correlations between observations taken at equally spaced time points are also constant. There is, however, a heterogeneous version of the AR(1) model, ARH(1), where variances are allowed to change with time (Núñez-Antón and Zimmerman, 2001). Differences between the AR(1) and CS models are very difficult to assess from the exploratory analysis or the individuals' profile plots, specially when there are only few observations available per subject (Fitzmaurice et al., 2009).

The autoregressive with moving average model or order (1,1), ARMA (1,1), represents a generalization of the previous two models, CS and AR(1). In this model, the correlation between consecutive observations of the same observational unit is given by:

$$\text{Corr}(Y_{ij},Y_{il}) = \begin{cases} \phi & |j-l| = 1 \\ \phi\rho^{|t_{ij}-t_{il}|-1} & |j-l| > 1, \end{cases} \tag{2}$$

with $\text{Var}(Y_{ij}) = \sigma^2,\ j = 1\ldots,n_i$, and where ϕ, $0 < \phi < 1$, is the correlation between consecutive observations of the same observational unit, ρ, $0 < \rho < 1$, is an additional parameter, which allows the correlation to feature an exponential decreasing behaviour.

As can be seen the ARMA(1,1) model reduces to the previous models: to the CS model if $\rho = 1$, and to the AR(1) model if $\phi = \rho$; and a moving average model of order 1, MA(1), if $\rho = 0$ (Weiss, 2005). In addition, the ARMA(1,1) model also assumes homogeneous variances and stationary correlations.

The concept of antedependence was originally introduced in 1962 (Gabriel, 1962), and the antedependence models within the longitudinal data settings first defined in 2010 (Zimmerman and Núñez-Antón, 2010). Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$ be the vector of measurements taken on the i-th subject, which is assumed to follow a multivariate normal distribution. The antedependence longitudinal model of order s, AD(s) with an autoregressive specification (Zimmerman and Núñez-Antón, 2010), is defined as:

$$\begin{aligned} Y_{i1} &= \mathbf{x}_{i1}^{\mathsf{T}}\boldsymbol{\beta} + \epsilon_{i1} \\ Y_{ij} &= \mathbf{x}_{ij}^{\mathsf{T}}\boldsymbol{\beta} + \sum_{k=1}^{s^*} \phi_{j,j-k}(Y_{i,j-k} - \mathbf{x}_{i,j-k}^{\mathsf{T}}\boldsymbol{\beta}) + \epsilon_{ij} \qquad j = 2, \ldots, n_i, \end{aligned} \tag{3}$$

where $\mathbf{x}_{ij}$ be a q-vector of covariates associated to Y_{ij}, $s^* = \min(s, j-1)$, the ϵ_{ij}'s are independent $N(0, \sigma_j^2)$ random variables, and σ_j^2 and $\phi_{j,j-k}$ are unstructured parameters. In this model, each variable is regressed on the previous s^* predecessors in the ordered list and, in addition, it is also allowed that autoregressive coefficients vary with time (i.e., that they depend upon j). In this sense, AD models are nostationary in both variance and correlation, because variances may vary with time and correlations between equidistant observations in time are not necessarily assumed to be constant. Specific elements of the variance-covariance matrix $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ in this model can be recursively obtained by using the well known Yule-Walker equations approach, so that, if an AD(1) model with a covariance specification is assumed (Zimmerman and Núñez-Antón, 2010), $\mathrm{Var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ can be specified as:

$$\begin{aligned} [\boldsymbol{\Sigma}_i(\boldsymbol{\theta})]_{kk} &= \sigma_k^2, \;\; k = 1, \ldots, n_i \\ [\boldsymbol{\Sigma}_i(\boldsymbol{\theta})]_{kl} &= \sigma_k \sigma_l \prod_{j=k}^{j=l-1} \rho_j, \;\; k < l, \;\; k, l = 1, \ldots, n_i \\ [\boldsymbol{\Sigma}_i(\boldsymbol{\theta})]_{kl} &= \boldsymbol{\Sigma}_i(\boldsymbol{\theta})_{lk}, \;\; k > l, \;\; k, l = 1, \ldots, n_i, \end{aligned} \tag{4}$$

with $\rho_j = \rho_{j,j+1}$. Antedependence models of order s can be not so parsimonious mainly because the vector of variance-covariance parameters $\boldsymbol{\theta}$ has, for $n_i = n, \forall i$, $(s+1)(2n - s)/2$ parameters (Zimmerman and Núñez-Antón, 1997). In addition, as the autoregressive coefficients and the variances of the ϵ_{ij}'s in (3) depend on the time at which measurements are taken, variances in this model are heterogeneous and correlations are nonstationary. That is, variances are allowed to change over time and correlations between observations taken at equally spaced time points are not constant and, thus, are allowed to vary. The same holds for the AD(1) model in (4).

Zimmerman and Núñez-Antón (1997) originally proposed the structured antedependence (SAD) models in 1997. Their proposed models specify that the correlation parameters are determined by a Box-Cox power function and the variances for each time point are determined by a polynomial function of not so many parameters, were able to model nonstationary correlations and variances. Moreover, Núñez-Antón and Woodworth (1994) and Zimmerman and Núñez-Antón (1997) have previously defined the specific commonly used functions for the parameters in model (3), for a general structured antedependence model or order s, SAD(s), with $n_i = n$, as:

$$\phi_{j,j-k} = \phi_k^{f(t_{ij},\lambda_k)-f(t_{i,j-k},\lambda_k)},\ j = s+1,\ldots,n;\ k = 1,\ldots,s \tag{5}$$

$$\sigma_j^2 = \sigma^2 G(t_{ij},\boldsymbol{\psi}),\ \ j = 1,\ldots,n, \tag{6}$$

or equivalently, for (4), with:

$$\rho_{j,j-k} = \rho_k^{f(t_{ij},\lambda_k)-f(t_{i,j-k},\lambda_k)},\ j = s+1,\ldots,n;\ k = 1,\ldots,s \tag{7}$$

$$\sigma_j^2 = \sigma^2 G(t_{ij},\boldsymbol{\psi}),\ \ j = 1,\ldots,n, \tag{8}$$

where

$$f(t_{ij},\lambda_k) = \begin{cases} (t_{ij}^{\lambda_k}-1)/\lambda_k, & \text{if } \lambda_k \neq 0 \\ \log(t_{ij}), & \text{if } \lambda_k = 0, \end{cases} \tag{9}$$

with $\phi_k > 0, 0 < \rho_k < 1, \forall k$, $\sigma_j^2 > 0, \forall j$, and $\{\boldsymbol{\psi} : G(t_{ij},\boldsymbol{\psi}) > 0\}$, in such a way that the variance-covariance matrix for the i-th subject, $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$, is positive definite. Here, as will be seen in the applications in Section 5, $G(t_{ij},\boldsymbol{\psi})$ is usually assumed to be a positive power or step function of time. In addition, given that the SAD models are special cases of the AD models, variances in these models are heterogeneous and correlations are also nonstationary. Equation (9) represents a Box-Cox power law. Moreover, in the SAD(1) model settings, and if measurement times are equally spaced, then the lag-one correlations (and, as a matter of fact, all same-lag correlations) are a monotone function of t: they increase if $\lambda < 1$ and decrease if $\lambda > 1$. For $\lambda = 1$, same-lag correlations remain constant and, in addition, they coincide with those of the AR(1) model. That is, the Box-Cox power law can be seen as a transformation to the time scale that effects a nonlinear deformation upon the time axis, such that correlations between measurements equidistant in the deformed scale remain constant. There are some specific special cases of the SAD(1) model that are worth mentioning:

1. **Type 1 - SAD model**: We assume an SAD(1) model as in (4), such that (8) holds with $G(t_{ij},\boldsymbol{\psi}) \equiv 1$, with $\rho_j = \rho_{j,j+1} = \rho^{f(t_{i,j+1},\lambda)-f(t_{i,j},\lambda)}$, and $f(t,\lambda)$ given by equation (9). This model assumes homogeneous variances and nonstationary correlations.
2. **Type 2 - SAD model**: We assume an SAD(1) model as in (4), such that (8) holds, with

$$G(t,\psi) = \begin{cases} 1, & \text{if } t = 1 \\ \psi, & \text{otherwise,} \end{cases} \tag{10}$$

and ρ_j as in the previous model. This model assumes a specific case of heterogeneous variances and nonstationary correlations.

3. **Type 3 - SAD model**: We propose an SAD(1) model as in (4), with ρ_j defined as in the previous models and

$$h(\sigma_j^2) = \psi_0 + \psi_1 t_{ij} + \cdots + \psi_r t_{ij}^r, \tag{11}$$

where h is an appropriately chosen link function so that the σ_j^2 variances are positive. Some authors (Zimmerman et al., 1998) have previously proposed h to be the identity link function, whereas we propose, without loss of generality, to use the logarithmic link function instead. Our proposal is more general in the sense that it does not require any additional constraints on the parameters for the variances to be positive. Moreover, this model assumes heterogeneous variances and nonstationary correlations.

3 Bayesian longitudinal model methodological proposals

Let $\mathbf{t}_i = (t_{i1}, t_{i2}, \ldots, t_{in_i})^\top$, represent the times at which observations on the i-th subject were taken, and Y_{ij} represent the observation taken on subject i at time t_{ij}, $j = 1, \ldots, n_i$. Let $\mathbf{x}_{ij}$ be a q-vector of covariates associated to Y_{ij}, so that $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in_i})^\top$ is the $n_i \times q$ design matrix of rank q. In this way, we have that model (1) holds. Thus, if $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m)^\top$ denotes the vector of measurements for all of the m individuals in the study, having a design matrix $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top, \ldots, \mathbf{X}_m^\top)^\top$, containing the values for the covariates for all individuals, we have that:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{12}$$

where $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_m)^\top$ is a vector of random errors associated to the corresponding component in the responses vector $\boldsymbol{Y}$, so that the $\boldsymbol{\epsilon}_i$'s are assumed to be independent from each other, $\boldsymbol{\epsilon} \sim MVN$ with mean $\mathbf{0}$ and block diagonal variance-covariance matrices, so that $\text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ will be a block diagonal matrix with diagonal components $\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{\Sigma}_m(\boldsymbol{\theta})$.

3.1 Prior parameter distributions

In order to provide the required details for our proposed Bayesian longitudinal method, prior distribution should be assumed for the mean and for the variance-covariance regression structure parameters (Gelman, 2006). For the mean regression parameters, we assume a q-multivariate normal distribution, so that $p(\boldsymbol{\beta}) \sim N(\mathbf{b}_0, \mathbf{B}_0)$. As for the

variance-covariance parameters, we assume a prior distribution $p(\boldsymbol{\theta})$ that will depend on the assumed covariance structure. More specifically:

1. For the CS and AR(1) models, if we let $\varphi = 1/\sigma^2$, the variance-covariance vector parameter in these models is $\boldsymbol{\theta} = (\varphi, \rho)^{\intercal}$, so that its assumed prior distribution is $p(\boldsymbol{\theta}) = p(\varphi)p(\rho)$, where:

$$p(\varphi) \equiv \text{Gamma}\left(\frac{g_0}{2}, \frac{g_0\sigma_0^2}{2}\right) \tag{13}$$

$$p(\rho) \equiv \text{Beta}(a,b), \tag{14}$$

   where g_0, σ_0^2, a and b are assumed to be known hyperparameter values (Gelman, 2006).

2. In the ARMA(1,1) structure, given that $0 < \phi < 1$ and $0 < \rho < 1$, the parameter vector is $\boldsymbol{\theta} = (\varphi, \rho, \phi)^{\intercal}$, so that its assumed prior distribution is $p(\boldsymbol{\theta}) = p(\varphi)p(\rho)p(\phi)$, where $p(\phi) \equiv \text{Beta}(a_1, b_1)$, $p(\rho) \equiv \text{Beta}(a_2, b_2)$, and, for $\varphi = 1/\sigma^2$, we have the same prior distributional assumption as in (13).

3. For the structured antedependence models, we assume the following independent prior distributions:

   (a) **Type 1 - SAD model**: In this model, assumed prior distributions for σ^2 and ρ, are as above. That is, for $\varphi = 1/\sigma^2$, we have $p(\varphi) \equiv \text{Gamma}\left(\frac{g_0}{2}, \frac{g_0\sigma_0^2}{2}\right)$ and, for ρ, we have $p(\rho) \equiv \text{Beta}(a,b)$. For λ, we assume a uniform prior distribution, so that $p(\lambda) \equiv U(-a,a)$.

   (b) **Type 2 - SAD model**: For this model, the same prior distributions as above are assumed for $\varphi = 1/\sigma^2$, ρ and λ. For ψ, if we let $\psi = \exp(\eta)$, we then assume that the prior distribution for η is such that $p(\eta) \equiv N(0, \nu^2)$.

   (c) **Type 3 - SAD model**: In this model, the same prior distributions as above are assumed for ρ and λ. For $\boldsymbol{\psi} = (\psi_0, \psi_1, \ldots, \psi_r)^{\intercal}$, we assume a multivariate prior normal distribution, so that $p(\boldsymbol{\psi}) \equiv MVN(\boldsymbol{\psi}_0, \mathbf{K}_0)$.

A final comment related to the aforementioned assumed prior distributions for the different covariance models: we believe that it is relevant to mention that, given that we do not really have prior information related to the parameters in the models, which will be the ones that will to be estimated in the applications in Section 5, we have decided to assume vague prior distributions so we do not include any prior unknown information that can generate unjustified and unnecessary changes in the posterior distributions that will be used for inferential purposes in Section 5. However, if we have prior information available for the mean regression parameters, it can be easily incorporated in the model, by assuming appropriate values for $\mathbf{b}_0$ and $\mathbf{B}_0$ in $p(\boldsymbol{\beta}) \sim N(\mathbf{b}_0, \mathbf{B}_0)$. With regard to the

variance parameters, our recommendation is that the prior information is specified as follows: (i) the gamma prior distribution for $\varphi = 1/\sigma^2$ can be specified by the mean g_0 and the variance σ_0^2 from the prior information for the parameter σ^2 available in the specific application; (ii) the parameters a and b in the beta prior distribution for ρ can be specified from their prior mean and variance for ρ available in the specific application; (iii) the parameters a and b in a more general uniform prior distribution for λ (i.e., $U(a,b)$), can be also specified from the prior information for λ available in the specific application; (iv) the prior distributions for η and ψ, or $\boldsymbol{\psi}$, in the Type 2 and Type 3 SAD models, respectively, can be directly specified from their corresponding prior means and variances information for these parameters available in the specific application.

3.2 Posterior conditional distributions and estimation proposals

Under the model assumptions, apart from a constant term, the likelihood function is given by:

$$\mathbb{L}(\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{Y}) \propto \prod_{i=1}^{m} |\boldsymbol{\Sigma}_i(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\intercal}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right]\right\}, \tag{15}$$

where $\boldsymbol{\theta} = (\sigma^2,\rho)^{\intercal}$ in the CS and AR(1) models, $\boldsymbol{\theta} = (\sigma^2,\rho,\phi)^{\intercal}$ in the ARMA(1,1) model, and $\boldsymbol{\theta} = (\sigma^2,\boldsymbol{\rho},\lambda,\boldsymbol{\psi})$ in the SAD models. Thus, the posterior parameter distribution is given by $p(\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{Y}) \propto \mathbb{L}(\boldsymbol{\beta},\boldsymbol{\theta}|\mathbf{Y})p(\boldsymbol{\beta})p(\boldsymbol{\theta})$. Moreover, given that, under the assumed prior distribution for $\boldsymbol{\beta}$ we have that:

$$p(\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mathbf{b}_0)^{\intercal}\mathbf{B}_0^{-1}(\boldsymbol{\beta}-\mathbf{b}_0)\right\}, \tag{16}$$

the posterior conditional distribution of $\boldsymbol{\beta}$ will be $p(\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{Y}) \equiv N(\mathbf{b}^*,\mathbf{B}^*)$, where (Cepeda-Cuervo, 2001; Cepeda and Gamerman, 2004):

$$\mathbf{b}^* = \mathbf{B}^*(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}^{\intercal}\boldsymbol{\Sigma}^{-1}\mathbf{Y}) \tag{17}$$

$$\mathbf{B}^* = (\mathbf{B}_0^{-1} + \mathbf{X}^{\intercal}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \tag{18}$$

Samples of $\boldsymbol{\beta}$ are are taken from the conditional posterior distribution $p(\boldsymbol{\beta}|\boldsymbol{\theta},\mathbf{Y}) \equiv N(\mathbf{b}^*,\mathbf{B}^*)$, and accepted with probability one (Gamerman and Lopes, 2006; Gelman et al., 2014b).

3.3 Posterior conditional distributions for σ^2 and ρ in the CS and AR(1) models

Taking into account that, in the in CS and AR(1) models, the variance-covariance matrix can be written as $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \frac{1}{\varphi}\boldsymbol{C}(\rho)$, with $\varphi = 1/\sigma^2$, samples of φ and ρ are obtained from their conditional posterior distributions. More specifically, samples of φ are obtained

from its posterior distribution, given by:

$$p(\varphi) \propto \varphi^{\left(\frac{m+g_0}{2}-1\right)} \exp\left(-\varphi\frac{g_0\sigma_0^2+R}{2}\right), \tag{19}$$

where $R = (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\intercal}\mathbf{C}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$. That is, values for φ can be obtained from the conditional gamma posterior distribution Gamma $\left(\frac{m+g_0}{2}, \frac{g_0\sigma_0^2+R}{2}\right)$. For the parameter ρ, and given that its posterior distribution $p(\rho|\boldsymbol{\beta},\varphi)$ is analytically intractable for these covariance models, we propose that samples generating the posterior distribution for ρ be obtained, using the MCMC algorithm (Gamerman and Lopes, 2006; Gelman et al., 2014b), from the kernel transition function:

$$q(\rho^{(*)}|\rho^{(k)}) = \begin{cases} \rho^{(*)} \sim U(0, 2\rho^{(k)}) & \rho^{(k)} \leq 0.5 \\ \rho^{(*)} \sim U(2\rho^{(k)}-1, 1) & \rho^{(k)} > 0.5 \end{cases} \tag{20}$$

3.4 Posterior conditional distributions for σ^2, ρ and ϕ in the ARMA(1,1) model

We assume a longitudinal model as in (12) with variance-covariance structure similar to (2). As in Section 3.3, samples of $\varphi = 1/\sigma^2$ are obtained from the posterior distribution described before; that is, from the corresponding gamma posterior distribution Gamma $\left(\frac{m+g_0}{2}, \frac{g_0\sigma_0^2+R}{2}\right)$. However, for the parameters ρ and ϕ, and given that their posterior distributions are analytically intractable for the ARMA(1,1) model, we propose that samples generating the posterior distribution for ρ be obtained as before, from (20), and samples for the posterior distribution of ϕ be obtained, using the MCMC algorithm (Gamerman and Lopes, 2006; Gelman et al., 2014b), from the kernel transition function:

$$q(\phi^{(*)}|\phi^{(k)}) = \begin{cases} \phi^{(*)} \sim U(0, 2\phi^{(k)}) & \phi^{(k)} \leq 0.5 \\ \phi^{(*)} \sim U(2\phi^{(k)}-1, 1) & \phi^{(k)} > 0.5 \end{cases} \tag{21}$$

3.5 Posterior conditional distributions for σ^2, ρ, λ and ψ in the SAD(1) models

As the number of parameters in the variance-covariance matrix for the structured antedependence models of order one, SAD(1), depend on the specific selected $G(\boldsymbol{\psi},t)$ function in (8) for the type 1 and 2 models, or (11) for the type 3 model, we have to propose specific Bayesian estimation modelling approaches for each of them, which will depend on the type of SAD model being considered. Based on the types of SAD models described in Section 3.1 above, we describe the different distributions and estimation algorithms for each of them in what follows.

1. **Type 1 - SAD model**: To estimate σ^2, we use the proposal in Section 3.3, so that samples of $\varphi = 1/\sigma^2$ are obtained from the posterior distribution described above. As for the ρ parameter, we propose that samples generating their posterior distribution be obtained, using MCMC and a transition kernel such as the one in (20). Given that we have assumed a $U(-a,a)$ uniform prior distribution for the parameter λ, samples from its posterior conditional distribution are obtained by using an MCMC algorithm (Gamerman and Lopes, 2006; Gelman et al., 2014b), assuming that $\lambda^{(*)} = a(2\nu^{(*)} - 1)$, where $\nu^{(*)}$ is obtained from a kernel transition function similar to those previously defined in equations (20) and (21).

2. **Type 2 - SAD model**: Our Bayesian proposal to estimate σ^2, ρ and λ is similar to the one described for the type 1 SAD model. As for the parameter ψ, we let $\psi = \exp(\eta)$ and, in addition, assume that the prior distribution for η is such that $p(\eta) \sim N(0,\nu^2)$. In this way, the complete conditional posterior distribution is not known, so that samples for the posterior distribution of ψ can be obtained, using the MCMC algorithm (Gamerman and Lopes, 2006; Gelman et al., 2014b), from the kernel transition function:

$$\begin{aligned} q(\psi^{(*)}, \psi^{(k)}) &= \psi^{(k)} + N(0,\nu^2) \\ \psi^{(k)} &= \exp(\eta^{(k)}) \end{aligned} \tag{22}$$

3. **Type 3 - SAD model**: Given that the posterior conditional distribution for $\boldsymbol{\psi}$, $p(\boldsymbol{\psi}|\boldsymbol{\beta},\lambda,\rho,\mathbf{Y})$, is analytically intractable, we propose a kernel transition function given by the observational model obtained from $\tilde{Y}_j = \frac{1}{m-1}\sum_{i=1}^{m}(Y_{ij} - \bar{Y}_j)^2$, where $\bar{Y}_j = \frac{1}{m}\sum_{i=1}^{m} Y_{ij}$, and by assuming, without loss of generality, that $n_i = n$, and that the working observational model

$$\widetilde{w}_j = \log(\tilde{Y}_j) = \psi_0 + \psi_1 X_{1j} + \psi_2 X_{2j} + \varepsilon_j \tag{23}$$

follows a normal distribution, where $\varepsilon_j \in N(0,\sigma^2)$, with σ^2 known, and such that $\tilde{\mathbf{X}}_j = (1, X_{1j}, X_{2j})$ and $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1^\intercal, \ldots, \tilde{\mathbf{X}}_n^\intercal)^\intercal$. Thus, the kernel transition function for $q(\boldsymbol{\psi})$ is obtained from the combination of the normal prior distribution and the observational model in (23). That is,

$$q(\boldsymbol{\psi}|\mathbf{Y}) \equiv N(\boldsymbol{\mu}_\psi, \mathbf{K}_\psi), \tag{24}$$

where $\boldsymbol{\mu}_\psi = \mathbf{K}_\psi(\mathbf{K}_0^{-1}\boldsymbol{\psi}_0 + \tilde{\mathbf{X}}^\intercal\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{W}})$, with $\tilde{\boldsymbol{\Sigma}} = \text{diag}(\sigma^2)$, $\tilde{\mathbf{W}} = (\widetilde{w}_1, \ldots, \widetilde{w}_n)^\intercal$ and $\mathbf{K}_\psi = (\mathbf{K}_0^{-1} + \tilde{\mathbf{X}}^\intercal\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{X}})^{-1}$ (Gelman et al., 2014b).

As a final comment to the posterior inferences related to all of the models described in Section 3, and given that the proposed Bayesian inference is based on the posterior distribution of the parameters and, in addition, given that, for the models considered

here, the posterior distributions do not have a closed form expression, inferences are based on the simulation of the posterior distributions obtained by applying the MCMC algorithm (Gamerman and Lopes, 2006; Gelman et al., 2014b). A first approach to simulate samples of the posterior distribution may be to apply the Gibbs sampler algorithm, but this is possible only if all conditional posterior distributions are known. If some or all of the conditional posterior distributions are not known, as is the case in the longitudinal Bayesian models proposed here, kernel transition functions should be built in order to be able obtain samples of the unknown conditional distributions using the Metropolis Hastings algorithm. Therefore, a Metropolis-Hastings-within the Gibbs algorithm is used to be able to draw samples of the posterior distributions, from which the posterior inferences can be straightforwardly obtained. For example, in CS and AR(1) models, samples of the mean regression parameters $\boldsymbol{\beta}$ are proposed from a normal distribution with mean and variance given by equations (17) and (18), respectively, and samples of $\varphi = 1/\sigma^2$ are obtained from the gamma distribution given in equation (19), where in both cases the Gibbs sampler algorithm is used. Moreover, samples of ρ are proposed from the kernel transition function in equation (20), by applying the Metropolis Hastings algorithm. Therefore, samples of the target posterior distributions are obtained by applying an iterative algorithm, so that posterior inferences on the parameters in the models can be obtained. A proper construction of the kernel transition function is very important to improve the convergence of the chains and to be able to obtain better posterior inferences.

4 Data

4.1 Small Mice Data

The Small Mice data set (Izenman and Williams, 1989) was used to illustrate the proposals along the lines of spectral models for the analysis of longitudinal data. The study analysed more than 600 mice at birth, 7 days after birth (onset of growth), 14 days (when eyes open and consumption of solid food begins), 21 days (end of maternal influence for food) and 42 days (when most mice reach sexual maturity). Of these 600 observations a particular group of 35 male mice was divided into 4 groups. The Small Mice data constitutes a balanced set of longitudinal data with the weights in milligrams of 14 mice which make up groups 3 and 4 of the original study (Izenman and Williams, 1989). These weights were taken in the days corresponding to $t = 2,5,8,11,14,17,20$ after birth by the same person using the same measurement scale. The objective is to find a parsimonious model describing in the best possible way how weight is related to the time at which measurements were taken, and weight on previous times.

4.2 Speech Recognition Data

This set of data comes from the audiological study presented in 1988 (Tyler et al., 1988). The general study was performed with five different types of implants, three single channel implants and two multichannel implants. The implants were surgically implanted five to six weeks before connecting electrically to an external voice processor. The data includes the scores obtained when performing a speech recognition test on patients with multichannel cochlear implants. These patients were divided into two groups depending on the type of implant received (namely A and B): 21 subjects received implant A and 21 subjects received implant B. The individuals in the study were bilaterally deaf, therefore the base values of the test were all equal to zero. Measurements were taken at 4 time points: 1,9, 18 and 30 months after having received the implant. In the study there was a variation in the actual follow-up times, so these times were not exact. In addition, some subjects did not show up in one or more of their programmed follow-ups, so some data were missing (there were eight missing observations at month 18 and twenty missing observations at month 30). It was assumed observations were missing at random (Zimmerman and Núñez-Antón, 2010). The interest of studying these data focuses on describing the audiological performance of the individuals who receive each type of implant and how their performance depends on the time elapsed since implantation, as well as on the type of implant. More specifically, the goal is to assess how the average means of the types of implants are compared to each other, and, secondarily, whether the audiologic performance of a subject tends to be more consistent over time (Zimmerman and Núñez-Antón, 2010).

4.3 100-Km Race Data

These data set was kindly provided by Ian Jollife of the University of Kent, and originally analysed in 1998 (Zimmerman et al., 1998). The data correspond to each of the partial times in minutes for each of the 80 competitors in each of the 10-kilometer sections of a 100-km race in the United Kingdom in 1984. In addition to the partial times, the data features the age of 76 of the 80 competitors. Some descriptive graphs and exploratory analyses of this data have been previously reported (Everitt, 1994a; Everitt, 1994b). The objective is to find a parsimonious model describing in the best possible way how competitor's performance on each 10-km section is related to the section number $(j = 1, 2, \ldots, 10)$, and performance on previous sections.

5 Applications

In this section we illustrate the usefulness of the proposed Bayesian methodology with the statistical analysis of the three data sets that were briefly introduced in Section 4. Longitudinal models with compound symmetry, CS, autoregressive of order one, AR(1), and autoregressive with moving average, ARMA(1,1), models for the variance-

covariance structure were fitted to the first data set, and structured antedependence models of order one, SAD(1), were fitted to the last two data sets. Unless indicated otherwise, in all of the analyses reported in this section, parameter estimates were obtained from 20000 iterations, after a burn-in of 10000 samples. As specific information on the parameters prior distributions is not available, a $N(\mathbf{b}_0, \mathbf{B}_0)$ distribution was assumed, where independence between the individual distributions for each one of the parameters was assumed, with $\mathbf{b}_0 = (0, \ldots, 0)^{\top}$ and variances for each one of the distributions being equal, so that $\mathbf{B}_0 = \text{diag}(10^k)$, where $k = 5$. In addition, Beta$(1,1)$ prior distributions were assumed for the correlation parameters ρ and ϕ, a Gamma $\left(\frac{g_0}{2}, \frac{g_0\sigma_0^2}{2}\right) \equiv$ Gamma$(10^{-k}, 10^{-k})$, $k = 1, 2, \ldots$ prior distribution was assumed for the variance parameter $\varphi = 1/\sigma^2$, a $U(-1,1)$ uniform prior distribution was assumed for the time-scale transforming parameter λ in equation (9), and a $N(0, \nu^2)$, with $\nu^2 = 1$, distribution was assumed for η, in the Type 2 - SAD model, with $\varphi = \exp(\eta)$. Given that the posterior estimates of φ may change significantly for different values of k, a sensitivity analysis was performed concluding that, for our specific applications, $k = 8$ is the appropriate value minimizing this effect.

5.1 Small Mice Data

From the correlation matrix reported in Table A.1 in the Supplementary Material, it is worth mentioning that there exists a high correlation between consecutive observations or lag-one correlations, with the smallest correlation being the one corresponding to the weights taken between days 5 and 8, and the remaining ones featuring similar values. Moreover, the values for the correlations outside the super diagonal are smaller, but not negligible at all. More specifically, the lag-one correlations range from 0.77 to 0.96, with correlations not being exactly equal, but quite similar to one another, except for the 0.77 value, which is smaller than the others. Thus, it seems reasonable to consider the initial hypothesis that the lag-one correlations are approximately equal. If we move to the lag-two and lag-three correlations, they seem to be quite similar for their first two values, and then their values suddenly increase for the later values. A close analysis of this matrix seems to suggest that there may be two groups of observations, the early ones, corresponding to times 2, 5, 8 and 11, and the late ones, corresponding to times 14, 17 and 20. The former feature a pattern of high lag-one correlations, intermediate values for lag-two correlations, and low lag-three correlations, whereas the latter, if we consider observations at times 11, 14, 17 and 20, all feature high correlations. Based on the above, variance-covariance models such as the CS, AR(1) and ARMA(1,1), as described in Section 2, should be considered for this specific data set. From the correlation matrix reported in Table A.1 (see Supplementary Material), it is worth mentioning that there exists a high correlation between consecutive observations or lag-one correlations, with the smallest correlation being the one corresponding to the weights taken between days 5 and 8, and the remaining ones featuring similar values. Moreover, the values for

the correlations outside the super diagonal are smaller, but not negligible at all. More specifically, the lag-one correlations range from 0.77 to 0.96, with correlations not being exactly equal, but quite similar to one another, except for the 0.77 value, which is smaller than the others. Thus, it seems reasonable to consider the initial hypothesis that the lag-one correlations are approximately equal. If we move to the lag-two and lag-three correlations, they seem to be quite similar for their first two values, and then their values suddenly increase for the later values. A close analysis of this matrix seems to suggest that there may be two groups of observations, the early ones, corresponding to times 2, 5, 8 and 11, and the late ones, corresponding to times 14, 17 and 20. The former feature a pattern of high lag-one correlations, intermediate values for lag-two correlations, and low lag-three correlations, whereas the latter, if we consider observations at times 11, 14, 17 and 20, all feature high correlations. Based on the above, variance-covariance models such as the CS, AR(1) and ARMA(1,1), as described in Section 2, should be considered for this specific data set. Figure A.1 in the Supplementary Material displays the profiles for the different mice in the data set, where we can see that there is an increasing trend for their weights. Given the increasing structure featured by the Small Mice Data (Weiss, 2005), we assume a longitudinal model with the following mean regression structure:

$$Y_{ij} = \beta_0 + \beta_1 \text{Day} + \beta_2 \text{Day}^2 + \epsilon_{ij}, \tag{25}$$

with CS, AR(1) and ARMA(1,1) variance-covariance structures. Parameter estimates are compared to those obtained by applying restricted maximum likelihood methods and reported in Weiss (2005). Tables 1, 2 and 3 include the parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations, and including median values, as well as estimates obtained by restricted maximum Likelihood methods (REML), using the SPSS statistical software package, for the CS, AR(1) and ARMA(1,1) variance-covariance structures, respectively. To implement and obtain the estimates under the Bayesian proposal we have used OpenBugs (Spiegelhalter et al., 2003), together with R (R Core Team, 2013). Based on the estimated parameter values for the different variance-covariance models fitted to the data, we can conclude that estimates and standard deviations under the Bayesian proposals and those obtained by REML are quite similar, which can be used as evidence supporting the fact that the proposed method is behaving as expected and its results are stable under the prior distributional assumptions. However, we should be careful about these conclusions in the sense that this is a very simple, well behaved and balanced data set, and the considered variance-covariance models are, in terms of complexity, very simple and very parsimonious models. More complex variance-covariance models, such as the AD or SAD models, cannot be fitted in most statistical packages and, thus, specific programming is required to fit these models. Selection of the model that best fits the data will be assessed by using the well-known and commonly used Akaike Information Criterion (AIC) (Akaike, 1974), the Bayesian Information Criterion (BIC) (Schwarz, 1978) and

the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). Smaller values of AIC, BIC or DIC indicate better fitting models.

Table 1: *Parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations within parentheses, including median values, and parameter estimates under REML methods for the CS variance-covariance structure for the Small Mice Data.*

Parameter	Mean	Median	REML-estimates
β_0	65.200 (30.158)	64.715	65.745 (29.459)
β_1	70.776 (4.469)	70.766	70.328 (4.406)
β_2	-1.351 (0.198)	-1.350	-1.349 (0.195)
σ^2	9889.113 (2663.290)	9348.963	9671.253 (2621.890)
ρ	0.603 (0.095)	0.606	0.626 (0.107)

Table 2: *Parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations within parentheses, including median values, and parameter estimates under REML methods for the AR(1) variance-covariance structure for the Small Mice Data.*

Parameter	Mean	Median	REML-estimates
β_0	73.843 (28.129)	73.518	74.083 (28.186)
β_1	68.613 (4.089)	68.629	68.588 (3.994)
β_2	-1.252 (0.173)	-1.254	-1.251 (0.169)
σ^2	8622.617 (2663.290)	8130.642	8796.697 (2488)
ρ	0.856 (0.037)	0.857	0.874 (0.038)

Table 3: *Parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations within parentheses, including median values, and parameter estimates under REML methods for the ARMA(1,1) variance-covariance structure for the Small Mice Data.*

Parameter	Mean	Median	REML-estimates
β_0	77.556 (28.099)	77.5126	77.418 (28.494)
β_1	67.687 (4.534)	67.735	67.620 (4.317)
β_2	-1.208 (0.194)	-1.210	-1.204 (0.183)
σ^2	8169.547 (2293.082)	7684.294	8796.697 (2488)
ρ	0.792 (0.0539)	0.797	0.832 (0.055)
ϕ	0.842 (0.034)	0.845	0.8732 (0.035)

Table 4: *Goodness-of-fit AIC, BIC and DIC values for the CS, Bayesian CS-BCS, AR(1), Bayesian AR(1)-BAR(1), ARMA(1,1) and Bayesian ARMA(1,1)-BARMA(1,1) variance-covariance structures for the Small Mice Data.*

Model	AIC	BIC	DIC
CS	1109.3	1110.6	–
BCS	1124.3	1137.2	1122.9
AR(1)	1039.3	1040.5	–
BAR(1)	1054.7	1067.6	1053.3
ARMA(1,1)	1038.9	1040.8	–
BARMA(1,1)	1054.6	1070.1	1052.7

Alternative more recent model selection criteria include the Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010). Given that the main objective here is the proposal of a Bayesian methodology and its comparison with previous maximum likelihood estimation methods, we have used the AIC and BIC criteria to be able to assess and compare the performance of our models with those previously fitted. In any case, and given that the DIC and WAIC are standard model evaluation tools and considered more appropriate criteria for model selection purposes within the Bayesian framework (Watanabe, 2010; Choi, Jang and Alemi, 2018), we have also provided the DIC values for some of the models fitted here. Moreover, we believe that model selection, within the Bayesian framework, and given the common use of the DIC criterion and the well known advantages of the WAIC criterion, should be proposed together with the use of both model selection criteria (Piironen and Vehtari, 2017; Vehtari, Gelman and Gabry, 2017). However, in our view and given that our main objective is to compare and assess the behaviour of our Bayesian proposals with previous non-Bayesian approaches, we consider that the reported AIC and BIC criteria are appropriate within this context, specially taking into account that we have assumed vague prior distributions. Table 4 includes the AIC, BIC and DIC values for the CS, Bayesian CS-BCS, AR(1), Bayesian AR(1)-BAR(1), ARMA(1,1) and Bayesian ARMA(1,1) variance-covariance structures for the Small Mice Data. Based on these values and keeping in mind that these are simple models than can be fitted by using REML methods in SPSS or other alternative statistical packages, the best fitting model based on AIC is the ARMA(1,1) model, with the AR(1) model being a close competitor. If we use BIC, the best fitting model is the AR(1), with the ARMA(1,1) model being also a close competitor. The same conclusion is reached if DIC is used as a model selection criterion, with the BARMA(1,1) model being the best fitting one, and the BAR(1) following quite closely, a fact that is also supported is we use the AIC or BIC criteria for the Bayesian model proposals. In summary, for the Small Mice data, the best fitting models are the autoregressive model of order one and the autoregressive model with moving average, ARMA(1,1), model.

A final remark on the basis of a comment raised by an anonymous reviewer is that practitioners may consider using the logarithm of the weight as a response variable instead. They should be aware that when we have higher variance sample values for the different time points it may be convenient to use this transformation so that these variance values may be more parsimoniously modelled. However, given that our main objective was to compare our results to those in previous analyses (see, e.g., Weiss, 2005), and also to assess if the proposed methodology was able to model specific variance and correlation behaviours, such as the ones in the small mice data, we decided to use weight as the response variable for the analyses reported here.

5.2 Speech Recognition Data

Previous analyses (Zimmerman et al., 1998) reported that the likelihood-ratio test for the equality of the within-group covariance matrices indicated that it was reasonable to

pool them ($p = 0.35$). From the pooled correlation matrix reported in Table A.2 in the Supplementary Material, it is worth mentioning that correlations are positive and quite large, that correlations between test scores at times t and $t+k$ seem to decrease monotonically as k increases, and that correlations between test scores at adjacent measurement times increase over time. This latter statement is somehow consistent with a prior belief that subjects may "learn" over time, as with the result that responses equidistant in time become more highly correlated as the study progressed, which is a clear sign of nonstationary correlation structures, such as the one modelled by SAD-type models and the proposed time-transforming scale in equation (9). In addition, variances seem to be homogeneous at all times points except for the first one. Based on the above, variance-covariance models such as the SAD, as described in Section 2, should be considered for this specific data set. Figure A.2 in the Supplementary Material displays the profiles for the different individuals for each type of implant. As Zimmerman and Núñez-Antón (2010) have already mentioned in previously presented exploratory analyses, these plots suggest that there is an increasing trend for the mean audiologic performance, at least for the initial months, and that audiologic performance seems to stabilize for the later months, which provides some empirical evidence for the consistency of audiologic performance over time. These plots also suggest that variances seem to increase slightly from the first to the second measurement, but remain constant thereafter. Several previous different models were fitted (Zimmerman and Núñez-Antón, 2010), such as, for example, homogeneous and heterogeneous versions of the CS and AR(1) models, but finally concluded that the best fitting models models for this data are the structured antedependence model of order one or SAD(1) models. Given the increasing mean featured by this data (Núñez-Antón and Woodworth, 1994), in order to be able to compare our results when fitting the Type 1 - SAD model, we initially propose the mean regression structure:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \epsilon_{ij}, t_{i1} = 1,\ t_{i2} = 9,\ t_{i3} = 18,\ t_{i4} = 30 \tag{26}$$

with the Type 1 - SAD model variance-covariance structure described in Section 2. Table 5 includes the parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations, and including median values, as well as estimates obtained by restricted maximum Likelihood methods (REML), previously reported by Núñez-Antón and Woodworth (1994), where standard deviations for the variance-covariance parameters were not provided. Initial values for the regression parameters were assumed so that $\boldsymbol{\beta}_0 = (20, 1, 0)^{\intercal}$. In addition, initial values for the Bayesian estimation were assumed so that $\rho_0 = 0.50$, $\lambda_0 = 0.50$ and $\sigma_0^2 = 100$. The acceptance rates for ρ and λ were equal to 39% and 32%, respectively. The corresponding goodness-of-fit information criteria values for this model were AIC=1322.752, BIC=1341.351, and DIC=1320.714. Núñez-Antón and Woodworth (1994) did not report these values in earlier analyses, and their computation was not straightforward unless specific programs to fit the proposed model are implemented. This issue is clearly out of the scope

of this paper. Based on the estimated parameter values reported in Table A.2 in the Supplementary Material, we can conclude that estimates and standard deviations under the Bayesian proposal and those obtained by REML are quite similar, which can be used as evidence supporting the fact that the proposed method is behaving as expected and its results are stable under the prior distributional assumptions. In addition, fitting of this not so parameterized and parsimonious model by REML methods requires a more specific and complex programming and maximization than the ones proposed in this paper. As an illustration of fitting the Type 2 - SAD model and in order to be able to compare the results obtained with our proposed methodology, we fitted the same model previously proposed (Zimmerman et al., 1998), with mean regression structure given by:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 z_i + \beta_4 z_i t_{ij} + \beta_5 z_i t^2 + \epsilon_{ij}, \tag{27}$$

and also with $t_{i1} = 1$, $t_{i2} = 9$, $t_{i3} = 18$, $t_{i4} = 30$, and $z_i = 1$ if the i-th individual received implant type A, and $z_i = 0$, otherwise. As for the variance-covariance structure, we assume a Type 2 - SAD model given by (4) and (10). Table 6 includes the parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations, as well as median values. Values previously obtained by restricted maximum likelihood methods (REML) were not reported (Zimmerman et al., 1998). Initial values for the regression parameters were assumed so that $\boldsymbol{\beta}_0 = (20,1,0,8,1,0)^{\top}$. In addition, initial values for the Bayesian estimation were assumed so that $\rho_0 = 0.50$, $\lambda_0 = 0.50$, $\psi_0 = 1$ and $\sigma_0^2 = 100$. Table 7 includes the parameter mean values under the Bayesian proposal for the variance-covariance parameters, together with their standard deviations, including median values, for the Type 2 - SAD variance-covariance structure, as well as those obtained by restricted maximum Likelihood methods (REML) (Zimmerman et al., 1998), where standard deviations for the variance-covariance parameters were not provided.

Table 5: *Parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations within parentheses, including median values, and parameter estimates under REML methods for the Type 1 - SAD variance-covariance structure for the Speech Recognition Data.*

Parameter	Mean	Median	REML-estimates
β_0	22.330 (4.294)	22.375	22.850 (4.260)
β_1	2.537 (0.313)	2.540	2.520 (0.340)
β_2	-0.048 (0.008)	-0.048	-0.04695 (0.009)
σ^2	602.028 (112.636)	585.562	587.15
ρ	0.933 (0.025)	0.9395	0.940
λ	0.297 (0.144)	0.300	0.300

The acceptance rates for ρ, λ and ψ were equal to 35%, 32% and 34%, respectively. The corresponding goodness-of-fit information criteria values for this model were AIC=1287.364, BIC=1318.362, and DIC=1283.554. These values were not originally reported in previous analyses (Zimmerman et al., 1998), and their computation is not

straightforward unless specific programs to fit the proposed model are implemented. This issue is clearly out of the scope of this paper. Based on the estimated parameter values reported in Table 7, we can conclude that estimates under the Bayesian proposal and those obtained by REML are comparable, except for parameter σ^2, which can be used as evidence supporting the fact that the proposed method is behaving as expected and its results are stable under the prior distributional assumptions. In addition, fitting of this not so parameterized and parsimonious model by REML methods requires a more specific and complex programming and maximization than the ones proposed in this paper. Moreover, estimates reported in Table 6 for the mean regression structure do not support the conclusions previously reported (Zimmerman et al., 1998) with regard to the significance of the parameter β_5 in (27). Given the robustness of the proposed methodology, the above differences could question the appropriateness of estimates obtained by REML methods.

Table 6: *Parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations within parentheses, including median values, for the Type 2 - SAD variance-covariance structure for the Speech Recognition Data.*

Parameter	**Mean**	**Median**
β_0	13.827 (3.959)	13.837
β_1	2.249 (0.386)	2.246
β_2	-0.044 (0.010)	-0.044
β_3	14.719 (5.663)	14.634
β_4	0.395 (0.551)	0.398
β_5	-0.009 (0.015)	-0.009

Table 7: *Parameter estimated mean values for the variance-covariance structure under the Bayesian proposal, together with their respective standard deviations in parentheses, including median values, for the Type 2 - SAD variance-covariance structure for the Speech Recognition Data.*

Parameter	σ^2	ρ	λ	ψ
Mean	334.046	0.928	0.323	1.773
Standard Deviation	71.858	0.0381	0.192	0.316
Median	325.052	0.936	0.330	1.746
REML-estimates	388.7	0.935	0.240	1.615

Table 8: *Parameter estimated mean values under the Bayesian proposal for the Type 3 - SAD variance-covariance structure, together with their respective standard deviations within parentheses, including median values, and parameter estimates under REML-methods for the 100-Km Race Data.*

Parameter	**Mean**	**Median**	**REML-estimates**
β_0	44.585 (1.632)	44.573	43.428
β_1	-2.410 (2.102)	-2.421	1.354
β_2	1.327 (0.752)	1.326	0.253
β_3	-0.097 (0.072)	-0.097	-0.017

5.3 100-Km Race Data

From the correlation matrix and sample variance values reported in Table A.3 in the Supplementary Material, it can be observed that variances tend to increase as the race progresses, that the correlations among split times are positive and quite large, that the correlations between the split time for a fixed 10-Km section and split times for successive sections tend to decrease monotonically, and that the correlations between split times for adjacent sections are smaller in the later sections of the race than in the earlier sections. In addition, variances seem to increase as the race progresses, with the exception of the seventh section of the race. Based on the above, variance-covariance models such as the SAD, as described in Section 2, should be considered for this specific data set. Figure A.3 in the Supplementary Material displays the profiles for the individuals in the data set, where we can see that there is an increasing trend for the times as the race progresses. In addition, variances for the different sections also seem to increase monotonically. Based on the above, some authors (Zimmerman et al., 1998) have previously suggested the fitting of an SAD model of order one, as well as a cubic in time mean regression model, so that:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + \epsilon_{ij},\ i = 1,\ldots,80;\ j = 1,\ldots,10 \tag{28}$$

As for the variance-covariance structure, we assume a Type 3 - SAD model with variances given by $\sigma_j^2 = \exp(\psi_0 + \psi_1 t_{ij} + \psi_2 t_{ij}^2),\ \ j = 1,\ldots,10,$, and covariance structure given by (4). In the model proposal for the Type 3 - SAD model, there is a slight difference with that in previous analyses (Zimmerman et al., 1998), where the proposed variance function was, instead, $\sigma_j^2 = \sigma^2(1 + \psi_1 t_{ij} + \psi_2 t_{ij}^2),\ \ j = 1,\ldots,10$. Therefore, parameter estimates are not directly comparable. Our variance model variation was necessary for the Bayesian proposal in this paper. In the data analysis reported here, parameter estimates were obtained from 15000 iterations, after a burn-in of 5000 samples. Initial values for the regression parameters were assumed so that $\boldsymbol{\beta}_0 = (20,1,0,0)^{\intercal}$. In addition, initial values for the Bayesian estimation were assumed so that $\rho_0 = 0.50$, $\lambda_0 = 0.50$, $\boldsymbol{\psi}_0 = (1,1,1)^{\intercal}$ and $\mathbf{K}_0 = \text{diag}(k_0, k_0, k_0)$, with $k_0 = 0.1384$. Table 8 includes the regression parameter estimated mean values under the Bayesian proposal, together with their respective standard deviations, and including median values, as well as estimates obtained by restricted maximum Likelihood methods (REML) (Zimmerman et al., 1998), where standard deviations for the variance-covariance parameters were not provided. It is worth mentioning that, even though there are differences between the REML estimates and those obtained by the Bayesian proposal, the values previously reported (Zimmerman et al., 1998) for the regression parameters are all within the 95% credibility intervals listed here: $\text{CI}(0.95)_{\beta_1} = (-11.77, 6.95)$, $\text{CI}(0.95)_{\beta_2} = (-4.90, 7.547)$, and $\text{CI}(0.95)_{\beta_3} = (-0.5433, 0.3513)$, which were generated with the obtained estimated values under the Bayesian proposal. Table 9 includes the estimated values for the variance-covariance parameters under the Bayesian proposal, together with their respective standard deviations, and including median values, as well as estimates obtained by

restricted maximum Likelihood methods (REML), when available (Zimmerman et al., 1998), where standard deviations for the variance-covariance parameters were not provided. In any case and in order to be able to compare the estimated variances at each split time, we also include their REML-estimates for the variance parameters: $\hat{\sigma}^2 = 16.952$, $\hat{\psi}_1 = 0.590$, and $\hat{\psi}_2 = 0.450$.

Table 9: *Parameter estimated mean values for the variance-covariance parameters under the Bayesian proposal for the Type 3 - SAD variance-covariance structure, together with their respective standard deviations within parentheses, including median values, and parameter estimates under REML-methods, when available, for the 100-Km Race Data.*

Parameter	Mean	Median	REML-estimates
ρ	0.918 (0.031)	0.924	0.929
λ	1.680 (0.261)	1.684	1.600
ψ_0	2.771 (0.308)	2.767	–
ψ_1	0.677 (2.128)	0.683	–
ψ_2	-0.034 (0.021)	-0.034	–

The acceptance rates for ρ, λ and $\boldsymbol{\psi}$, the latter resulting from the working variable in equation (23), were equal to 34%, 37% and 84%, respectively. The corresponding goodness-of-fit information criteria values for this model were AIC = 1401.88, BIC = 1425.31, and DIC = 1403.078. These values were not reported in previous analyses (Zimmerman et al., 1998), and their computation is not straightforward unless specific programs to fit the proposed model are implemented. This issue is clearly out of the scope of this paper. In order to better assess the behaviour of the estimated split time variances obtained under the Bayesian proposal, we have computed the estimated variances under our proposal and under the REML method proposal (Zimmerman et al., 1998) and report this information, as well as the estimated sample variance values obtained from the data, in Table 10. In our opinion, it is clear that the Bayesian and REML estimated values for each of the sections in the race differ from each other, as well as from the reported sample values. However, it is worth mentioning that the observed increase for the estimated variances under the Bayesian proposal is smaller than that obtained under the REML methods.

Based on the estimated parameter values reported in Tables 8, 9 and 10, we can conclude that estimates under the Bayesian proposal and those obtained by REML are not exactly similar, but show a similar behaviour, which can be used as evidence supporting the fact that the proposed method is behaving as expected and its results are quite stable even under very general prior distributional assumptions. In addition, fitting of this not so parameterized and parsimonious model by REML methods requires a more specific and complex programming and maximization than the ones proposed in this paper. Given the robustness of the proposed methodology, the above differences could question the appropriateness of estimates obtained by REML methods. In addition, to be able to better compare mean split times estimated values under the Bayesian proposal with the corresponding fitted values that can be obtained from the estimates previously

reported (Zimmerman et al., 1998), Figure A.4 in the Supplementary Material shows the residuals obtained for the 100-Km Race Data obtained under the Bayesian and classic REML methods for the Type 3 - SAD variance-covariance structure. As can be seen from this figure, there are no significant differences between the residuals resulting from the model estimation by classic REML estimation and Bayesian estimation methods. Moreover, the residual sum of squares computed on the model estimated by restricted maximum likelihood methods is RSS $= 83455.21$, whereas the corresponding one for the proposed Bayesian method is RSS $= 94598.3$. As an additional way of comparing the behaviour of the residuals for each section of the race, Figures A.5 and A.6 in the Supplementary Material include the corresponding boxplots for the residuals resulting from the REML and Bayesian method proposals. Conclusions that can be obtained from the information provided in these figures suggest that residuals for the different sections of the race obtained by the two methods do not significantly differ from each other, which supports the claim that results obtained by the REML classic methodology can be well approximated by means of a simpler and more flexible Bayesian method, such as the one included in this manuscript.

Table 10: *Estimated sample variances, and parameter estimated variances under the Bayesian proposal for the Type 3 - SAD variance-covariance structure, and REML-methods for the 100-Km Race Data.*

Parameter	Sample values	REML-estimates	Bayesian estimates
σ_1^2	26.89	34.58	31.01
σ_2^2	34.78	67.48	54.60
σ_3^2	49.01	115.61	90.08
σ_4^2	58.89	179.013	139.28
σ_5^2	91.41	257.67	201.82
σ_6^2	149.90	351.58	274.05
σ_7^2	107.85	460.75	348.73
σ_8^2	152.22	585.18	415.86
σ_9^2	144.99	724.86	464.73
σ_{10}^2	167.21	879.80	486.70

5.4 Sensitivity analysis

In this section we study the behaviour of the Bayesian estimate for the variance σ^2 under different values of the hyperparameters used in the assumed prior distribution for $\varphi = 1/\sigma^2$, which, as already mentioned in previous sections, was assumed to be a Gamma(k,k) distribution, with $k = 10^{-5}$. In this case, we illustrate this behaviour in the analysis of the three different models (i.e., CS, AR(1) and ARMA(1,1)) fitted to the Small Mice Data (SMD) and the two SAD (i.e., Type 1 and Type 2) models fitted to the Speech Recognition Data (SRD). Changes in the estimated values for σ^2 are observed

for different values of k in the gamma distribution, such that the assumed values for the hyperparameter k for this analysis are $k = 1 \times 10^{-3}$, $k = 1 \times 10^{-5}$, $k = 1 \times 10^{-8}$ and $k = 1 \times 10^{-10}$. Table 11 includes the average variance estimated value of the chains by means of a Gibbs sample of the resulted conditional posterior distribution, together with their corresponding standard deviations in parentheses, for different values of the hyperparameter k in the prior Gamma(k,k) distribution assumed for $\varphi = 1/\sigma^2$. Fitted models correspond to the SC, AR(1) and ARMA(1,1) models for the Small Mice Data (SMD) and to the Type 1 and Type 2 - SAD models for the Speech Recognition Data (SRD). From the information reported in Table 11, we can conclude that, as the hyperparameter in the assumed gamma distribution becomes smaller, the standard deviation and estimated values obtained under the Bayesian proposal approach those obtained by the REML estimating method. In addition, and given that variance estimates and their standard deviations obtained for values of $k = 1 \times 10^{-8}$ and $k = 1 \times 10^{-10}$ are quite similar, we can conclude that once the value of k in the prior distribution is equal to 1×10^{-8}, changes in the means of the corresponding chains are negligible, and this was the main reason for the use of this specific hyperparameter value in the prior distribution assumed for the analysis of the three data sets in Section 5.

Table 11: *Estimated variances, together with their corresponding standard deviations within parentheses, for different values of the hyperparameter k in the prior* Gamma(k,k) *distribution assumed for* $\varphi = 1/\sigma^2$. *Fitted models correspond to the CS, AR(1) and ARMA(1,1) models for the Small Mice Data (SMD) and to the Type 1 and Type 2 - SAD models for the Speech Recognition Data (SRD).*

k	1×10^{-3}	1×10^{-5}	1×10^{-8}	1×10^{-10}
SMD-CS	12302.35 (4103.29)	10672.29 (3504.29)	9889.11 (2663.29)	9890.02 (2662.79)
SMD-AR(1)	10025.26 (3234.567)	9989.30 (3012.23)	8622.617 (2663.29)	8621.85 (2661.95)
SMD-ARMA(1.1)	10054.53 (2997.72)	9867.56 (2900.32)	8169.55 (2488.56)	8169.00 (2488.32)
SRD-Type 1 SAD	768.34 (172.3452)	727.02 (156.14)	602.03 (112.64)	602.24 (112.45)
SRD-Type 2 SAD	380.65 (83.579)	372.99 (78.93)	334.05 (71.86)	334.05 (71.81)

6 Conclusions and final recommendations

We have proposed alternative Bayesian longitudinal models for fitting compound symmetry, autoregressive or order one, autoregressive with moving averages, as well as unstructured and structured antedependence models for nonstationary in variance and/or correlation longitudinal data settings. Very flexible distributional prior assumptions were proposed, and the specific methods to obtain the conditional posterior distribution were

described. The usefulness of the proposed method was illustrated with the analysis of the Small Mice Data, the Speech Recognition Data and the 100-Km Race Data, and results were compared to those obtained by restricted maximum likelihood methods. Results suggested that the proposed methods behave well under general conditions, and estimated values are in line with with those obtained by classic methods. However, classic methods require specific programming, whereas the proposed Bayesian methods can be easily adjusted to the data sets under study by using very flexible and easy programming, as well as general available software, such as R and OpenBugs. Future work includes extending these proposals to more complex unstructured and structured antedependence higher order models. Finally, we would like to mention that, even though the proposed Bayesian methodology has been shown to have a fast convergence and reasonable acceptance rates, our future research in the area includes the study of acceptance rates improvements in terms of making the proposed methodology more efficient, providing at the same time recommendations useful for researchers in the area. In practice, acceptance rates are known to improve with the adequate selection of initial values from the information available in the data, as well as from the appropriate parameter selection for the prior distributions. They do so by making use of a thorough analysis of the available prior information, such as, for example, variables rank or proposed models motivation and/or parameterization. For example, if we let $\varphi' = \log(\varphi)$ and we assume a normal prior distribution for φ' instead in the CS and AR(1) models. Convergence rates and acceptance rates can also be improved by applying alternative Monte Carlo resampling methods, such as the reduced-rejection-rate method (Baldassi, 2017). This parametrization of φ can be also important for the aforementioned problem of sensitivity of the posterior variance (i.e., precision) estimates to the gamma prior distributions Gamma$(10^{-k}, 10^{-k})$ assumption, for $k = 1,2,3,4,\ldots$. Thus, when this prior distribution is assumed and no prior information on φ is available, a sensitivity analysis, like the one described in Section 5.4, should always be included in any statistical data analysis, so that the sensitivity of φ to the smallest changes in the value of k is minimized.

Acknowledgments

This work was supported by the Department of Statistics, Faculty of Sciences of the Universidad Nacional de Colombia, Ministerio de Economía y Competitividad, Agencia Estatal de Investigación (AEI), Fondo Europeo de Desarrollo Regional (FEDER), the Department of Education of the Basque Government (UPV/EHU Econometrics Research Group), and Universidad del País Vasco UPV/EHU under research grants MTM2013-40941-P (AEI/FEDER, UE), MTM2016-74931-P (AEI/FEDER, UE), IT-642-13, IT1359-19 and UFI11/03.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.

Baldassi, C. (2017). A method to reduce the rejection rate in Monte Carlo Markov chains. *Journal of Statistical Mechanics: Theory and Experiment*, 2017, 033301.

Brown, P.J., Kenward, M.G. and Bassett, E.E. (2001). Bayesian discrimination with longitudinal data. *Biostatistics*, 2, 417–432.

Cepeda-Cuervo, E. (2001). Modelagem da Variabilidade em Modelos Lineares Generalizados. Unpublished Math Ph.D. Thesis. Mathematics Institute, Universidade Federal do Río de Janeiro, Brazil.

Cepeda-Cuervo, E. (2011). Generalized spatio-temporal models. *SORT*, 35, 165–178.

Cepeda, E.C. and Gamerman, D. (2004). Bayesian modeling of joint regressions for the mean and covariance matrix. *Biometrical Journal*, 46, 430–440.

Cepeda-Cuervo, E. and Núñez-Antón, V. (2007). Bayesian joint modelling of the mean and covariance structures for normal longitudinal data. *SORT*, 31, 181–200.

Cepeda-Cuervo, E. and Núñez-Antón, V. (2009). Bayesian modelling of the mean and covariance matrix in normal nonlinear models. *Journal of Statistical Computation and Simulation*, 79, 837–853.

Choi, H., Jang, E. and Alemi, A.A. (2018). WAIC, but Why? Generative ensembles for robust anomaly detection. arXiv preprint arXiv:1810.01392.

Congdon, P. (2020). *Bayesian Hierarchical Models With Applications Using R*. Second edition. CRC Press, London.

Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger, S. (2002). *Analysis of Longitudinal Data (2nd edition)*. Oxford University Press, Oxford.

Everitt, B. (1994a). Exploring multivariate data graphically: A brief review with examples. *Journal of Applied Statistics*, 21, 63–94.

Everitt, B. (1994b). *A Handbook of Statistical Analyses Using S-Plus*. Chapman and Hall, London.

Fahrmeir, L., Kneib, T. and Lang, S. (2013). Bayesian multilevel models. In: Scott, M.A., Simonoff, J.S. and Marx. B.D. (eds). *The SAGE Handbook of Multilevel Modeling*. SAGE, London, 53–71.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2009). *Longitudinal Data Analysis*. CRC Press, New York.

Gabriel, K.R. (1962). Ante-dependence analysis of an ordered set of variables. *Annals of Mathematical Statistics*, 33, 201–212.

Gamerman, D. and Lopes, H.F. (2006). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference (2nd edition)*. CRC Press, New York.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515–533.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., (2014a). *Hierarchical Linear Models in Bayesian Data Analysis*. Chapman and Hall/CRC Press, Boca Raton, FL.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2014b). *Bayesian Data Analysis, vol. 2*. Chapman and Hall/CRC Press, Boca Raton, FL.

Gill, J. (2014). *Bayesian Hierarchical Models in Bayesian Methods: A Social and Behavioral Sciences Approach*. CRC Press, Boca Raton, FL.

Hui, S.L. and Berger, J.O. (1983). Empirical Bayes estimation of rates in longitudinal studies. *Journal of the American Statistical Association*, 78, 753–760.

Izenman, A.J. and Williams, J.S. (1989). A class of linear spectral models and analyses for the study of longitudinal data. *Biometrics*, 45, 831–849.

Jaffrézic, F. and Pletcher, S.D. (2000). Statistical Models for estimating the genetic basis of repeated measures and other function-valued traits. *Genetics*, 156, 913–922.

Jaffrézic, F., White, I.M.S., Thompson. R. and Visscher, P.M. (2002). Contrasting models for lactation curve analysis. *Journal of Dairy Science*, 85, 968–975.

Jiang, J., Zhang, Q., Ma, L., Li, J., Wang, Z. and Liu, J.-F. (2015). Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity*, 115, 29–36.

Laird, N.M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, 7, 305–315.

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

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.

Núñez-Antón, V. and Zimmerman, D.L. (2001). Modelización de datos longitudinales con estructuras de covarianza no estacionarias: Modelos de coeficientes aleatorios frente a modelos alternativos. *Qüestiió*, 25, 225–262.

Piironen, J. and Vehtari, A. (2017). Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, 27, 711–735.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86, 677–690.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society - Series B*, 64, 583–616.

Spiegelhalter, D.J., Thomas, A., Best, N. and Lunn, D. (2003). *WinBUGS User Manual*. MRC Biostatistics Unit, Cambridge, UK. www.mrc-bsu.cam.ac.uk/bugs.

Tyler, R.S., Abbas, P., Tye-Murray, N., Gantz, B.J., Knutson, J.F., McCabe, B.F., Lansing, C., Brown, C., Woodworth, G., Hinrichs, J. and Kuk, F. (1988). Evaluation of five different cochlear implant designs: Audiologic assessment and predictors of performance. *The Laryngoscope*, 98, 1100–1106.

Vehtari, A., Gelman, A. and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out-cross-validation and WAIC. *Statistics and Computing*, 27, 1413–1432.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Ware, J.H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, 39, 95–101.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.

Weiss, R.E. (2005). *Modeling Longitudinal Data*. Springer, New York.

Yang, W. and Tempelman, R.J. (2012). A Bayesian Antedependence Model for Whole Genome Prediction. *Genetics*, 190, 1491–1501.

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). Modelling longitudinal and spatially correlated data: Antedependence models for longitudinal data. In: Gregoire, T.G., Brillinger, D.R., Diggle, P.J., Russell-Cohen, E., Warren, W.G. and Wolfinger, E. (eds). *Modelling Longitudinal and Spatially Correlated Data. Methods, Applications, and Future Directions*. Springer-Verlag, New York. Lecture Notes in Statistics No. 122, 63–76.

Zimmerman, D.L and Núñez-Antón, V. (2010). *Antedependence Models for Longitudinal Data*. CRC Press, New York.

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.