Biometrika (2002), **89,** 3, *pp.* 553–566
© 2002 Biometrika Trust
Printed in Great Britain

Bayesian analysis of covariance matrices and dynamic models for longitudinal data

BY MICHAEL J. DANIELS

Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.

mdaniels@iastate.edu

AND MOHSEN POURAHMADI

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

pourahm@math.niu.edu

SUMMARY

Parsimonious modelling of the within-subject covariance structure while heeding its positive-definiteness is of great importance in the analysis of longitudinal data. Using the Cholesky decomposition and the ensuing unconstrained and statistically meaningful reparameterisation, we provide a convenient and intuitive framework for developing conditionally conjugate prior distributions for covariance matrices and show their connections with generalised inverse Wishart priors. Our priors offer many advantages with regard to elicitation, positive definiteness, computations using Gibbs sampling, shrinking covariances toward a particular structure with considerable flexibility, and modelling covariances using covariates. Bayesian estimation methods are developed and the results are compared using two simulation studies. These simulations suggest simpler and more suitable priors for the covariance structure of longitudinal data.

Some key words: Antedependence and autoregressive models; Bayes estimate; Hierarchical model; Markov chain Monte Carlo; Mixed model; Shrinkage estimator; Time series model; Unconstrained parameterisation; Wishart distribution.

1. INTRODUCTION

Repeated measures on the same subject over time are typically correlated and there are several approaches in the literature for handling stationary correlation in longitudinal data (Diggle et al., 1994). A powerful approach to modelling nonstationary and unbalanced longitudinal data, however, is based on the linear mixed model (Laird & Ware, 1982)

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \quad (i = 1, \ldots, m), \tag{1}$$

where Y_i is an $n_i \times 1$ vector of responses measured on the ith subject, β is a $p \times 1$ vector of unknown fixed effects parameters, b_i is a $q \times 1$ vector of unknown random effect parameters, and X_i and Z_i are known $n_i \times p$ and $n_i \times q$ design matrices. The dual role of the random term $Z_i b_i$ in modelling covariances using covariates and individual effects, i.e. conditional means, is one of the most important features of mixed models (Searle et al., 1992, Ch. 7) which we would like to emulate here.

Our goal in this follow-up paper to Pourahmadi (1999, 2000) is to introduce new priors

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

for a covariance matrix and Bayesian hierarchical models for shrinking a covariance matrix towards structure and/or for introducing covariates using models of the form (1). The key idea is the unconstrained reparameterisation of the covariance matrix Σ of a random vector $Y=(y_1,\ldots,y_n)'$ which hinges on diagonalising Σ using a triangular matrix:

$$T\Sigma T'=D, \tag{2}$$

where T is a unique unit lower triangular matrix having 1's on its diagonal and D is diagonal with positive diagonal entries. The (T, D)-reparameterisation of Σ in (2) along with $p(\Sigma)=p(T|D)p(D)$ provide a convenient framework for developing prior distributions for covariance matrices. The standard prior distribution for the covariance matrix is the inverse Wishart which offers the advantage of being the conjugate family to the covariance matrix of a normal distribution. The lack of flexibility of this distribution is often overlooked in view of the ease of computation it allows. More flexible priors have been developed in recent years, for example by Leonard & Hsu (1992), Daniels & Kass (1999) and Barnard et al. (2000) and in a University of New South Wales technical report by M. Smith and R. Kohn, but their use has been limited by the computational burden and statistical interpretation. We develop a conditionally conjugate prior and show its relationship with the generalised inverse Wishart prior in Brown et al. (1994) and the inverse Wishart prior. We also develop a hierarchical shrinkage prior, similar to those developed in Daniels & Kass (1999), which offers considerable advantages in terms of computations.

A reinterpretation of the factorisation (2) leads to dynamic models for the vector Y_i of repeated measurements on the ith subject which are similar to (1), but, in sharp contrast to (1), the Z_i is random depending on the response Y_i and β, and b_i is fixed and nonrandom. This generic model appears to be universal, in the sense that it can be used to model an unstructured covariance matrix, and subsumes a plethora of dynamic models (Anderson, 1978; Jones, 1993, Ch. 5; Tsimikas & Ledolter, 1998; West & Harrison, 1997).

Section 2 reviews the dynamic model representations of longitudinal data. Several examples illuminate how these dynamic models subsume general linear models, stationary models, antedependence (Gabriel, 1962) and other nonstationary models. An iteratively reweighted least-squares method for estimating their parameters is given in Appendix 1. In § 3, we illustrate the connection to, and considerable advantages of these models for Bayesian modelling of a covariance matrix. In both frequentist and Bayesian approaches to covariance modelling based on (2), the estimated covariance matrix is guaranteed to be positive definite. We offer extensions in § 4.

2. DYNAMIC REPRESENTATION OF LONGITUDINAL DATA

2·1. *Dynamic models for a random vector*

For a time-ordered random vector $Y=(y_1,\ldots,y_n)'$ with zero mean and positive definite covariance matrix Σ, let $\hat{y}_t$ be the linear least-squares predictor of y_t based on its predecessors $y_{t-1},\ldots,y_1$ and let $\varepsilon_t=y_t-\hat{y}_t$ be its prediction error with variance $\sigma_t^2=\mathrm{var}(\varepsilon_t)$. Then, for some unique scalars ϕ_{tj} we have

$$y_t=\sum_{j=1}^{t-1}\phi_{tj}y_j+\varepsilon_t, \tag{3}$$

where the ϕ_{tj}'s and the variances σ_t^2 are easily computed and are statistically meaningful (Pourahmadi, 1999). With $D=\mathrm{diag}(\sigma_1^2,\ldots,\sigma_n^2)$, T the unit lower triangular matrix with $-\phi_{tj}$ as its (t,j)th entry and (3) written in matrix form, it follows that (2) is satisfied, so

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

that T and D are the components of the square-root-free Cholesky decomposition of Σ (Tanabe & Sagae, 1992).

When Σ is unstructured, the nonredundant entries of T and

$$\log D = \mathrm{diag}(\log \sigma_1^2, \ldots, \log \sigma_n^2)$$

are unconstrained, and the $n(n+1)/2$ constrained and hard-to-model parameters of Σ can be swapped with the same number of unconstrained parameters of the unique pair $(T, \log D)$. The new parameters ϕ_{tj} and σ_t^2 are referred to as the generalised autoregressive parameters and the innovation variances of Σ. Thus, the dimension of the parameter space can be reduced considerably by using covariates (Pourahmadi, 1999, 2000). To this end, for $t = 1, \ldots, n$ and $j = 1, \ldots, t-1$, let

$$\phi_{tj} = z_{tj}'\gamma, \quad \log \sigma_t^2 = z_t'\lambda, \tag{4}$$

where z_{tj} and z_t are $d \times 1$ and $q \times 1$ vectors of covariates and γ and λ are $d \times 1$ and $q \times 1$ vectors of unknown parameters for the dependence and variance of Y, respectively. Substituting for the ϕ_{tj}'s from (4) in (3) we obtain

$$y_t = \left(\sum_{j=1}^{t-1} z_{tj} y_j \right)' \gamma + \varepsilon_t = Z(t)'\gamma + \varepsilon_t, \tag{5}$$

where the design vector

$$Z(t) = \sum_{j=1}^{t-1} z_{tj} y_j \tag{6}$$

is a stochastically weighted sum of the covariates z_{tj} with the lagged y_j's as the weights. Note that $Z(1) = 0$ and, with $Z = (Z(1), \ldots, Z(n))'$, (5) written in the matrix form is a dynamic linear model:

$$Y = Z\gamma + \varepsilon, \tag{7}$$

where the design matrix Z depends only on the lagged response variable (Anderson, 1978). However, when the mean of Y is nonzero and modelled as $\mu = E(Y) = X\beta$, then replacing Y by $Y - X\beta$ in (7) leads to

$$Y = X\beta + Z\gamma + \varepsilon, \tag{8}$$

where Z does now depend on β since it is obtained from (6) upon replacing y_j by $y_j - \mu_j$.

In general, (8) is a dynamic nonlinear model in β and γ; it is linear in γ for fixed β. Heuristically, such decomposition of Y into three components involving its mean, dependence and variance could make the task of statistical modelling of these components more manageable and suggest appropriate estimation strategies. For example, developing an iteratively reweighted least-squares method for estimating $(\beta', \gamma')'$ and λ is now within easy reach; see Appendix 1. Also, the similarity of (1) and (8) puts the latter within the scope of linear mixed and state-space models (Searle et al., 1992, Ch. 6; Jones, 1993; West & Harrison, 1997, Ch. 16; Tsimikas & Ledolter, 1998). Thus, in principle, one could rely on PROC MIXED (SAS Institute, 1997) and BUGS (Spiegelhalter et al., 1996) to fit such models.

2·2. *Examples*

The following examples shed light on the interplay between Z and the dependence in Y, and on the interpretation of the ϕ_{tj}'s and σ_t^2's for some structured covariances arising from time series analysis and categorical data.

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

Example 1: *Variance heterogeneity.* Let the entries of Y be independent random variables. Then $\phi_{tj} \equiv 0$ or $\gamma = 0$ and (8) reduces to a general linear model with heterogeneous variances. When the σ_t^2's are modelled as in (4), estimates of β and λ can be computed as in Verbyla (1993).

Example 2: *Variable-order and time-varying parameter autoregression.* A useful generalisation of (3) is

$$y_t = \sum_{j=1}^{p_t} \phi_{t,t-j} y_{t-j} + \varepsilon_t, \tag{9}$$

where $p_t \geqslant 0$ is an integer. Of course, for $p_t = t-1$ this reduces to (3) and for $p_t \equiv 0$ it is Example 1. Otherwise, (9) is called the variable-order antedependence model (Macchiavelli & Arnold, 1994; Gabriel, 1962). Finally, when $p_t \equiv p$, (9) represents the time-varying parameter autoregressive models in West et al. (1999). In this case, for $\phi_{tj} \equiv \phi_j$ and $\sigma_t^2 \equiv \sigma^2$, (9) reduces to the standard autoregressive models.

Example 3: *Stationary covariance matrix.* For Y stationary, its covariance matrix Σ is a Toeplitz matrix, i.e. all the entries of a given diagonal are constant. Consequently, Σ is structured and the ϕ_{tj}'s and $\log \sigma_t^2$'s are constrained and cannot necessarily be modelled as in (4) using covariates (Pourahmadi, 1999, § 2.6). Nevertheless, a great deal is known about these parameters since they appear as the coefficients of the Szegö polynomials orthogonal on the unit circle (Grenander & Szegö, 1958, Ch. 2). In this case, $\phi_{21}, \phi_{31}, \ldots, \phi_{n-1,1}$, the negatives of the entries of the first column of T, are precisely the first $n-1$ partial autocorrelation coefficients of Y. These are useful not only in ARMA model identification, but also as reflection coefficients in geophysics and as Schur parameters in digital signal processing and numerical linear algebra (Kailath, 1987).

Example 4. *Multinomial covariance matrix.* Let $p = (p_1, \ldots, p_n)'$ be such that $p_i > 0$, $\sum_{i=1}^n p_i = 1$, and define the diagonal matrix $P_l = \mathrm{diag}(p_1, \ldots, p_n)$. Then, for an integer N the covariance matrix of the multinomial $\mathrm{Mu}(N; p_1, \ldots, p_n)$ distribution, given by $N(P - pp')$, is singular of rank $n-1$. The nonredundant entries of the Cholesky decomposition of $P - pp'$ are given, for $t = 2, \ldots, n$ and $j = 1, \ldots, t-1$, by

$$\phi_{tj} = \phi_t = p_t / q_{t-1}, \quad \sigma_t^2 = p_t q_t / q_{t-1}, \tag{10}$$

where $q_t = 1 - \sum_{k=1}^t p_k$, $q_0 = 1$ and $q_n = 0$ (Tanabe & Sagae, 1992). In addition to the regression-based interpretation of the ϕ_{tj}'s and σ_t^2's from (3), these do have substantive meaning in the multinomial case. Consider a group of size N and let p_t be the probability of leaving the group at time t. Then q_t is the probability of staying beyond time t and $\phi_t = p_t / q_{t-1}$ is the conditional probability of a person of age $t-1$ leaving in the next year, i.e. the hazard rate. Furthermore, the tth diagonal entry of ND, namely

$$N\sigma_t^2 = N q_{t-1} \frac{p_t}{q_{t-1}} \left(1 - \frac{p_t}{q_{t-1}}\right) = N q_{t-1} \phi_t (1 - \phi_t),$$

is the expected value of the conditional variance of the number who leave at age t, given the number who are still there at time $t-1$. For other interpretations and applications of (10) to Pederson's method of generating multinomial random samples via a normal random generator and multivariate normal density approximations to multinomial distributions, see Tanabe & Sagae (1992).

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

Other concrete examples of X and Z in (8) suggested by two real datasets can be found in Pourahmadi (1999) and Pourahmadi & Daniels (2002).

2·3. *Dynamic representation of longitudinal data*

We introduce the analogue of (8) for random vectors of repeated measures on several subjects. Consider a clinical trial where m subjects enter the study. Let $Y_i = (y_{i1}, \ldots, y_{in_i})'$ stand for the vector of repeated measurements on subject i, let y_{it} be the tth response measured at not necessarily equally spaced times indexed by $t = 1, \ldots, n_i$ $(i = 1, \ldots, m)$ and let $x_{it} = (x_{1it}, \ldots, x_{pit})'$ be the covariates. To model the dependence among the repeated measurements, let $T_i \Sigma_i T_i' = D_i$ be the square-root-free Cholesky decomposition of Σ_i, the covariance matrix of the measurements on the ith subject. The nonredundant entries of T_i and D_i can be modelled as in (4) using subject-specific covariates with γ and λ as the common parameters for all Σ_i's.

There are two distinct ways of approaching the simultaneous modelling of the mean and covariance matrix in this situation depending on whether or not Z_i is a function of β. In the first method we simply add $\beta' x_{it}$ to the right-hand side of (3), keeping Z_i free of β for each subject. Then attaching the subscript i to the relevant terms in (8) leads to

$$Y_i = X_i \beta + Z_i \gamma + \varepsilon_i = (X_i, Z_i) \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \varepsilon_i, \tag{11}$$

a dynamic linear model for the data where $\mathrm{cov}(\varepsilon_i) = D_i = D_i(\lambda)$, for $i = 1, 2, \ldots, m$. In the second approach, replacing y_j by $y_j - \mu_j$ in (6) leads to dynamic nonlinear models in β and γ, since in this case Z_i does depend on β. The basic steps for likelihood-based inferences when the Y_i's are normally distributed are discussed in Pourahmadi (2000). The algorithm in Appendix 1 makes these steps more transparent for normal data and provides a viable method for estimation in the nonnormal and nonlinear cases, a topic which is currently under study by the authors.

3. Bayesian modelling of covariance matrices

3·1. *A new prior for a covariance matrix*

The most common prior for a covariance matrix is the inverse Wishart prior. However, this prior is very restrictive and lacks flexibility. Recently, several authors have proposed more flexible priors for a covariance matrix, but which lack conjugacy and can make computations difficult (Leonard & Hsu, 1992; Daniels & Kass, 1999; Barnard et al., 2000). Other priors appearing in the literature in the 1990's include a reference 'noninformative' prior (Yang & Berger, 1994) and a constrained Wishart prior (Everson & Morris, 2000). We show that factorising the covariance matrix as in (2) provides considerable flexibility in specifying prior beliefs and also offers conditional conjugacy for computations. This was recognised first by Bartlett who partitioned a covariance matrix into two blocks, the Bartlett decomposition, and was later used and extended by Dawid (1988), Richard & Steele (1988), Brown et al. (1994) and Roverato (2000).

We place a multivariate normal prior on the unconstrained nonredundant entries of T, denoted by ϕ from here on, and inverse gamma priors on the innovation variances, σ_t^2 $(t = 1, \ldots, n)$, the diagonal entries of D. This conjugacy facilitates computations using Gibbs sampling algorithms, by recognising that, conditional on ϕ, the full conditional distribution of the diagonal elements of D are inverse gamma and similarly, conditional

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

on D, the full conditional of ϕ is multivariate normal. We refer to this form of conjugacy as conditional conjugacy. Using the finest partition of the data and for a particular block diagonal structure on the covariance matrix for the multivariate normal prior on ϕ we show below that this simplifies to the generalised inverse Wishart prior proposed by Brown et al. (1994). In addition, for specific choices for the hyperparameters of the priors on ϕ and D, this simplifies to the inverse Wishart prior.

To be specific, our prior takes the following form:

$$\phi \sim N(\phi^*, \Sigma_\phi), \tag{12}$$

$$\sigma_t^2 \sim \text{IG}(\delta_t, a_t) \quad (t = 1, \ldots, n), \tag{13}$$

where ϕ and $\{\sigma_t^2\}$ are independent. The prior of Brown et al. (1994) is identical for σ_t^2, but imposes a block diagonal structure on the prior covariance matrix for ϕ:

$$\Sigma_\phi = \text{diag}(\sigma_n^2 \Sigma_n, \sigma_{n-1}^2 \Sigma_{n-1}, \ldots, \sigma_2^2 \Sigma_2), \tag{14}$$

where $\Sigma_t = \text{var}(\phi_t)/\sigma_t^2$ $(t = 2, \ldots, n)$ and $\phi_t = (\phi_{t,t-1}, \ldots, \phi_{t,1})'$. Under an inverse Wishart specification for Σ with parameters (δ, G), we have that

$$\sigma_1^2 \sim \text{IG}(\delta, g_{11}), \tag{15}$$

$$\sigma_t^2 \sim \text{IG}(\delta + 1, g_{tt.t-1}) \quad (t = 2, \ldots, n), \tag{16}$$

$$\phi_t \sim N(\phi_t^{**}, \Sigma_{\phi_t}^{**}) \quad (t = 2, \ldots, n), \tag{17}$$

where $g_{tt.t-1} = g_{tt} - g_t' G_t^{-1} g_t$ with g_{tt} the tth diagonal element of G, g_t the column vector composed of the first $t-1$ entries of the tth column of G, G_t the $(t-1) \times (t-1)$ leading principal minor of G, $\phi_t^{**} = G_t^{-1} g_t$ and $\Delta_{\phi_t}^{**} = \sigma_t^2 G_t^{-1}$. Our parameterisation provides a sensible scale for introducing or eliciting prior information as it focuses on the relationship between the 'current' y and the previous y's defined by the regression parameters ϕ, and how well we can do this prediction, as measured by the innovation variances σ_t^2; strategies for eliciting priors in the context of the generalised inverse Wishart prior can be found in Garthwaite & Al-Awadhi (2001). This prior also provides great flexibility in expressing prior belief through a plethora of hyperparameters and like the generalised inverse Wishart has the marginal invariance property discussed in Brown et al. (1994); when we obtain an additional observation, at time-point $n+1$, our prior can be extended to preserve the original specification of the marginal prior of the first n time-points. Our prior appears to offer many advantages relating to elicitation, positive definiteness and computations for both Gibbs sampling and introducing covariates; see § 3·3.

3·2. *Shrinking toward a specific structure*

Several authors have suggested placing priors on a covariance matrix so as to 'shrink' it toward some structure (Chen, 1979; Daniels & Kass, 1999). Such an approach provides robustuess to misspecification of the structure and offers stability over assuming no structure. The prior discussed in § 3·1 provides a natural formulation for such shrinkage and offers three distinct advantages over previous approaches, namely computational simplicity, the flexibility of many parameters to control shrinkage and sensible scale parameters for longitudinal data. The first advantage is lacking in the priors of Daniels & Kass (1999), which are based on decomposing a matrix into the eigenvalues and Givens angles or into the variance and correlations and shrinking these. Regarding the second, the approaches of Daniels & Kass (1999) offer two parameters for controlling the amount of

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

Table 1. *Risks from the first simulation for sample covariance matrix* (Un), (*T*,*D*), *rotation* (Rot) *and correlation* (Corr) *shrinkage estimators. See Appendix* 3 *for a description of the true covariance matrices of Scenarios* I–III, IIR_1, IIIR_1, IIR_2 *and* IIIR_2 *and* § 3·2 *for further details of the simulations.*

Scenario	m	Un	(T, D)	Rot	Corr	Scenario	m	Un	(T, D)	Rot	Corr
I	10	1·89	0·98	0·55	0·73	IIR_1	10	1·99	1·14	0·78	0·94
	20	0·84	0·48	0·24	0·32		20	0·82	0·58	0·38	0·46
	100	0·15	0·09	0·05	0·06		100	0·16	0·14	0·13	0·15
II	10	1·93	1·04	0·52	0·72	IIIR_1	10	1·86	1·86	1·65	1·76
	20	0·79	0·46	0·22	0·30		20	0·81	0·81	0·68	0·80
	100	0·15	0·10	0·05	0·06		100	0·16	0·16	0·14	0·15
III	10	1·83	1·00	0·81	0·71	IIR_2	10	1·80	1·14	0·63	0·84
	20	0·81	0·45	0·29	0·30		20	0·85	0·57	0·39	0·51
	100	0·14	0·09	0·06	0·06		100	0·14	0·13	0·11	0·13
						IIIR_2	10	1·90	1·67	1·69	1·65
							20	0·85	0·77	0·68	0·66
							100	0·15	0·15	0·13	0·11

shrinkage, and shrinking with an inverse Wishart prior only offers one (Chen, 1979), while our prior allows for at least two.

To assess the utility of shrinking under the (T, D)-parameterisation, we first replicated the simulations from Daniels & Kass (1999, § 3.2), see Appendix 3, and computed the risk using Stein's loss, $L(\hat{\Sigma}, \Sigma) = \mathrm{tr}(\hat{\Sigma}\Sigma^{-1}) - \log|\hat{\Sigma}\Sigma^{-1}| - p$ (Yang & Berger, 1994). The forms of true covariance matrices are given in Appendix 3 in terms of eigenvalues/angles and T/D, and the results appear in Table 1. We only report the two best shrinkage estimators, the rotation and correlation shrinkage estimators from Daniels & Kass (1999). To offer a fair comparison, in shrinking towards a diagonal structure we only shrink the ϕ and set $\Sigma_\phi = \tau^2 I$ in (12); we do not shrink the σ^2's. We can see that the shrinkage estimator using the (T, D)-parameterisation offers large gains over the sample covariance matrix when the matrix is not far from diagonal, but not as large as those offered by the rotation and correlation priors. When the matrix is very far from diagonal, as in Scenarios IIIR_1 and IIIR_2, similar small but uniformly positive gains are seen over the sample covariance matrix. However, we believe the computational advantages of the (T, D) shrinkage prior can offset the slight decrease in gains compared to the estimators proposed in Daniels & Kass (1999). In addition, the (T, D) shrinkage prior outperforms the estimators in Daniels & Kass (2001), which were proposed as computationally simpler alternatives to those in Daniels & Kass (1999), results not shown.

In the light of the promising results of the first simulation, we conducted another simulation in which we considered structures more suitable for longitudinal data, and compared estimators for the covariance matrix based on the structured, unstructured and shrinkage estimators; here, we shrink both the ϕ's and σ^2's again using $\Sigma_\phi = \tau^2 I$ and setting $\delta_t = \delta$ from (13). We consider the following three true matrices for the simulation.

Scenario IA. Correct structure: $\sigma_t^2 = 1$ $(t = 1, \ldots, 5)$, $\phi_{t,t-1} = 0{\cdot}75$, $\phi_{tj} = 0$ $(j < t-1)$.

Scenario IIA. Close to correct structure: $\sigma_t^2 = 1$ $(t = 1, \ldots, 5)$, $\phi_{t,t-1} = 0{\cdot}3 + 0{\cdot}1(t-1)$, $\phi_{t,t-2} = 0{\cdot}2$, $\phi_{t,t-3} = 0{\cdot}1$, $\phi_{t,t-4} = 0$.

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

Scenario IIIA. Far from the correct structure: $\sigma^2 = (0{\cdot}5, 0{\cdot}7, 1, 3, 5)$,

$$\phi_{t,t-1} = 0{\cdot}75 + 0{\cdot}02(t-1), \quad \phi_{t,t-2} = 0{\cdot}4, \quad \phi_{t,t-3} = 0{\cdot}2, \quad \phi_{t,t-4} = 0{\cdot}1.$$

For the shrinkage and structured estimators, we shrunk towards or assumed the first structure, respectively. We chose sample sizes of $m = 10, 20, 100$ to see how well the prior performs in both small, $m = 10$, and large, $m = 100$, sample sizes. The three true matrices were chosen to determine how the prior performs when shrinking toward the correct structure and when shrinking toward incorrect structures.

Table 2 shows the results for this simulation using the same loss function as in the first simulation. The shrinkage estimator is superior to the unstructured estimator for all three scenarios, including the robustness to misspecification for Scenario IIIA. The structured estimator did very well when the structure was correct, but was dominated by the unstructured and shrinkage when the structure is far from correct and by the shrinkage when the structure was close, except for the smallest sample size.

Table 2. *Risks from the second simulation for unstructured* (Un), *shrinkage* (Shrink) *and structured* (Struc) *estimators. See Appendix* 3 *for a description of the true covariance matrices of Scenarios* IA–IIIA, *and* § 3·2 *for further details of the simulation.*

Scenario	m	Un	Shrink	Struc
IA	10	1·90	0·58	0·21
	20	0·83	0·28	0·11
	100	0·15	0·06	0·02
IIA	10	1·90	0·69	0·54
	20	0·83	0·39	0·42
	100	0·15	0·10	0·34
IIIA	10	1·90	1·16	4·50
	20	0·83	0·58	4·23
	100	0·15	0·13	4·20

These simulations led us to a simpler form for the hyperparameters of the prior discussed in § 3·1 for longitudinal data. Use the (T, D)-prior, centred at a sensible structure,

$$\phi_{tj} \sim N(\phi^s_{tj}, \tau^2_{tj}), \tag{18}$$

$$\sigma^2_t \sim \text{IG}(\delta m^s_t, \delta), \tag{19}$$

where ϕ^s_{tj} and m^s_t are the unknown prior means for ϕ_{tj} and σ^2_t, respectively, under the hypothesised structure and τ^2_{tj} and δ are the unknown parameters which control the amount of shrinkage; for simplicity we can set $\tau^2_{tj} = \tau^2$ or set $\tau^2_{tj} = \tau^2_t$, that is a separate shrinkage parameter for each ϕ_t, or set $\tau^2_{tj} = \tau^2_{t-j}$, that is a separate shrinkage parameter for each lag. This prior offers many advantages and is a reasonable choice for longitudinal data for the following reasons: a wide array of structures exist for longitudinal data, many more than for general correlated data; the strategy is consistent with searching for and finding a parsimonious structure; and this prior offers stability in small samples, but still is consistent in large samples, as the simulations show. A sensible structure might be chosen by using the Bayesian information criterion or by examining the regressograms (Pourahmadi, 1999), so that the prior would be data-dependent. As a generic strategy, we

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

might shrink towards 'stationarity', shrinking ϕ_t to ϕ^*_{t-j} and σ_t^2 to σ^{2*}. In addition, to simplify computations further, the hyperparameters ϕ^s and σ^{2s} can be estimated by fitting the structured model directly. This would constitute an empirical Bayes approach.

3·3. *Covariance models with covariates*

The (T, D)-reparameterisation of Σ in (2) suggests a computationally simple, and easily interpretable, way of modelling the covariance matrix as a function of covariates with considerable flexibility. When ϕ is modelled as in (4), a normal prior on γ provides a conjugate prior for the regression coefficient, γ; this is analogous to modelling the mean of a normal distribution using covariates, by linear regression. Thus, as discussed in the context of the prior in § 3·2, modelling with covariates under this reparameterisation guarantees positive definiteness and provides a sensible metric for interpretation and conditional conjugacy for computations. Alternative approaches for introducing covariates include through the logarithm of the covariance matrix (Chiu et al., 1996) or through the variances and/or correlations (Verbyla, 1993; Barnard et al., 2000). The former lacks both interpretation and conjugacy while the latter is computer-intensive for moderate and large n.

To formulate the model with covariates, we start with (11); that is

$$Y_i = X_i\beta + Z_i(\beta)\gamma + \varepsilon_i, \tag{20}$$

with $\mathrm{cov}(\varepsilon_i) = D_i(\lambda) = \mathrm{diag}\{\exp(l'_{i,1}\lambda), \ldots, \exp(l'_{i,n}\lambda)\}$, for $i = 1, \ldots, m$. The design matrix for the dependence component, $Z_i(\beta)$, is constructed as in (6), using a subject-specific design vector $z_{i,tj}$ as a replacement for z_{tj} given in (4) and with y_j replaced by $y_{ij} - x_{ij}\beta$. This matrix can be used to specify a particular structure for the dependence, such as those forms given in § 2·2, and/or to include subject-specific covariate effects. Similarly, $l_{i,t}$ is the design vector for the innovation variances and can be chosen to represent a particular covariance structure and/or to model these variances as a function of subject-specific covariates. We specify normal priors for β, γ and λ, of which the priors for β and γ are conjugate:

$$\beta \sim N(\beta^*, \Sigma_\beta), \quad \gamma \sim N(\gamma^*, \Sigma_\gamma), \quad \lambda \sim N(\lambda^*, \Sigma_\lambda). \tag{21}$$

The hyperparameters in (21) can be chosen based on prior information or to be 'non-informative'. We note that the models in this section and in § 3·2 could be combined, for example by having the structures shrunk towards those in § 3·2 depend on subject-specific covariates.

So far as computation is concerned, the models in (20) and (21) and/or models using the priors in §§ 3·1 and 3·2 are well suited to Gibbs sampling as most of the full conditional distributions take known forms. We discuss some important issues in the Gibbs algorithm here, but leave the details to Appendix 2. In addition, we focus on the case where Z_i is a function of β, as in the dynamic nonlinear model. Otherwise, the X_i and Z_i, and β and γ, can be combined as on the right-hand side of (11).

First, considering the case with all hyperparameters known, we need to construct the full conditional distributions for β, γ and λ. The first two are normally distributed. The third can be sampled by using a Metropolis–Hastings algorithm. For the models and priors in § 3·2, we need to sample from the full conditional distribution of σ_t^2; this will follow an inverse gamma distribution. For the unknown hyperparameters, we use Metropolis–Hastings algorithms.

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

4. DISCUSSION

An attractive feature of the priors in § 3 is their marginal invariance property. Further research on recursive computation of the posterior quantities using this property is of interest in predicting a future response and could bring valuable theoretical insights as well as computational saving to Bayesian analysis of covariance matrices for longitudinal data.

Model (3) can be generalised to a dynamic mixed model that includes random coefficient autoregressive models for longitudinal data (Rahiala, 1999). Such models for normal data are introduced in Pourahmadi & Daniels (2002) and are applied to an actual dataset. We are extending these dynamic mixed models to nonnormal data through dynamic generalised linear mixed models. For estimation of such models the iteratively reweighted least squares algorithm presented in Appendix 1 is expected to play a key role.

ACKNOWLEDGEMENT

M. J. Daniels was supported in part by a grant from the National Institutes of Health, and M. Pourahmadi by a grant from the Graduate School of Northern Illinois University.

APPENDIX 1

Iteratively reweighted least squares

We develop a method for estimating the parameters β, γ and λ in (11) without assuming normality for the response Y. If we set $A_i = (X_i, Z_i)$ and $\theta_i = (\beta', \gamma')'$, the vector of combined mean and dependence parameters, then (11) can be viewed as a general linear model with a diagonal covariance matrix parameterised by λ. For the time being assume that the Z_i's do not depend on β. Then, the following iterative procedure is natural and much simpler than Pourahmadi's (2000) algorithm.

Step 1. Select an initial value $\tilde{\theta}$ for θ.

Step 2. Compute $\tilde{\varepsilon}_i = Y_i - A_i\tilde{\theta}$ $(i = 1, \ldots, m)$ to estimate the variance parameter $\tilde{\lambda}$ of D_i using the algorithm in Verbyla (1993).

Step 3. Update $\tilde{\theta}$ using $\tilde{D}_i = D_i(\tilde{\lambda})$ and

$$\hat{\theta} = \left(\sum_{i=1}^{m} A_i' \tilde{D}_i^{-1} A_i \right)^{-1} \sum_{i=1}^{m} A_i' \tilde{D}_i^{-1} Y_i.$$

Step 4. Stop the process if $\hat{\theta} \doteq \tilde{\theta}$ and take $\hat{\theta}$ and $\tilde{\lambda}$ as the estimates of θ and λ.

The ordinary least-squares estimate of θ provides a simple initial value. Note that updating θ is less complicated than computing the weighted least-squares estimates of regression parameters since the $\tilde{D}_i$'s are diagonal matrices. However, Step 2 is the most computationally intensive part since it relies on an iterative Newton–Raphson algorithm as an inner loop.

The above method for estimating the parameters β, γ and λ in (11) is not applicable when Z_i depends on β. To deal with this, we suggest the following simple modification of Step 3.

Step 3′. Update $\tilde{\theta} = (\tilde{\beta}', \tilde{\gamma}')'$ using

$$\hat{\beta} = \left(\sum_{i=1}^{m} X_i' \tilde{\Sigma}_i^{-1} X_i \right)^{-1} \sum_{i=1}^{m} X_i' \tilde{\Sigma}_i^{-1} Y_i,$$

$$\hat{\gamma} = \left\{ \sum_{i=1}^{m} Z_i'(\hat{\beta}) \tilde{D}_i^{-1} Z_i(\hat{\beta}) \right\}^{-1} \sum_{i=1}^{m} Z_i'(\hat{\beta}) \tilde{D}_i^{-1} (Y_i - X_i\hat{\beta}),$$

where $\tilde{\Sigma}_i = T_i(\tilde{\gamma}) D_i(\tilde{\lambda}) T_i'(\tilde{\gamma})$.

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

Initial values can be obtained by computing $\hat{\beta} = (\sum_{i=1}^{m} X_i' X_i)^{-1} \sum_{i=1}^{m} X_i' Y_i$ and then using $Z_i(\hat{\beta})$ and the earlier algorithm to obtain $\tilde{\theta}$ as an initial value.

To obtain standard errors for $\hat{\theta}$, we recommend either ignoring the dependence of $Z_i(\beta)$ on β and using $\text{var}(\hat{\theta}) = \sum_{i=1}^{m} A_i'(\hat{\beta}) \tilde{D}_i^{-1} A_i(\hat{\beta})$ or using the inverse of the information matrix in Pourahmadi (2000).

Appendix 2

Full conditional distributions for models specified in § 3

The full conditional for β. We exploit the connection between $\phi(\gamma)$ and the covariance matrix for the Y_i discussed in § 2. Recall that we can re-write equation (11) as

$$Y_i \sim N(X_i \beta, \Sigma_i),$$

where $\Sigma_i = \Sigma_i(\gamma, \lambda)$. Under this formulation, the full conditional for β is normal with covariance matrix and mean

$$\Omega_\beta = \left\{ \Sigma_\beta^{-1} + \left(\sum X_i \Sigma_i^{-1} X_i \right) \right\}^{-1}, \quad \mu_\beta = \Omega_\beta \left(\Sigma_\beta^{-1} \beta^* + \sum X_i \Sigma_i^{-1} Y_i \right),$$

The full conditional for γ. This full conditional will be normal with covariance matrix and mean

$$\Omega_\gamma = \left[\Sigma_\gamma^{-1} + \left\{ \sum_i Z_i(\beta) D_i^{-1} Z_i(\beta) \right\} \right]^{-1}, \quad \mu_\gamma = \Omega_\gamma \left\{ \Sigma_\gamma^{-1} \gamma^* + \sum Z_i(\beta) D_i^{-1} (Y_i - X_i \beta) \right\},$$

respectively.

The full conditional for λ. As this full conditional takes a nonstandard form, we use a normal approximation to the full conditional distribution as a candidate distribution for a Metropolis–Hastings algorithm. To compute the candidate distribution, we use Newton–Raphson to find the mode of the full conditional distribution using the results in Verbyla (1993). Using the inverse of the negative of the Hessian matrix evaluated at the mode as our estimate of the covariance matrix, we form a normal approximation to the full conditional, denoted by $\hat{p}(.)$. The acceptance probability at the kth iteration of the Metropolis–Hastings algorithm is $\alpha = \min(1, \alpha_0)$, where

$$\alpha_0 = \frac{p(\lambda^{(k)} | \beta, \gamma, y) / \hat{p}(\lambda^{(k)} | \beta, \gamma, y)}{p(\lambda^{(k-1)} | \beta, \gamma, y) / \hat{p}(\lambda^{(k-1)} | \beta, \gamma, y)}.$$

Missing data. To 'impute' missing data, we use data augmentation (Tanner & Wong, 1987). If we define $Y_{\text{mis},i}$ to be the missing data and $Y_{\text{obs},i}$ to be the observed data on subject i, the conditional distribution of $Y_{\text{mis},i} | Y_{\text{obs},i}, \beta, \Sigma_i$, will be multivariate normal. We impute missing data here under a missing at random assumption (Little & Rubin, 1976). For further details, see Pourahmadi & Daniels (2002).

The full conditional for σ_t^2, *for models using the prior in* § 3·2. The full conditional for $(\sigma_t^2)^{-1}$ will be $\text{Ga}(n/2 + \delta m_t^s, [\delta + \frac{1}{2} \sum_i \{ y_{it} - X_{it} \beta - Z_i(t; \beta) \gamma \}^2])$, where here $\gamma = \phi$.

The full conditional for the hyperparameters of the priors for σ_t^2 *and* ϕ *in* § 3·2. Uniform shrinkage priors can be specified for the shrinkage parameters, τ_{tj}^2 and δ (Daniels, 1999); priors for the other hyperparameters will depend on the structure to which one is shrinking. Metropolis–Hastings algorithms can be used to sample from all these hyperparameters using either a candidate distribution constructed as in generating λ above or a random walk candidate distribution.

Appendix 3

Details for the first simulation

We present the true covariance matrices used in the first simulation. The first three are diagonal matrices and the last four are rotations of Matrices II and III, with the same eigenvalues. The

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

rotation matrices are uniquely defined by the Givens angles; see Daniels & Kass (1999) for more details.

Matrix I: diag(1, 1, 1, 1, 1).
Matrix II: diag$\{1, 0{\cdot}75, (0{\cdot}75)^2, (0{\cdot}75)^3, (0{\cdot}75)^4\} = \mathrm{diag}(1, 0{\cdot}75, 0{\cdot}56, 0{\cdot}42, 0{\cdot}32)$.
Matrix III: diag$\{1, 0{\cdot}75, (0{\cdot}75)^2, (0{\cdot}75)^{10}, (0{\cdot}75)^{20}\} = \mathrm{diag}(1, 0{\cdot}75, 0{\cdot}56, 0{\cdot}06, 0{\cdot}003)$.
Matrix IIR_1: Matrix II with Givens angles all set to $\pi/4$.
Matrix IIIR_1: Matrix III with Givens angles all set to $\pi/4$.
Matrix IIR_2: Matrix II with Givens angles evenly spaced between $(-\pi/4, \pi/4)$.
Matrix IIIR_2: Matrix III with Givens angles evenly spaced between $(-\pi/4, \pi/4)$.

We now give the forms of the T and D matrices for the nondiagonal matrices, corresponding to Matrices IIR_1, IIIR_1, IIR_2 and IIIR_2.

Matrix IIR_1:

$$T = \begin{pmatrix} 1{\cdot}00 & & & & \\ -0{\cdot}36 & 1{\cdot}00 & & & \\ -0{\cdot}14 & 0{\cdot}06 & 1{\cdot}00 & & \\ -0{\cdot}04 & 0{\cdot}07 & 0{\cdot}22 & 1{\cdot}00 & \\ -0{\cdot}21 & -0{\cdot}11 & 0{\cdot}01 & 0{\cdot}35 & 1{\cdot}00 \end{pmatrix},$$

Matrix IIIR_1:

$$T = \begin{pmatrix} 1{\cdot}00 & & & & \\ -1{\cdot}73 & 1{\cdot}00 & & & \\ -1{\cdot}05 & -0{\cdot}37 & 1{\cdot}00 & & \\ -0{\cdot}82 & -0{\cdot}27 & 0{\cdot}44 & 1{\cdot}00 & \\ -2{\cdot}89 & -1{\cdot}39 & -0{\cdot}78 & 1{\cdot}15 & 1{\cdot}00 \end{pmatrix},$$

Matrix IIR_2:

$$T = \begin{pmatrix} 1{\cdot}00 & & & & \\ -1{\cdot}04 & 1{\cdot}00 & & & \\ 0{\cdot}22 & 0{\cdot}22 & 1{\cdot}00 & & \\ 0{\cdot}22 & 0{\cdot}18 & 0{\cdot}03 & 1{\cdot}00 & \\ 0{\cdot}25 & 0{\cdot}11 & -0{\cdot}11 & 0{\cdot}04 & 1{\cdot}00 \end{pmatrix},$$

Matrix IIIR_2:

$$T = \begin{pmatrix} 1{\cdot}00 & & & & \\ -0{\cdot}16 & 1{\cdot}00 & & & \\ 0{\cdot}41 & 0{\cdot}34 & 1{\cdot}00 & & \\ 0{\cdot}51 & 0{\cdot}43 & -0{\cdot}17 & 1{\cdot}0 & \\ 0{\cdot}97 & 0{\cdot}63 & -0{\cdot}66 & -0{\cdot}55 & 1{\cdot}00 \end{pmatrix},$$

Matrix IIR_1: $D = \mathrm{diag}(0{\cdot}42, 0{\cdot}61, 0{\cdot}58, 0{\cdot}53, 0{\cdot}71)$,
Matrix IIR_2: $D = \mathrm{diag}(0{\cdot}68, 0{\cdot}73, 0{\cdot}57, 0{\cdot}45, 0{\cdot}44)$,

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

Matrix IIIR_1: $D = \text{diag}(0{\cdot}15, 0{\cdot}15, 0{\cdot}48, 0{\cdot}16, 0{\cdot}04)$,
Matrix IIIR_2: $D = \text{diag}(0{\cdot}58, 0{\cdot}66, 0{\cdot}43, 0{\cdot}05, 0{\cdot}01)$.

References

Anderson, T. W. (1978). Repeated measurements on autoregressive processes. *J. Am. Statist. Assoc.* **73**, 371–8.

Barnard, J., McCulloch, R. & Meng, X. (2000). A natural strategy for modeling covariance matrices with application to shrinkage. *Statist. Sinica* **10**, 1281–311.

Brown, P. J., Le, N. D. & Zidek, J. V. (1994). Inference for a covariance matrix. In *Aspects of Uncertainty*, Ed. P. R. Freeman and A. F. M. Smith, pp. 77–92. Chichester: John Wiley.

Chen, C. F. (1979). Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis *J. R. Statist. Soc.* B **41**, 235–48.

Chin, T. Y. M., Leonard, T. & Tsui, K.-W. (1996). The matrix-logarithmic covariance model. *J. Am. Statist. Assoc.* **91**, 198–210.

Daniels, M. J. (1999). A prior for the variance in hierarchical models. *Can. J. Statist.* **27**, 569–80.

Daniels, M. J. & Kass, R. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J. Am. Statist. Assoc.* **94**, 1254–63.

Daniels, M. J. & Kass, R. (2001). Shrinkage estimators for a covariance matrix. *Biometrics* **57**, 1173–84.

Dawid, A. P. (1988). The infinite regress and its conjugate analysis. In *Bayesian Statistics 3*, Ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, pp. 95–110. Oxford: Clarendon Press.

Diggle, P. J., Liang, K. Y. & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

Everson, P. & Morris, C. (2000). Inference for multivariate normal hierarchical models. *J. R. Statist. Soc.* B **62**, 399–412.

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

Garthwaite, P. H. & Al-Awadhi, S. A. (2001). Nonconjugate prior distribution assessment for multivariate normal sampling. *J. R. Statist. Soc.* B **63**, 95–110.

Grenander, U. & Szegö, G. (1958). *Toeplitz Forms and their Applications*. Berkeley, CA: University of California Press.

Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach*. London: Chapman and Hall.

Kailath, T. (1987). Signal processing applications of some moment problems. In *Proc. Symposia in Applied Mathematics*. Ed. H. J. Landau, pp. 71–109. Providence, RI: Am. Math. Soc.

Laird, N. M. & Ware, J. J. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 973–9.

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

Little, R. J. A. & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

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

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**, 677–90.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87**, 425–35.

Pourahmadi, M. & Daniels, M. (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics*, **58**, 82–8.

Rahiala, M. (1999). Random coefficient autoregressive models for longitudinal data. *Biometrika* **86**, 718–22.

Richard, J. F. & Steel, M. F. J. (1988). Bayesian analysis of systems of seemingly unrelated regression equations under a recursive extended natural conjugate prior density. *J. Economet.* **38**, 7–37.

Roverato, A. (2000). Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika* **87**, 99–112.

SAS Institute Inc. (1997). *SAS/STAT Software: Changes and Enhancements through Release 6.12*. Cary, NC: SAS Institute Inc.

Searle, S., Casella, G. & McCulloch, C. (1992). *Variance Components*. New York: John Wiley.

Spiegelhalter, D. J., Best, N. G., Gilks, W. R. & Inskip, H. (1996). Hepatitis B: A case study in MCMC methods. In *Markov Chain Monte Carlo in Practice*, Ed. W. R. Gilks, S. Richardson and D. J. Spiegelhalter, pp. 339–58. London: Chapman and Hall.

Tanabe, K. & Sagae, M. (1992). An exact Cholesky decomposition and the generalized inverse of the variance-covariance matrix of the multinomial distribution, with applications. *J. R. Statist. Soc.* B **54**, 211–9.

Tanner, M. A. & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Statist. Assoc.* **82**, 528–40.

Tsimikas, J. V. & Ledolter, J. (1998). Analysis of multi-unit variance components models with state space profiles. *Ann. Inst. Statist. Math.* **50**, 147–64.

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013

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

WEST, M. & HARRISON, P. J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer-Verlag.

WEST, M., PRADO, R. & KRYSTAL, A. D. (1999). Evaluation and comparison of EEG traces: Latent structure in nonstationary time series. *J. Am. Statist. Assoc.* **94**, 1083–95.

YANG, R. & BERGER, J. O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22**, 1195–211.

[*Received January* 2001. *Revised December* 2001]

Downloaded from http://biomet.oxfordjournals.org/ at UNIVERSIDAD DEL PAIS VASCO on April 4, 2013