ELSEVIER

Journal of Statistical Planning and Inference 76 (1999) 41–55

journal of statistical planning and inference

On an unbalanced growth curve model with random effects and AR(1) errors from a Bayesian and the ML points of view

Jack C. Lee *, Wei-Fang Niu

Institute of Statistics, National Chiao Tung University, 1001 TA Hsuch Road, Hsinchu, Taiwan

Received 2 October 1997; received in revised form 1 July 1998

Abstract

In this paper we consider a Bayesian analysis of unbalanced (general) growth curve model with random effects and AR(1) errors. Three priors are proposed and put into comparisons in parameter estimation and prediction problems.

1. Introduction

In this paper we mainly devote ourselves to the growth curve model with random effects and AR(1) errors via a Bayesian approach. The model considered here is:

$$\underset{p_{ij}\times 1}{Y_{ij}} = \underset{p_{ij}\times m_i}{X_{ij}}\ \underset{m_i\times 1}{\tau_i} + \underset{p_{ij}\times w_i}{D_{ij}}\ \underset{w_i\times 1}{u_{ij}} + \underset{p_{ij}\times 1}{\varepsilon_{ij}}, \quad j=1,2,\ldots,N_i;\ i=1,2,\ldots,r, \tag{1.1}$$

where Y_{ij} is the measurement and with unequal lengths, τ_i is an unknown vector of regression coefficients of group i, X_{ij} and D_{ij} are known design matrices, u_{ij} is the random effects error and possesses w_i-variate normal distribution with mean vector 0 and covariance matrix $\sigma^2\Gamma_i$, and the disturbance terms ε_{ij}'s are independent p_{ij}-variate normal with mean vector 0 and AR(1) covariance matrix $\sigma^2 C_{ij}$, where $C_{ij}=(\rho^{|a-b|})$, $a,b=1,\ldots,p_{ij}$.

The covariance matrix of Y_{ij} can be written as

$$\begin{aligned}\Sigma_{ij} &= \sigma^2(D_{ij}\Gamma_i D_{ij}' + C_{ij})\\ &\equiv \sigma^2\Lambda_{ij}.\end{aligned} \tag{1.2}$$

It is noted that the random effects are characterized by D_{ij}. For example, D_{ij} equals to $1_{p_{ij}}$ for the random intercept model and equals to X_{ij} when both intercept and slope are random.

* Corresponding author.

0378-3758/99/$ – see front matter
PII: S0378-3758(98)00161-X

The growth curve model was first proposed by Potthoff and Roy (1964) and subsequently considered by many authors, including Rao (1987), Khatri (1966), Geisser (1970), Lee and Geisser (1972), Lee (1988, 1991), Lee and Hsu (1997), among others. Laird and Ware (1982) considered the random effects model with white noise errors and obtained maximum likelihood (ML) and restricted maximum likelihood (REML) estimates of variance–covariance components via EM algorithm. Jennrich and Schluchter (1986) discussed various types of covariance structures, including random effects models and the AR(1) model separately. In Lee (1988, 1991) and Keramidas and Lee (1990) the importance of the AR(1) dependence was demonstrated using some real data in prediction of future observations. Chi and Reinsel (1989) considered the ML estimates for the model with both random effects and AR(1) errors by the scoring method. A Bayesian analysis of the AR(1) dependence was considered by Lee and Hsu (1997).

Three types of prior for Bayesian analysis of the model (1.1) are proposed in this paper. In addition to parameter estimation, we also consider two types of prediction problem which are practically useful. In Section 2, the motivations of the priors will be discussed in detail. In Sections 3 and 4, Bayesian methods for parameter estimation and prediction of future values are developed. Some approximate methods are suggested since the exact forms of the posterior distribution functions of a few parameters are not easy to obtain. Some numerical results with real-data analyzed by the methods developed in Sections 3 and 4 are illustrated in Section 5. Simulations are also done to clarify some properties of the proposed methods. Finally, in Section 6 some concluding remarks are given.

2. Choices of prior

For convenience let $\tau=(\tau_1,\ldots,\tau_r)$ and $\Gamma=(\Gamma_1,\ldots,\Gamma_r)$. For model (1.1), the likelihood function of τ, σ^2, Γ and ρ is

$$L(\tau,\sigma^2,\Gamma,\rho|Y)\propto(\sigma^2)^{-(1/2)n}\prod_{i=1}^{r}\prod_{j=1}^{N_i}|\Lambda_{ij}|^{-1/2}\,\mathrm{e}^{-1/2\sigma^2(Y_{ij}-X_{ij}\tau_i)'\Lambda_{ij}^{-1}(Y_{ij}-X_{ij}\tau_i)}, \qquad (2.1)$$

where $n=\sum_{i=1}^{r}\sum_{j=1}^{N_i}p_{ij}$.

To choose appropriate prior densities, we will assume τ,σ^2,Γ and ρ are independent a priori. We next present the three different priors considered for this model.

2.1. Prior 1

This joint prior is noninformative in nature. The noninformative prior distributions for τ and ρ are easily seen to be proportional to constants since all of the elements of τ are just real numbers and ρ is confined within $(-1,1)$. For σ^2, we choose σ^{-2} as its prior because $\sigma^2>0$. For Γ, we may consider each of its elements as a ratio of the variance–covariance components of the covariance matrix of u_{ij} to the main variation σ^2, so every element of Γ should be within a finite interval such that neither random effects nor the AR(1) errors dominate the other. Utilizing the 'principle of

stable estimation' suggested by Edwards et al. (1963), uniform prior is appropriate for Γ. Therefore, coupling with the spirits of Zellner and Tiao (1964), the joint prior for τ, σ^2, Γ and ρ can be written as

$$\pi(\tau,\sigma^2,\Gamma,\rho)\propto\pi(\tau)\pi(\sigma^2)\pi(\Gamma)\pi(\rho)\propto\sigma^{-2}. \tag{2.2}$$

It is worth noting that the usual noninformative prior for Γ_i, $|\Gamma_i|^{-(w_i+1)/2}$, is not appropriate here since it induces nonintegrable posterior densities. An explanation for this is that such a prior just down-weighs one of the two components (random effects) of the covariance. However, the properties of being strictly positive definite of Σ_{ij} is guaranteed by $\sigma^2 C_{ij}$ so there's no mechanism in the likelihood to cease the intendency for Γ_i to decrease. Finally $|\Gamma_i|$ approaches 0 and thus reduces the model simply to the AR(1) model. This also could lead to the question whether it is reasonable to assume Γ and ρ to have independent distributions, since both of them are used to explain the 'variations' of the data whereas the total 'variations' of the data is just fixed!

2.2. Prior 2

Following Box and Tiao (1973), assume $r=1$, $w_i=w$ and omitting subscripts for C_{ij} and D_{ij}, we may also choose the prior as

$$\pi(\tau,\sigma^2,\Gamma,\rho)\propto\sigma^{-2}|\Gamma+(D'C^{-1}D)^{-1}|^{-(w+1)/2}. \tag{2.3}$$

2.3. Prior 3

Another approach to determine the prior is to introduce an informative prior such as the inverse Wishart distribution, $\mathrm{IW}(\Omega,\nu)$, for Γ. The parameters Ω and ν may be roughly estimated from the data. We will set ν as small as possible, i.e., $m+2$. The Ω could be set as diagonal with diagonal elements being the sample variances of the corresponding regression coefficients when each individual regresses on the design matrix X_{ij} with white noise errors.

In the following Bayesian inference, we shall denote the prior of Γ and ρ as $\pi(\Gamma,\rho)$ without specifying which prior being used in the theoretical development. However, three different priors will be utilized for $\pi(\Gamma,\rho)$ in numerical illustrations.

3. Parameter estimation

Combining Eq. (2.1) with the prior and integrating w.r.t. $\sigma^2,\tau_1,\tau_2,\ldots,\tau_r$, we get the following joint posterior of Γ and ρ:

$$P(\Gamma,\rho|Y)\propto\pi(\Gamma,\rho)\left[\prod_{i=1}^{r}\prod_{j=1}^{N_i}|\Lambda_{ij}|^{-1/2}\right]B^{-1/2(n-\sum_{i=1}^{r}m_i)}\times\prod_{i=1}^{r}\left|\sum_{j=1}^{N_i}X_{ij}'\Lambda_{ij}^{-1}X_{ij}\right|^{-1/2}, \tag{3.1}$$

where

$$B = \sum_{i=1}^{r} \sum_{j=1}^{N_i} (Y_{ij} - X_{ij}\hat{\tau}_i)' \Lambda_{ij}^{-1} (Y_{ij} - X_{ij}\hat{\tau}_i),$$

$$\hat{\tau}_i = \left(\sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} X_{ij} \right)^{-1} \sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} Y_{ij}. \tag{3.2}$$

The posterior density of τ_i can be approximated by

$$P(\tau_i|Y) \doteq P(\tau_i|\hat{\Gamma}, \hat{\rho}, , Y), \tag{3.3}$$

where $(\hat{\Gamma}, \hat{\rho})$ is the mode of $P(\Gamma, \rho|Y)$, if $P(\Gamma, \rho|Y)$ is concentrated and nearly symmetric, as pointed out by Ljung and Box (1980).

Thus, we have the following approximate posterior distribution for τ_i:

$$\tau_i|Y \dot{\sim} T_{m_i} \left(\hat{\tau}_i^*, \hat{B} \left(\left(n - \sum_{j=1}^{r} m_j \right) \sum_{j=1}^{N_i} X_{ij}' \hat{\Lambda}_{ij}^{-1} X_{ij} \right)^{-1}, \; n - \sum_{j=1}^{r} m_j \right), \tag{3.4}$$

where $\hat{\tau}_i^*, \hat{B}, \hat{\Lambda}_{ij}$ and $\hat{C}_{ij}$ are the $\hat{\tau}_i, B, \Lambda_{ij}$ and C_{ij}, defined earlier, with ρ and Γ replaced by $\hat{\rho}$ and $\hat{\Gamma}$, respectively, and $T_p(\mu, \Sigma, v)$ has the density

$$f(y) = K(v, p)|\Sigma|^{-1/2}[1 + v^{-1}(y - \mu)'\Sigma^{-1}(y - \mu)]^{-(p+v)/2}, \tag{3.5}$$

$K(v, p) = \Gamma((v + p)/2)/\Gamma(v/2)(\pi v)^{p/2}$, and $\hat{\rho}$ and $\hat{\Gamma}$ maximized $P(\Gamma, \rho|Y)$, as given in Eq. (3.1).

Hence, an approximate posterior region for τ_i can be obtained from

$$(\tau_i - \hat{\tau}_i^*)' \left(\sum_{j=1}^{N_i} X_{ij}' \hat{\Lambda}_{ij}^{-1} X_{ij} \right) (\tau_i - \hat{\tau}_i^*) \leqslant \frac{m_i \hat{B}}{n - \sum_{i=1}^{r} m_i} F_\alpha \left(m_i, n - \sum_{i=1}^{r} m_i \right), \tag{3.6}$$

where $F_\alpha(v_1, v_2)$ is the upper $100\alpha\%$ point of the F distribution with v_1 and v_2 degrees of freedom.

Also,

$$\sigma^2|Y \dot{\sim} \text{IG} \left(\frac{n - \sum_{i=1}^{r} m_i}{2}, \frac{\hat{B}}{2} \right). \tag{3.7}$$

4. Prediction of partially observed future values

In this section we consider the prediction of y_{lk}, a future q-dimensional values of measurement Y_{lk}. To accomplish this task, the covariance structure must be extendable to the future values of all observed individuals. The covariance structure considered in

this paper satisfies this requirement. We will first state the following lemma which is useful in this section.

Lemma 4.1. *Let $X^*_{p\times q}$ and $Z^*_{p\times(p-q)}$ be matrices of ranks $q<p$ and $(p-q)$, respectively, such that $X^{*\prime}Z^*=0$. If $\Lambda^*_{p\times p}$ is a positive matrix, then*

$$\Lambda^{*^{-1}}=\Lambda^{*^{-1}}X^*(X^{*\prime}\Lambda^{*^{-1}}X^*)^{-1}X^{*\prime}\Lambda^{*^{-1}}+Z^*(Z^{*\prime}\Lambda^*Z^*)^{-1}Z^{*\prime}.$$

Let x and d be the $q\times m_l$ and $q\times w_l$ design matrices corresponding to y_{lk}. Then we have

$$E\begin{pmatrix}Y_{lk}\\ y_{lk}\end{pmatrix}=\begin{pmatrix}X_{lk}\\ x\end{pmatrix}\tau_l,$$

$$\mathrm{Cov}\begin{pmatrix}Y_{lk}\\ y_{lk}\end{pmatrix}=\sigma^2(D^*\Gamma_lD^{*\prime}+C^*)=\sigma^2\Lambda^*=\sigma^2\begin{bmatrix}\Lambda^*_{11} & \Lambda^*_{12}\\ \Lambda^*_{21} & \Lambda^*_{22}\end{bmatrix},$$

where

$$D^*=\begin{pmatrix}D_{lk}\\ d\end{pmatrix},\quad C^*=(\rho^{|a-b|}),\ a,b=1,\ldots,p_{lk}+q.$$

Let

$$Y^*_{lk}=\begin{pmatrix}Y_{lk}\\ y_{lk}\end{pmatrix},\quad X^*=\begin{pmatrix}X_{lk}\\ x\end{pmatrix}.$$

Combining the conditional density function of y_{lk} given Y_{lk},τ_l, σ^2, Γ_l and ρ with the posterior density of τ, σ^2, Γ and ρ, and integrating w.r.t. σ^2 and Γ, applying Lemma 4.1 and after some algebraic manipulations we have the following joint posterior density of y_{lk}, Γ and ρ:

$$P(y_{lk},\Gamma,\rho|Y)\propto\pi(\Gamma,\rho)\left[\prod_{i=1}^{r}\prod_{j=1}^{N_i}|\Lambda_{ij}|^{-1/2}\right]|\Lambda^*_{22.1}|^{-1/2}\prod_{i\neq l}\left|\sum_{j=1}^{N_i}X'_{ij}\Lambda_{ij}^{-1}X_{ij}\right|^{-1/2}$$
$$\times|Q|^{-1/2}[B_1+B_2+B_3+(y_{lk}-\mu)'G_{22}(y_{lk}-\mu)]^{-1/2(n+q-\sum_{i=1}^{r}m_i)},\tag{4.1}$$

where

$$G=\Lambda^{*^{-1}}X^*Q_2^{-1}Q_1Q^{-1}Q_2Q_2^{-1}X^{*\prime}\Lambda^{*^{-1}}+Z^*(Z^{*\prime}\Lambda^*Z^*)^{-1}Z^{*\prime}$$
$$=\begin{pmatrix}G_{11} & G_{12}\\ G_{21} & G_{22}\end{pmatrix},$$

$$G_{11.2}=G_{11}-G_{12}G_{22}^{-1}G_{21},$$

$$\mu=x\hat{\tau}_{l1}-G_{22}^{-1}G_{21}(Y_{lk}-X_{lk}\hat{\tau}_{l1}),$$

$$Q = Q_1 + Q_2,$$

$$Q_1 = \sum_{j\neq k} X'_{lj}\Lambda_{lj}^{-1}X_{lj},$$

$$Q_2 = X^{*\prime}\Lambda^{*^{-1}}X^*,$$

$$B_1 = \sum_{i\neq l}\sum_{j=1}^{N_i}(Y_{ij} - X_{ij}\hat{\tau}_i)'\Lambda_{ij}^{-1}(Y_{ij} - X_{ij}\hat{\tau}_i),$$

$$B_2 = \sum_{j\neq k}(Y_{lj} - X_{lj}\hat{\tau}_{l1})'\Lambda_{lj}^{-1}(Y_{lj} - X_{lj}\hat{\tau}_{l1}),$$

$$B_3 = (Y_{lk} - X\hat{\tau}_{l1})'G_{11.2}(Y_{lk} - X\hat{\tau}_{l1}),$$

$$\hat{\tau}_i = \left(\sum_{j=1}^{N_i} X'_{ij}\Lambda_{ij}^{-1}X_{ij}\right)^{-1}\sum_{j=1}^{N_i} X'_{ij}\Lambda_{ij}^{-1}Y_{ij}, \quad i \neq j,$$

$$\hat{\tau}_{l1} = \left(\sum_{j\neq k} X'_{lj}\Lambda_{lj}^{-1}X_{lj}\right)^{-1}\sum_{j\neq k} X'_{lj}\Lambda_{lj}^{-1}Y_{lj},$$

$$\hat{\tau}_{l2} = (X^{*\prime}\Lambda^{*^{-1}}X^*)^{-1}X^{*\prime}\Lambda^{*^{-1}}Y^*_{lk},$$

$$\hat{\tau}_l = Q^{-1}(Q_1\hat{\tau}_{l1} + Q_2\hat{\tau}_{l2}).$$

Integrating w.r.t. y_{lk} we have the following posterior density of Γ and ρ:

$$P(\Gamma,\rho|Y) \propto \pi(\Gamma,\rho)\left[\prod_{i=1}^{r}\prod_{j=1}^{N_i}|\Lambda_{ij}|^{-1/2}\right]|\Lambda^*_{22.1}|^{-1/2} \times \prod_{i\neq l}\left|\sum_{j=1}^{N_i}X'_{ij}\Lambda_{ij}^{-1}X_{ij}\right|^{-1/2}|Q|^{-1/2}|G_{22}|^{-1/2} \times [B_1 + B_2 + B_3]^{-1/2(n-\sum_{i=1}^{r}m_i)}. \tag{4.2}$$

With arguments similar to those in Eq. (3.4), we also have the following approximate predictive density of y_{lk}:

$$y_{lk}|Y \dot{\sim} T_q\left(\hat{\mu}, (\hat{B}_1 + \hat{B}_2 + \hat{B}_3)\left(\left(n - \sum_{i=1}^{r} m_i\right)\hat{G}_{22}\right)^{-1}, n - \sum_{i=1}^{r} m_i\right), \tag{4.3}$$

where $(\hat{\Gamma},\hat{\rho})$ is the mode of $P(\Gamma,\rho|Y)$, as given in Eq. (4.2), and the quantities $\hat{\mu}$, $\hat{B}_1$, $\hat{B}_2$, $\hat{B}_3$ and $\hat{G}$ are, respectively, μ, B_1, B_2, B_3, and G evaluated at $(\hat{\Gamma},\hat{\rho})$.

An approximate $1-\alpha$ predictive region for y_{lk} can be obtained from

$$(y_{lk} - \hat{\mu})'\hat{G}_{22}(y_{lk} - \hat{\mu}) \leqslant \frac{q(\hat{B}_1 + \hat{B}_2 + \hat{B}_3)}{n - \sum_{i=1}^{r} m_i}F_\alpha\left(q, n - \sum_{i=1}^{r} m_i\right),$$

where $F_\alpha(v_1, v_2)$ is the upper $100\alpha\%$ point of the F distribution with v_1 and v_2 degrees of freedom.

5. Numerical illustrations

Some of the results developed in this paper are illustrated with a real-data set and some simulations. The real-data set, the obese data, was analyzed by Reinsel (1984) with a random effects model and by Chi and Reinsel (1989) with the model considered in this paper when the parameters are estimated by the ML method. Similar to Chi and Reinsel (1989) we will consider the random slope and AR(1) model for this data set. The moderate simulations are performed with input parameters roughly matching the obese data for the model considered.

In the illustrations, all the observations are assumed to be in the same group, i.e., $r = 1$, for both real data and simulations.

5.1. Parameter estimation

The parameter estimates of the obese data are listed in Table 1. For practical calculations, Γ are transformed into $f'f$ where f is a $w \times w$ upper-triangular matrix. It is seen that estimates from different priors lead to very different estimate for the components of the covariance structure. Generally speaking, prior 1 gives estimates closer to those by the ML method. Priors 2 and 3 put more weights on the variations from AR(1) errors and random effects, respectively. But all the methods considered here give similar estimates for τ.

In Fig. 1, the posterior density functions of f and ρ, as given in Eq. (3.1), corresponding to the three priors for the obese data are plotted. It is noted that all three are unimodal and concentrated, so we may expect that Eq. (3.4) would give good approximations to the exact posterior density functions of τ. Fig. 2 shows the posterior and confidence regions of τ constructed by the Bayesian methods, as given in Eq. (3.6), and by the ML method. Since τ is a 3-variate vector, only projections on the τ_3 plane is presented. The posterior regions constructed with priors 1 and 2 are just a bit larger than and surrounds the region by the ML method. The posterior region constructed by prior 3 is the largest and very different in orientation from the other three. Results from simulations with 1000 replications for the coverage probabilities of these regions are shown in Tables 2 and 3. The input parameters are roughly the ML estimates for the obese data. It is seen that all the coverage probabilities do not reach 0.95 but increase with sample size except those by prior 3. The confidence regions by the ML method are always the smallest and have the smallest coverage probabilities. For smaller

Table 1
Parameter estimates of obese data

	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$	$\hat{\sigma}^2$	$\hat{f}$	$\hat{\rho}$
ML	4.5062	−0.6742	0.2787	0.2511	1.0264	0.6683
Prior 1	4.5065	−0.6738	0.2784	0.2604	1.0167	0.6830
Prior 2	4.5077	−0.6728	0.2776	0.3010	0.8135	0.7237
Prior 3	4.5050	−0.6756	0.2799	0.2081	1.5240	0.6104

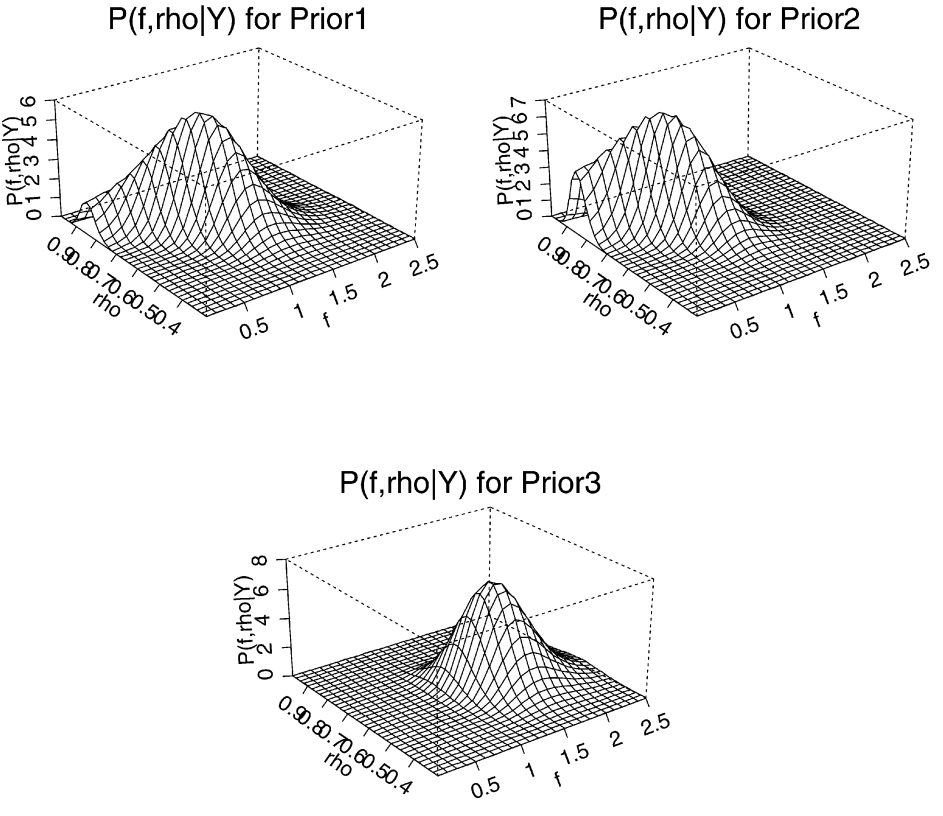

Fig. 1. Posterior densities of f and rho given Y (obese data).

sample sizes, the posterior regions by prior 3 have coverage probabilities closer to 0.95. But for larger sample sizes, the posterior regions by priors 2 and 3 have coverage probabilities larger than the other two.

Using Eq. (3.6), we may construct 0.95 confidence regions as shown in Fig. 3. For this data set these regions are different in sizes but uniform in orientation. Simulations for coverage probabilities reveal similar characteristics as previously discussed for the obese data.

5.2. Prediction of partially observed future value

Here, we consider two types of prediction problem, which are usually of practical interest. The first type, called conditional prediction, involves the following settings:

$$p_1 = \cdots = p_{i-1} = p_{i+1} = \cdots = p_{N-1} = p, \quad p_i = p - 1, \quad i \in \{1, \ldots, N\}$$

and the goal is to predict y_{ip}.

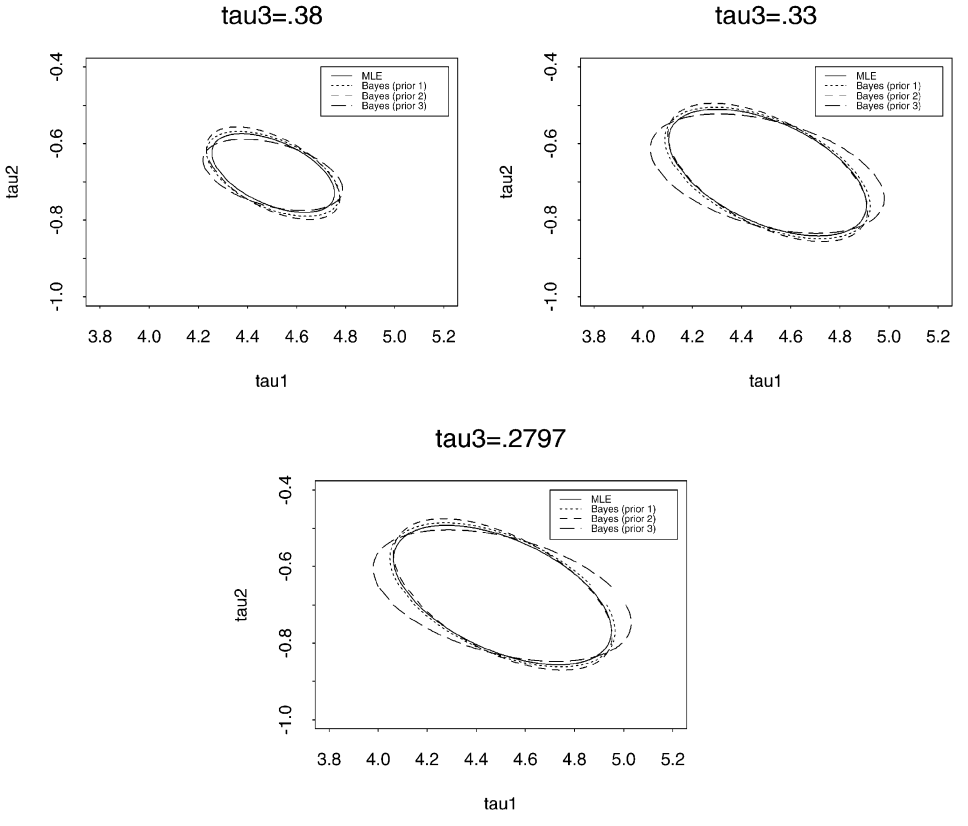

Fig. 2. Posterior and confidence regions of tau (obese data).

The other, called extended prediction, is set as

$$p_1 = \cdots = p_N = p - 1, \quad i \in \{1, \ldots, N\}$$

and y_{ip} is to be predicted.

In conditional prediction, we use the leave-one-out procedure. That is, the last point of each vector measurement is taken out as the true value and to be predicted each time. In extended prediction, the points of the last row are taken out to be predicted.

Fig. 4 shows comparisons among the conditional predictive densities of the last point of the first measurement in the obese data induced from the three priors. It is seen that prior 3 gives a slight location shift from the other two. In Fig. 5 comparisons are shown for exact and approximate predictive density functions using all three priors. It convinces us of the adequacy of the approximation given in Eq. (4.3).

To compare the performances of different prediction methods, we use the following three criteria as measures of discrepancy: mean-square deviation (MSD), mean absolute deviation (MAD), and mean absolute relative deviation (MARD) of the predictions

Table 2
Comparison of coverage probabilities (Random intercept model)

	Prior 1	Prior 2	Prior 3	ML
$N=10$	91.5	91.3	92.5	88.3
$N=20$	93.8	93.4	93.9	92.4
$N=30$	95.2	95.4	95.2	94.4

Here $1-\alpha=0.95$,

$$\tau=\begin{pmatrix}4.5\\-0.68\\0.28\end{pmatrix},\qquad X=\begin{pmatrix}1&1&1&1&1&1&1&1\\0&0.5&1&1.5&2&2&2&2\\0&0&0&0&0&1&2&3\end{pmatrix}',$$

$D=(1\ 1\ 1\ 1\ 1\ 1\ 1\ 1)'$, $\sigma^2=0.25$, $f=1.02$, $\rho=0.6$. and the number of replications is 1000.

Table 3
Comparison of coverage probabilities (Both intercept and slope random)

	Prior 1	Prior 2	Prior 3	ML
$N=10$	86.8	85.3	91.9	83.8
$N=20$	91.0	91.0	94.1	89.7
$N=30$	92.9	92.3	94.6	92.5

Here $1-\alpha=0.95$, parameters as in Table 2 and the number of replications is 1000.

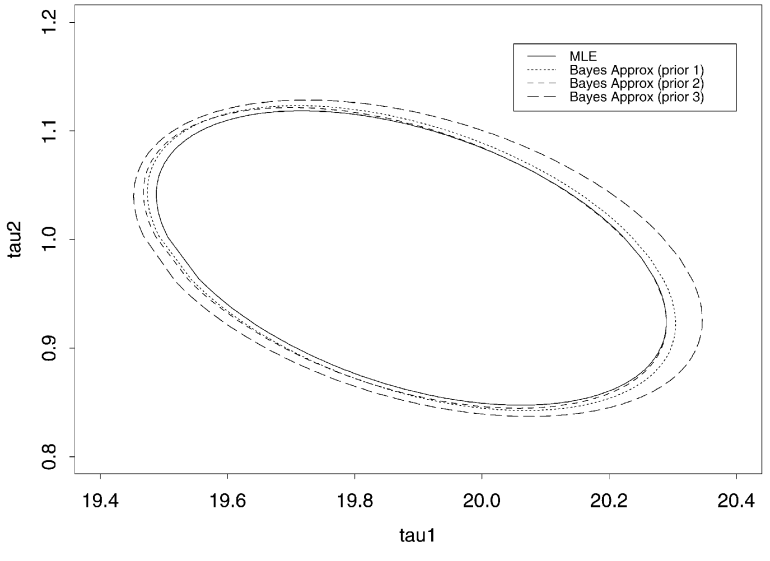

Fig. 3. Ninty-five percent confidence regions for tau (simulated data).

from the actuals. Tables 4 and 5 show all these quantities for all methods for the obese data. It is seen that for the obese data the Bayesian method with priors 1 and 2 and the ML method are somewhat comparable, while with prior 3 does not perform as well.

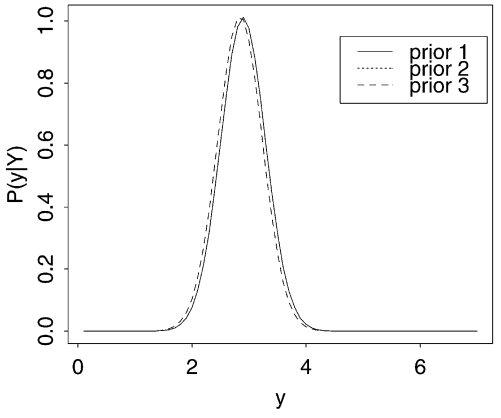

Fig. 4. Conditional predictive densities (8th point of 1st measurement in the obese data).

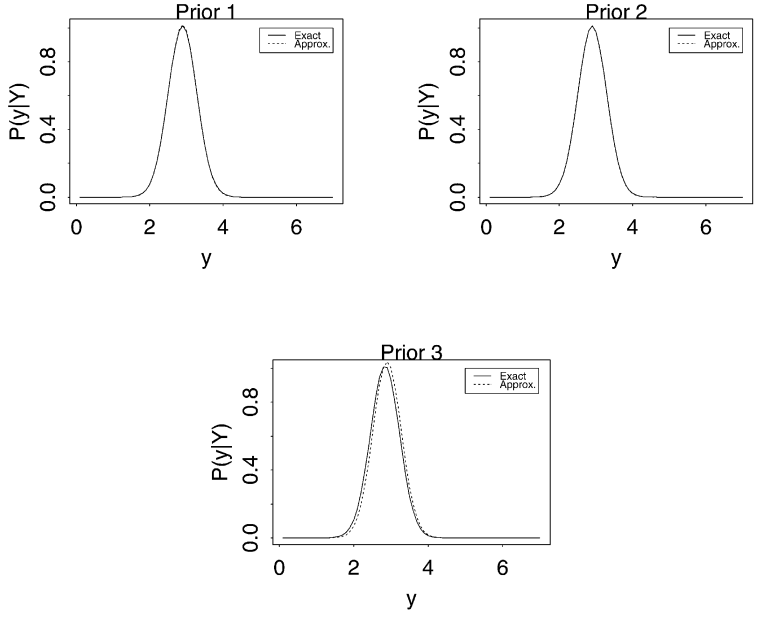

Fig. 5. Comparisons of exact and approximate posterior densities for conditional prediction (obese data).

Table 4
Comparison of prediction accuracy in conditional prediction (obese data)

	Prior 1	Prior 2	Prior 3	ML
MSD	0.0614	0.0598	0.0677	0.0624
MAD	0.2071	0.1992	0.2219	0.2170
MARD	0.0538	0.0517	0.0576	0.0563

Table 5
Comparison of prediction accuracy in extended prediction (obese data)

	Prior 1	Prior 2	Prior 3	ML
MSD	0.0648	0.0632	0.0720	0.0611
MAD	0.2093	0.1989	0.2275	0.2099
MARD	0.0532	0.0504	0.0581	0.0539

In addition to point prediction, we also address the interval prediction. Generally speaking, priors 1 and 2 perform comparably with or even better than the ML method both in point and interval predictions. Prior 3 does not perform as well in point prediction but still gives better interval predictions than the ML method.

6. Conclusions

The Bayesian method presented in this paper provides an alternative way of dealing with the general growth curve model having random effects and AR(1) errors. From the analysis in Section 5 it is evident that model (1.1) is very suitable for the obese data and some related biological measurements.

It is noted that the approximate Bayesian method presented here, especially with prior 1, provides a superior way of constructing more reliable predictive intervals for future values than the ML method, while the forecast accuracies for future values are at least comparable. Also, the approximate Bayesian method is obtained by simply substituting the posterior joint mode of the conditioning variables. The approximation turns out to be quite reasonable for this model.

Finally, it is fair to say that the proposed approximate Bayesian method should be very useful in dealing with growth curve data when the prior is appropriately chosen. Here we recommend prior 1, i.e., the uniform prior, since it is easy to execute and has good performance.

Appendix A. Estimation and prediction based on the ML method

For simplicity we consider balanced data set, in which $p_{ij} = p, Y_{ij} = Y_i, X_{ij} = X, Y = (Y_1, \ldots, Y_N)$.

A.1. Parameter estimation

We have the following lemma similar to Lee (1988).

Lemma A.1. *For model* (1.1) *with balanced design and covariance matrix satisfying Eq.* (1.2), *the MLEs of* τ *and* σ^2, *denoted by* $\hat{\tau}$ *and* $\hat{\sigma}^2$, *are*:

$$\hat{\tau} = \frac{1}{N}(X'\hat{\Lambda}^{-1}X)^{-1}X'\hat{\Lambda}^{-1}\sum_{i=1}^{N} Y_i, \tag{A.1}$$

$$\hat{\sigma}^2 = \frac{1}{pN}[\mathrm{tr}(X'\hat{\Lambda}^{-1}X)^{-1}X'\hat{\Lambda}^{-1}S\hat{\Lambda}^{-1}X + \mathrm{tr}(Z'\hat{\Lambda}Z)^{-1}Z'YY'Z],$$

where Z *is a known* $p \times (p-m)$ *matrix with rank* $p-m$ *such that* $X'Z = 0$,

$$S = Y(I - 11'/N)Y'$$

and $\hat{\Lambda}$ *is obtained with* $\hat{\Gamma}$ *and* $\hat{\rho}$ *maximizing the profile likelihood function*

$$L_{\max}(\Gamma, \rho) = (\hat{\sigma}^2)^{-pN/2}|D\Gamma D' + C|^{-N/2}$$

subject to $\hat{\sigma}^2$ *given by Eq.* (A.1) *with* $(\hat{\Gamma}, \hat{\rho})$ *replaced by* (Γ, ρ).

A.2. Confidence region for τ

Approximately, a $1-\alpha$ confidence region for τ can be constructed from

$$(\tau - \hat{\tau})' \left[\frac{\hat{\sigma}^2}{N}(X'\hat{\Lambda}^{-1}X)^{-1}\right]^{-1} (\tau - \hat{\tau}) \leqslant \chi_m^2(\alpha),$$

where $\chi_m^2(\alpha)$ is the $100(1-\alpha)$th the percentile of χ^2 distribution with m degree of freedom.

A.3. Conditional prediction

In this case, we estimate parameters as $\hat{\tau}_c$, $\hat{\sigma}_c^2$, $\hat{\Gamma}_c$ and $\hat{\rho}_c$ by Eq. (A.1) with sample $Y_c^{(i)} = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_N)$. The approximate mean, denoted by $\hat{y}_{ip}$, of the distribution of y_{ip} given $Y_c^{(i)}$ is

$$\hat{y}_{ip} = X^{(2)}\hat{\tau}_c + \hat{\Lambda}_{21}\hat{\Lambda}_{11}^{-1}(Y_i^{(1)} - X^{(1)}\hat{\tau}_c),$$

where

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix},$$

$X^{(1)}$ is $(p-1) \times m$, Λ_{11} is $(p-1) \times (p-1)$, Λ_{12} is $(p-1) \times 1$ and Λ_{22} is 1×1.

The variance of the prediction error for y_{ip} given σ^2, Γ, and ρ is

$$\sigma_{cf}^2 = \sigma^2 \left[\Lambda_{22} - \Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12} + \frac{1}{N}H_c\left(X'\Lambda^{-1}X\right)^{-1}H_c'\right],$$

where $H_c = X^{(2)} - \Lambda_{21}\Lambda_{11}^{-1}X^{(1)}$. An approximate interval is

$$\hat{y}_{ip} \pm z_{\alpha/2}\hat{\sigma}_{cf},$$

where $\hat{\sigma}_{cf}^2$ is σ_f^2 evaluated at $\hat{\sigma}_c^2$, $\hat{\Gamma}_c$ and $\hat{\rho}_c$, and $z_{\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard normal distribution.

A.4. Extended prediction

Here, we set

$$Y = (Y_1, \ldots, Y_N) = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix},$$

where $Y^{(1)}$ is $(p-1)\times N$, $Y^{(2)}$ is $1 \times N$ and y_{ip} is the ith element of $Y^{(2)}$. We use $Y^{(1)}$ as the sample to get parameter estimates $\hat{\tau}_e$, $\hat{\sigma}_e^2$, $\hat{\Gamma}_e$ and $\hat{\rho}_e$. Again the approximate mean of the distribution of y_{ip} given $Y^{(1)}$ is

$$\hat{y}_{ip} = X^{(2)}\hat{\tau}_e + \hat{\Lambda}_{21}\hat{\Lambda}_{11}^{-1}(Y^{(1)} - X^{(1)}\hat{\tau}_e).$$

The variance of the prediction error for y_{ip} given σ^2, Γ, and ρ is

$$\sigma_{ef}^2 = \sigma^2\left[\Lambda_{22} - \Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12} + \frac{1}{N}H_e\left(X^{(1)'}\Lambda_{11}^{-1}X^{(1)}\right)^{-1}H_e'\right],$$

where $H_e = X^{(2)} - \Lambda_{21}\Lambda_{11}^{-1}X^{(1)}$. An approximate interval is

$$\hat{y}_{ip} \pm z_{\alpha/2}\hat{\sigma}_{ef}^2 \ ,$$

where $\hat{\sigma}_{cf}^2$ is evaluated at $\hat{\sigma}_e^2$, $\hat{\Gamma}_e$ and $\hat{\rho}_e$.

References

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Wiley, New York.

Chi, E.M., Reinsel, G.C., 1989. Models for longitudinal data with random effects and AR(1) errors. J. Amer. Statist. Assoc. 84, 452–459.

Edwards, W., Lindman, H., Savage, L.J., 1963. Bayesian statistical inference for psychological research. Psychol. Rev. 70, 193–242.

Geisser, S., 1970. Bayesian analysis of growth curves. Sankhya Ser. A 32, 53–64.

Jennrich, R.I., Schlucter, M.D., 1986. Unbalanced repeated measures models with structured covariance matrices. Biometrics 42, 805–820.

Keramidas, E.M., Lee, J.C., 1990. Forecasting technological substitutions with concurrent short time series. J. Amer. Statist. Assoc. 85, 625–632.

Khatri, C.G., 1966. A note on a MANOVA model applied to problems in growth curves. Ann. Inst. Statist. Math. 18, 75–86.

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

Lee, J.C., 1988. Prediction and estimation of growth curve with special covariance structure. J. Amer. Statist. Assoc. 83, 432–440.

Lee, J.C., 1991. Test and model selection for the general growth curve model. Biometrics 47, 147–159.

Lee, J.C., Geisser, S., 1972. Growth curve prediction. Sankhya Ser. A 34, 393–412.

Lee, J.C., Hsu, Y.L., 1997. Bayesian analysis of growth curves with AR(1) dependence. J. Statist. Plan. Inference 64, 205–229.

Ljung, G.M., Box, G.E.P., 1980. Analysis of variance with autocorrelated observations. Scand. J. Statist. 7, 172–180.

Potthoff, R.F., Roy, S.N., 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 51, 313–326.

Rao, C.R., 1987. Prediction of future observations in growth curve models. Statist. Sci. 2, 434–471.

Reinsel, G.C., 1984. Estimation and prediction in a multivariate random effects generalized linear model. J. Amer. Statist. Assoc. 79, 406–413.

Zellner, A., Tiao, G.C., 1964. Bayesian analysis of regression model with autocorrelated errors. J. Amer. Statist. Assoc. 59, 763–778.