ELSEVIER

Journal of Statistical Planning and Inference 87 (2000) 87–104

journal of statistical planning and inference

www.elsevier.com/locate/jspi

On estimation and prediction for temporally correlated longitudinal data

Jack C. Lee[a,*], R.C. Hwang[b]

[a] *Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan*

[b] *Department of Accounting and Statistics, Dahan Institute of Technology, Hualien, Taiwan*

Received 20 October 1998; received in revised form 12 August 1999; accepted 1 October 1999

Abstract

In this paper, from a Bayesian point of view, we consider estimation of parameters and prediction of future values for the longitudinal model proposed by Diggle (1988. Biometrics 44, 959–971). This model, called the repeated measures linear model, incorporates group mean, variability among individuals, serial correlation within an individual, and measurement error. Two different priors are employed by the Bayesian approach, one is the noninformative prior and the other is composed of inverse gamma distributions. Given the noninformative prior, it is shown that the resulting approximate estimates of the regression coefficients are the same as those derived by the restricted maximum likelihood estimation. Markov chain Monte Carlo methods are also used to obtain more accurate Bayesian inference for parameters as well as prediction of future values. For parameter estimation and prediction of future values, the advantages of the Bayesian approach over the maximum likelihood method and the restricted maximum likelihood method are demonstrated by both real and simulated data.

MSC: 62F15; 62F10

Keywords: Approximate Bayesian method; Informative prior; Maximum likelihood estimation; Minimum accumulated prediction error; Noninformative prior; Repeated measures linear model; Restricted maximum likelihood estimation

1. Introduction

In this paper, from a Bayesian point of view, we consider estimation of parameters and prediction of future values for the longitudinal model proposed by Diggle (1988). This model, called the repeated measures linear model, incorporates group mean, variability among individuals, serial correlation within an individual, and measurement error, and is defined as

$$Y_{ij} = X_{ij}\beta_i + \tau_{ij}1_{ij} + V_{ij} + \varepsilon_{ij} \tag{1.1}$$

* Corresponding author.

0378-3758/00/$ - see front matter
PII: S0378-3758(99)00184-6

for $j=1,\ldots,N_i$ and $i=1,\ldots,r$. Here Y_{ij} is a $p_{ij}\times 1$ random vector representing p_{ij} observations made on the jth subject in the ith group. The X_{ij} is a known design matrix of order $p_{ij}\times m$, and β_i an unknown $m\times 1$ vector of regression coefficients of the ith group. The τ_{ij}'s are independent and identically distributed normal random variables each with mean 0 and variance σ_τ^2, and 1_{ij} is a $p_{ij}\times 1$ vector of 1's. The V_{ij}'s are independent $p_{ij}\times 1$ normal random vectors each with mean 0 and covariance matrix $\sigma_V^2 C_{ij}$, where C_{ij} stands for a correlation matrix. Finally, ε_{ij}'s are independent $p_{ij}\times 1$ normal random vectors each with mean 0 and covariance matrix $\sigma_\varepsilon^2 I_{ij}$. It is noted that I_{ij} is a $p_{ij}\times p_{ij}$ identity matrix, and X_{ij} is of rank m, $2\leqslant m\leqslant p_{ij}$. The first component $X_{ij}\beta_i$ in (1.1) is interpreted as the ith group mean over individuals. The second component τ_{ij} stands for the individual random effect on the jth subject in the ith group. The third component V_{ij} represents serial correlation in Y_{ij}. The fourth component ε_{ij} denotes measurement errors which are uncorrelated.

By (1.1), the covariance matrix of Y_{ij} can be expressed as

$$\mathrm{Cov}(Y_{ij})=\sigma_V^2\Sigma_{ij}, \tag{1.2}$$

where $\Sigma_{ij}=\phi_1 1_{ij}1_{ij}'+C_{ij}+\phi_2 I_{ij}$, $\phi_1=\sigma_\tau^2/\sigma_V^2$ and $\phi_2=\sigma_\varepsilon^2/\sigma_V^2$. The dependence structure in V_{ij} considered in this paper is an autoregressive process of order 1, i.e., an AR(1) process. Hence $C_{ij}=(\phi_3^{|a-b|})$, for $a,b=1,\ldots,p_{ij}$, and $\phi_3\in(-1,1)$ is the AR(1) parameter. Note that the covariance structure of Y_{ij} consists of a uniform covariance and an AR(1) serial covariance. See Lee (1988) for a detailed explanation. It is also worth noting that this covariance structure, although quite flexible for longitudinal data, may not model well in some situations. In particular, it assumes variance components, and correlation structure, that apply to all subjects and is time independent. A residual variance that slightly increases over time might well be indistinguishable from one that does not in terms of model fit to the observations. However, it could make a large difference in prediction. Furthermore, a more complicated correlation structure such as AR(g) or ARMA could be more appropriate for the data at hand. These are beyond the scope of the current paper and will be subjects of future investigation. Finally, $\phi_1/(\phi_1+\phi_2)=\sigma_\tau^2/(\sigma_\tau^2+\sigma_\varepsilon^2)$ will tends to the intraclass correlation as ϕ_3 tends to zero.

The estimation of parameters for the model has been studied by Diggle (1988) using the methods of maximum likelihood (ML) and restricted maximum likelihood (REML). The prediction of future values for the model has been considered by Donnelly et al. (1995) by using the ML method.

In addition to estimation of parameters, two different prediction problems for the unbalanced repeated measures growth curve model are considered from a Bayesian point of view. For the first prediction problem, let $y_{i\ell}$ be a $q\times 1$ random vector representing q future observations made on the ℓth subject in the ith group, for $\ell=1,\ldots,N_i$ and $i=1,\ldots,r$. We are interested in predicting $y_{i\ell}$ given $Y=(Y_{11}',\ldots,Y_{1N_1}',\ldots,Y_{r1}',\ldots,Y_{rN_r}')'$. This is a time series prediction and thus is important in practice. This prediction problem is called extended prediction of $y_{i\ell}$, by Lee (1988), since the prediction is made beyond the observed time range of the sample. It is noted that in our setting extended prediction is identical to conditional prediction as studied in Lee and Geisser (1972,

1975), Fearn (1975), Rao (1987), Lee (1988, 1991), and Donnelly et al. (1995). For the second prediction problem, let F_i be p future values made on a new subject under study in the ith group. Note that the new subject under study is not contained in the original N_i subjects in the ith group in (1.1). Specifically, F_i is a multivariate normal random vector with mean $X\beta_i$ and covariance matrix $\sigma_V^2\Sigma_p$, where X is a known $p\times m$ design matrix for the new subject under study in the ith group, and Σ_p is of the form specified by (1.2). We will address the prediction of F_i given Y.

In order to estimate parameters and to predict future values for the growth curve problem, the model is considered in this paper from a Bayesian viewpoint with two different priors. The first prior is the noninformative prior (Edwards et al., 1963; Zellner and Tiao, 1964). Using this prior, it is shown that the resulting approximate estimates of the regression coefficients are the same as those obtained by the REML estimation. For a detailed discussion of the REML estimation, see, for example, Patterson and Thompson (1971) and Harville (1974). The second prior is composed of inverse gamma distributions, Gelfand et al. (1990). In order to employ the second prior, the hyperparameters of the inverse gamma distributions are selected by the idea of minimum accumulated prediction error (Rissanen, 1986; Lee and Tsao, 1992). In order to compare relative merits of different methods, we will compare the mean absolute deviation (MAD) and the mean absolute relative deviation (MARD) of the predicted values from the actual observations. For extended prediction of $y_{i\ell}$, it is shown by real data that the approximate Bayesian method employing the second prior performs better than those using the first prior, the ML method and the REML method. A possible advantage of this prior is demonstrated in Table 5 regarding the predictive accuracy of extended prediction of $y_{i\ell}$.

Given the model using two different priors, Bayesian estimation of parameters is considered in Section 2 and prediction problems are presented in Section 3. Bayesian inference via the Markov chain Monte Carlo (MCMC) methodology is considered in Section 4. The results developed in this paper are illustrated in Section 5 with real and simulated data. Finally, some concluding remarks are made in Section 6.

2. Bayesian estimation of the parameters

In this section, the posterior distributions and the posterior regions for the parameters in the model are studied.

For the joint prior of the parameters $\beta, \sigma_V^2, \phi_1, \phi_2$, and ϕ_3, two different priors (2.1) and (2.2) will be used in this paper:

$$\Pi(\beta, \sigma_V^2, \phi_1, \phi_2, \phi_3) \propto \sigma_V^{-2}, \tag{2.1}$$

$$\Pi(\beta, \sigma_V^2, \phi_1, \phi_2, \phi_3) \propto \sigma_V^{-2}\pi(\phi_1)\pi(\phi_2), \tag{2.2}$$

where $\beta = (\beta_1, \ldots, \beta_r)$ and the function π is the inverse gamma distribution with the hyperparameters $\eta > 1$ and $\theta > 0$. Specifically,

$$\pi(x) = \mathrm{IG}(\eta, \theta) \propto x^{-(\eta+1)}\exp(-\theta/x).$$

In both (2.1) and (2.2), we have assumed that $\beta_1,\ldots,\beta_r,\sigma_V^2,\phi_1,\phi_2$, and ϕ_3 have independent prior distributions. In (2.1), no information is available for each of the parameters and it is called noninformative prior (Edwards et al., 1963; Zellner and Tiao, 1964); whereas for (2.2) Gelfand et al. (1990) chose subjectively the hyperparameters η and θ of the inverse gamma distribution in advance. But in this paper we shall use the idea of minimum accumulated prediction error as proposed by Rissanen (1986) and Lee and Tsao (1992) to choose these two hyperparameters. In the accumulated prediction error criterion, the prediction is performed sequentially and was termed "prequential" by Dawid (1984). In this case, we avoid choosing these two hyperparameters subjectively. Note that the priors for ϕ_1 and ϕ_2 given on the right-hand side of (2.2) can be generalized by using two different sets of hyperparameters in their inverse gamma distributions. Here, for simplicity, the same set of hyperparameters is assumed for the two inverse gamma distributions.

We now introduce the idea of the minimum accumulated prediction error criterion for choosing the two hyperparameters (η,θ). This idea will take the minimizer of the accumulated prediction error $S(\eta,\theta)$ over (η,θ) as the selected hyperparameters for (η,θ). The function $S(\eta,\theta)$ is defined by

$$S(\eta,\theta)=\sum_{i=1}^{r}\sum_{j=1}^{N_i}\sum_{k=4}^{p_{ij}}|Y_{ijk}-\hat{Y}_{ijk}(\eta,\theta)|.$$

Here Y_{ijk} is the kth observation in Y_{ij} and $\hat{Y}_{ijk}(\eta,\theta)$ the predicted value of Y_{ijk} when η and θ are used in (2.2). The detailed formulation of $\hat{Y}_{ijk}(\eta,\theta)$ will be given in Section 3. Let $(\hat{\eta},\hat{\theta})$ denote the selected hyperparameters for (η,θ). In this paper, the prior given in (2.1) is called prior 1 and that given in (2.2) with the selected hyperparameters $(\hat{\eta},\hat{\theta})$ prior 2.

By applying the approximate method of Ljung and Box (1980), it can be shown that the approximate posterior distributions of β_i and σ_V^2, and an approximate $100(1-\alpha)\%$ posterior region for $\beta_i,\ i=1,\ldots,r$, are

$$\beta_i|Y\dot{\sim}T_m\left(\hat{\hat{\beta}}_i,\hat{B}_1\left((n-mr)\sum_{j=1}^{N_i}X_{ij}'\hat{\Sigma}_{ij}^{-1}X_{ij}\right)^{-1},n-mr\right), \tag{2.3}$$

$$\sigma_V^2|Y\dot{\sim}\mathrm{IG}((n-mr)/2,\hat{B}_1/2), \tag{2.4}$$

$$(\beta_i-\hat{\hat{\beta}}_i)'\left(\sum_{j=1}^{N_i}X_{ij}'\hat{\Sigma}_{ij}^{-1}X_{ij}\right)(\beta_i-\hat{\hat{\beta}}_i)\leqslant const_1, \tag{2.5}$$

where $n=\sum_{i=1}^{r}\sum_{j=1}^{N_i}p_{ij}$, $const_1=F(1-\alpha;m,n-mr)m(n-mr)^{-1}\hat{B}_1$, $T_p(\mu,\Sigma,n)$ denotes a p-variate T-distribution with mean μ and covariance matrix $n(n-2)^{-1}\Sigma$, and $\hat{\hat{\beta}}_i,\hat{B}_1$ and $\hat{\Sigma}_{ij}$ are the $\hat{\beta}_i,B_1$ and Σ_{ij} with ϕ_1,ϕ_2,ϕ_3 replaced by $\hat{\phi}_1,\hat{\phi}_2,\hat{\phi}_3$, respectively,

$$B_1=\sum_{i=1}^{r}\sum_{j=1}^{N_i}(Y_{ij}-X_{ij}\hat{\beta}_i)'\Sigma_{ij}^{-1}(Y_{ij}-X_{ij}\hat{\beta}_i),$$

$$\hat{\beta}_i = \left(\sum_{j=1}^{N_i} X_{ij}' \Sigma_{ij}^{-1} X_{ij} \right)^{-1} \left(\sum_{j=1}^{N_i} X_{ij}' \Sigma_{ij}^{-1} Y_{ij} \right),$$

and $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$ maximize

$$p(\phi_1, \phi_2, \phi_3 | Y, \text{prior 1}) \propto B_1^{-(n-mr)/2} \prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Sigma_{ij}|^{-1/2} \prod_{i=1}^{r} \left| \sum_{j=1}^{N_i} X_{ij}' \Sigma_{ij}^{-1} X_{ij} \right|^{-1/2}, \tag{2.6}$$

or

$$p(\phi_1, \phi_2, \phi_3 | Y, \text{prior 2}) \propto \phi_1^{-(\hat{\eta}+1)} \phi_2^{-(\hat{\eta}+1)} \exp(-\hat{\theta}/\phi_1 - \hat{\theta}/\phi_2) \times B_1^{-(n-mr)/2} \prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Sigma_{ij}|^{-1/2} \prod_{i=1}^{r} \left| \sum_{j=1}^{N_i} X_{ij}' \Sigma_{ij}^{-1} X_{ij} \right|^{-1/2}. \tag{2.7}$$

We now close this section with the following remarks. Firstly, the joint posterior distribution of ϕ_1, ϕ_2, and ϕ_3 given in (2.6) is proportional to the objective function for choosing the REML estimates of ϕ_1, ϕ_2, and ϕ_3. Hence, the REML estimates of ϕ_1, ϕ_2, and ϕ_3 are identical to the mode of the joint posterior distribution of ϕ_1, ϕ_2, and ϕ_3 as given in (2.6). Secondly, by (2.3), the expectation of the approximate posterior distribution of β_i is $\hat{\hat{\beta}}_i$. It is the same as the REML estimate of β_i. Hence, we know that when prior 1 is used, the REML estimate of β_i is an approximate Bayesian estimate, for each i. To the authors' knowledge, this fact regarding the estimation of β_i has not been given in the literature. Finally, by (2.4), the posterior expectation of σ_V^2 is $\hat{B}_1/(n-2-mr)$ while the REML estimate of σ_V^2 is $\hat{B}_1/(n-mr)$. Hence, when $n-mr$ is sufficiently large, the estimates of σ_V^2 derived by the two methods are very close. However, in small samples, the Bayesian estimate of σ_V^2 will be a bit more biased than the REML estimate. In fact, it is easily seen that

$$E\left(\frac{\hat{B}_1}{n-2-mr} - \sigma_V^2 \right) = \frac{n-mr}{n-2-mr} E\left(\frac{\hat{B}_1}{n-mr} - \sigma_V^2 \right) + \frac{2}{n-2-mr} \sigma_V^2, \tag{2.8}$$

and the REML estimate of σ_V^2 is positively biased as seen in Table 11.

Note that if the joint prior for the parameters $\beta, \sigma_V^2, \phi_1, \phi_2$, and ϕ_3 is proportional to a constant, then the mode of the resulting joint posterior distribution of $\sigma_V^2, \phi_1, \phi_2$, and ϕ_3 is the REML estimates of $\sigma_V^2, \phi_1, \phi_2$, and ϕ_3. This fact has been noted by Harville (1974) for a general variance component model. In this case, the approximate posterior distribution of β_i is a multivariate T distribution as given in (2.3), except that the degree of freedom $n-mr$ is replaced by $n-2-mr$.

3. Prediction

In this section, two different prediction problems for the model will be considered.

We shall first study extended prediction of $y_{i\ell}$, a future q-dimensional observation made on the ℓth subject in the ith group, given Y, for $\ell = 1, \ldots, N_i$ and $i = 1, \ldots, r$,

when each of priors 1 and 2 is employed. To solve this prediction problem, both mean and covariance structure generally have to be extendable to the future values of the individuals observed. We assume that the mean structure is extendable and the covariance structure considered in the paper also satisfies this requirement. Since the estimation of parameters, especially ϕ_3, can be highly sensitive to outliers and the proper dependence structure is important in the prediction of future observation, we would expect that the predictive inference will be quite sensitive to outliers. This topic is beyond the scope of the paper and will be a potential topic in the future.

Let x be a $q \times m$ design matrix corresponding to $y_{i\ell}$ and $\tilde{X}_{i\ell} = \binom{X_{i\ell}}{x}$. We have

$$\mathrm{Cov}\begin{pmatrix} Y_{i\ell} \\ y_{i\ell} \end{pmatrix} = \sigma_V^2(\phi_1 \mathbf{1}\mathbf{1}' + C + \phi_2 \mathbf{I}) = \sigma_V^2 \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}. \tag{3.1}$$

Here $\mathbf{1}$ is a $(p_{i\ell}+q)\times 1$ vector of 1's, $\mathbf{I}$ is a $(p_{i\ell}+q)\times(p_{i\ell}+q)$ identity matrix, and $C = (\phi_3^{|a-b|})$ for $a, b = 1, \ldots, (p_{i\ell}+q)$.

As in Section 2, it can be shown that the approximate predictive distribution and an approximate $100(1-\alpha)\%$ predictive region for $y_{i\ell}$ are

$$y_{i\ell}|Y \dot{\sim} T_q(\hat{\mu}_y, \hat{S}_y((n-mr)\hat{G}_{22})^{-1}, n-mr), \tag{3.2}$$

$$(y_{i\ell} - \hat{\mu}_y)'\hat{G}_{22}(y_{i\ell} - \hat{\mu}_y) \leqslant const_2, \tag{3.3}$$

where

$$Q_1 = \sum_{j\neq\ell}^{N_i} X_{ij}'\Sigma_{ij}^{-1}X_{ij}, \quad Q_2 = \tilde{X}_{i\ell}'\Lambda^{-1}\tilde{X}_{i\ell}, \quad Q = Q_1 + Q_2,$$

$$\hat{\beta}_{i1} = \left(\sum_{j\neq\ell}^{N_i} X_{ij}'\Sigma_{ij}^{-1}X_{ij}\right)^{-1}\left(\sum_{j\neq\ell}^{N_i} X_{ij}'\Sigma_{ij}^{-1}Y_{ij}\right),$$

$$B_2 = \sum_{k\neq i}^{r}\sum_{j=1}^{N_k}(Y_{kj} - X_{kj}\hat{\beta}_k)'\Sigma_{kj}^{-1}(Y_{kj} - X_{kj}\hat{\beta}_k) + \sum_{j\neq\ell}^{N_i}(Y_{ij} - X_{ij}\hat{\beta}_{i1})'\Sigma_{ij}^{-1}(Y_{ij} - X_{ij}\hat{\beta}_{i1}),$$

$$G = \Lambda^{-1}\tilde{X}_{i\ell}(\tilde{X}_{i\ell}'\Lambda^{-1}\tilde{X}_{i\ell})^{-1}Q_1Q^{-1}\tilde{X}_{i\ell}'\Lambda^{-1} + Z(Z'\Lambda Z)^{-1}Z' = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix},$$

$$\mu_y = x\hat{\beta}_{i1} - G_{22}^{-1}G_{21}(Y_{ij} - X_{ij}\hat{\beta}_{i1}), \quad S_y = B_2 + (Y_{i\ell} - X_{i\ell}\hat{\beta}_{i1})'G_{11.2}(Y_{i\ell} - X_{i\ell}\hat{\beta}_{i1}),$$

Z is a $(p_{i\ell}+q)\times(p_{i\ell}+q-m)$ matrix satisfying $\tilde{X}_{i\ell}'Z = 0$, G_{11}, G_{12}, and G_{22} are $p_{i\ell}\times p_{i\ell}$, $p_{i\ell}\times q$, and $q\times q$ matrices, respectively, $G_{21} = G_{12}'$, and $G_{11.2} = G_{11} - G_{12}G_{22}^{-1}G_{21}$, $const_2 = F(1-\alpha; q, n-mr)q(n-mr)^{-1}\hat{S}_y$, and $\hat{\mu}_y, \hat{S}_y$, and $\hat{G}_{22}$ are the μ_y, S_y, and G_{22} with ϕ_1, ϕ_2, and ϕ_3 replaced respectively by $\hat{\phi}_1, \hat{\phi}_2$, and $\hat{\phi}_3$ which maximize

$$p(\phi_1, \phi_2, \phi_3|Y, \text{prior } 1) \propto S_y^{-(n-mr)/2}\prod_{k\neq i}^{r}\prod_{j=1}^{N_k}|\Sigma_{kj}|^{-1/2}\prod_{j\neq\ell}^{N_i}|\Sigma_{ij}|^{-1/2}$$
$$\times(|\Lambda||Q||G_{22}|)^{-1/2}\prod_{k\neq i}^{r}\left|\sum_{j=1}^{N_k}X_{kj}'\Sigma_{kj}^{-1}X_{kj}\right|^{-1/2},$$

or

$$p(\phi_1,\phi_2,\phi_3|Y,\text{prior 2}) \propto \phi_1^{-(\hat{\eta}+1)}\phi_2^{-(\hat{\eta}+1)}\exp(-\hat{\theta}/\phi_1-\hat{\theta}/\phi_2) \times S_y^{-(n-mr)/2}\prod_{k\neq i}^{r}\prod_{j=1}^{N_k}|\Sigma_{kj}|^{-1/2}\prod_{j\neq \ell}^{N_i}|\Sigma_{ij}|^{-1/2} \times(|\Lambda||Q||G_{22}|)^{-1/2}\prod_{k\neq i}^{r}\left|\sum_{j=1}^{N_k}X_{kj}'\Sigma_{kj}^{-1}X_{kj}\right|^{-1/2}.$$

The prediction of future values F_i, a $p\times 1$ vector, given the sample Y is now considered. Similar to $y_{i\ell}$, the approximate predictive distribution and an approximate $100(1-\alpha)\%$ predictive region for F_i are

$$F_i|Y\sim T_p(X\tilde{\beta}_i,\tilde{B}_1((n-mr)\tilde{M})^{-1},n-mr), \tag{3.4}$$

$$(F_i-X\tilde{\beta}_i)'\tilde{M}(F_i-X\tilde{\beta}_i)\leqslant const_3, \tag{3.5}$$

where $const_3=F(1-\alpha;p,n-m)p(n-m)^{-1}\tilde{B}_1$,

$$W_1=\sum_{j=1}^{N_i}X_{ij}'\Sigma_{ij}^{-1}X_{ij},\quad W_2=X'\Sigma_p^{-1}X,\quad W=W_1+W_2,$$

$$M=\Sigma_p^{-1}X(X'\Sigma_p^{-1}X)^{-1}W_1W^{-1}X'\Sigma_p^{-1}+U(U'\Sigma_pU)^{-1}U',$$

U is a $p\times(p-m)$ matrix satisfying $X'U=0$, and $\tilde{\beta}_i$, $\tilde{B}_1$, and $\tilde{M}$ are the $\hat{\beta}_i$, B_1, and M with ϕ_1, ϕ_2, and ϕ_3 replaced respectively by $\hat{\phi}_1$, $\hat{\phi}_2$, and $\hat{\phi}_3$ which maximize

$$p(\phi_1,\phi_2,\phi_3\,|\,Y,\text{prior 1}) \propto B_1^{-(n-mr)/2}\prod_{i=1}^{r}\prod_{j=1}^{N_i}|\Sigma_{ij}|^{-1/2} \times(|\Sigma_p||W||M|)^{-1/2}\prod_{k\neq i}^{r}\left|\sum_{j=1}^{N_k}X_k'\Sigma_{kj}^{-1}X_{kj}\right|^{-1/2},$$

or

$$p(\phi_1,\phi_2,\phi_3\,|\,Y,\text{prior 2}) \propto \phi_1^{-(\hat{\eta}+1)}\phi_2^{-(\hat{\eta}+1)}\exp(-\hat{\theta}/\phi_1-\hat{\theta}/\phi_2)B_1^{-(n-mr)/2} \times\prod_{i=1}^{r}\prod_{j=1}^{N_i}|\Sigma_{ij}|^{-1/2}(|\Sigma_p||W||M|)^{-1/2} \times\prod_{k\neq i}^{r}\left|\sum_{j=1}^{N_k}X_k'\Sigma_{kj}^{-1}X_{kj}\right|^{-1/2}.$$

4. Bayesian inference via MCMC methodology

The MCMC methodology has been extremely popular in statistics since the publication of Gelfand and Smith (1990). For some details see Casella and George (1992) and Gilks et al. (1996). In the rest of this section we will show the results for prior 1 only. They can be adapted for prior 2 easily and are hence omitted.

4.1. Parameter estimation

The joint posterior of $\beta=(\beta_1,\ldots,\beta_r)$, σ_V^2, ϕ_1, ϕ_2, and ϕ_3 can be obtained from combining the likelihood function and the prior (2.1). MCMC proceeds as follows:

1. Generate β_i given β_j, $j\neq i$, σ_V^2, ϕ_1, ϕ_2, ϕ_3 and Y from $\mathrm{N}(\hat{\beta}_i, \sigma_V^2(\sum_{j=1}^{N_i} X_{ij}'\Sigma_{ij}^{-1}X_{ij})^{-1})$.
2. Generate σ_V^2 given β, ϕ_1, ϕ_2, ϕ_3, and Y from $\mathrm{IG}(n/2, S(\beta,\phi_1,\phi_2,\phi_3,Y))$ where $S(\beta,\phi_1,\phi_2,\phi_3,Y)=\sum_{i=1}^{r}\sum_{j=1}^{N_i}(Y_{ij}-X_{ij}\beta_{ij})'\Sigma_{ij}^{-1}(Y_{ij}-X_{ij}\beta_{ij})$.
3. Generate ϕ_1 given β, ϕ_2, ϕ_3 and Y via the Metropolis–Hastings (M–H) algorithm from $g_1(\phi_1'\mid\beta,\phi_2,\phi_3,Y)\propto\prod_{i=1}^{r}\prod_{j=1}^{N_i}|\Sigma_{ij}^*|^{-1/2}\exp[(1/2\sigma_V^2)\,S(\beta,\phi_1',\phi_2,\phi_3,Y)]\exp(\phi_1')$, where $\phi_1'=\log(\phi_1)$, Σ_{ij}^* is Σ_{ij} with ϕ_1 replaced by $\exp(\phi_1')$. Once ϕ_1' is generated, $\phi_1=\exp(\phi_1')$.
4. Generate ϕ_2 given β, ϕ_1, ϕ_3 and Y via the M–H algorithm from $g_2(\phi_2'\mid\beta,\phi_1,\phi_3,Y)$ which is defined as g_1 with ϕ_1' replaced by ϕ_2' and ϕ_2 replaced by ϕ_1.
5. Generate ϕ_3 given β, ϕ_1, ϕ_2 and Y via the M–H algorithm from

$$g_3(\phi_3'\mid\beta,\phi_1,\phi_2,Y)\propto\prod_{i=1}^{r}\prod_{j=1}^{N_i}|\Sigma_{ij}^*|^{-1/2}\exp\left[-\frac{1}{2\sigma_V^2}\,S(\beta,\phi_1,\phi_2,\phi_3',Y)\right]\times[2\exp(\phi_3')][1+\exp(\phi_3')]^{-2},$$

where $\phi_3'=\log(1+\phi_3)(1-\phi_3)^{-1}$, Σ_{ij}^* is Σ_{ij} with ϕ_3 replaced by $[\exp(\phi_3')-1][\exp(\phi_3')+1]^{-1}$. Once ϕ_3' is generated, $\phi_3=[\exp(\phi_3')-1][\exp(\phi_3')+1]^{-1}$.

In order to ensure that the samples are drawn from the domain of the entire density, Gelman and Rubin (1992) suggested using "overdispersed" starting values in multiple chains to assist in drawing from the domain of the entire density and assessing convergence by using the potential scale reduction measure $\sqrt{\hat{R}}$. According to this viewpoint, we implemented the MCMC sampling using seven chains with different starting values. The starting values for each variable of interest are the MLE and six others obtained from MLE $\pm\, k$ times standard deviation, for $k=1,3,5$. For each chain, after a sufficiently long burn-in iterations, we then use the remaining samples as simulated from the variable of interest.

4.2. Prediction

We have from (3.1) that $y_{ij}\mid\theta,Y\sim\mathrm{N}(\mu_{2.1},\sigma_V^2\Lambda_{22.1})$, where $\theta=(\beta,\sigma_V^2,\phi_1,\phi_2,\phi_3)$, $\mu_{2.1}=x\beta_i+\Lambda_{21}\Lambda_{11}^{-1}(Y_{i\ell}-X_{i\ell}\beta_i)$, $\Lambda_{22.1}=\Lambda_{22}-\Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12}$, and Λ_{ij} is defined in (3.1). Hence we can generate $y_{i\ell}^{(k,s)}$ from $f(y_{i\ell}\mid\theta^{(k,s)},Y)$ where $\theta^{(k,s)}$ is the kth iteration and sth replication of the MCMC sampler of θ. Thus, we can predict $y_{i\ell}$ by the mean

$$\hat{y}_{i\ell}=\frac{1}{mn}\sum_{s=1}^{m}\sum_{k=1}^{n}y_{i\ell}^{(k,s)}$$

or the median of the samples, i.e.,

$$\hat{y}_{i\ell}=\frac{1}{m}\sum_{s=1}^{m}y_{0.5}^{(s)},$$

where $y_{0.5}^{(s)}$ is the median of the last n iterations for each chain. Prediction intervals and quantiles of $y_{i\ell}$ can also be obtained from the sampler $y_{i\ell}^{(k,s)}$.

It is noted that the prediction of future values F_i can be done in a similar manner.

5. Numerical illustration

In order to get a further insight into the results obtained in Sections 2–4, we will illustrate with a real example and a simulation study.

5.1. A real example

In this section, the results obtained in Sections 2–4 are applied to a real example. The data, given in Table 1, consist of the weights of 23 calves, each observed from 0 to 18 weeks with an increment of 2 weeks. They are given in Group B of Table 6.1 of Diggle et al. (1994) with the weights on 19th week excluded and with the exception of individuals $1, 9, 11, 19, 23, 26$ and 28. The seven calves are excluded from consideration for this illustration because their growth patterns are irregular in the last two weeks which make them unsuitable for the model.

Table 1
Weights (100 kg) of calves

	Time in weeks									
	0	2	4	6	8	10	12	14	16	18
1	2.30	2.40	2.58	2.77	2.77	2.93	3.00	3.23	3.27	3.40
2	2.26	2.33	2.48	2.77	2.97	3.13	3.22	3.40	3.54	3.65
3	2.33	2.39	2.53	2.77	2.92	3.10	3.18	3.33	3.36	3.53
4	2.38	2.41	2.62	2.82	3.00	3.14	3.19	3.31	3.38	3.48
5	2.25	2.28	2.37	2.61	2.71	2.88	3.00	3.16	3.19	3.33
6	2.24	2.25	2.39	2.57	2.68	2.90	3.04	3.13	3.10	3.18
7	2.37	2.41	2.55	2.76	2.93	3.07	3.12	3.36	3.36	3.44
8	2.33	2.39	2.59	2.83	2.94	3.13	3.20	3.47	3.48	3.62
9	2.28	2.23	2.46	2.66	2.77	2.87	3.00	3.12	3.08	3.28
10	2.41	2.47	2.68	2.90	3.09	3.23	3.36	3.48	3.59	3.72
11	2.21	2.21	2.40	2.53	2.73	2.82	2.92	3.07	3.06	3.17
12	2.17	2.20	2.35	2.59	2.62	2.76	2.84	3.05	3.03	3.15
13	2.14	2.21	2.37	2.56	2.71	2.83	2.87	3.14	3.16	3.20
14	2.24	2.31	2.41	2.56	2.65	2.83	2.95	3.14	3.13	3.28
15	2.00	2.03	2.21	2.36	2.48	2.62	2.76	2.94	2.91	3.11
16	2.30	2.22	2.43	2.53	2.68	2.84	2.90	3.16	3.14	3.30
17	2.17	2.24	2.42	2.65	2.84	3.02	3.09	3.24	3.28	3.38
18	2.09	2.09	2.21	2.38	2.56	2.67	2.81	2.95	3.01	3.09
19	2.30	2.31	2.44	2.61	2.72	2.83	2.94	3.18	3.20	3.33
20	2.16	2.18	2.23	2.43	2.59	2.70	2.70	2.90	3.01	3.14
21	2.07	2.16	2.28	2.55	2.75	2.85	2.96	3.14	3.19	3.30
22	2.21	2.32	2.51	2.84	2.84	2.95	3.00	3.23	3.19	3.33
23	2.33	2.38	2.54	2.66	2.82	2.94	2.95	3.10	3.20	3.27

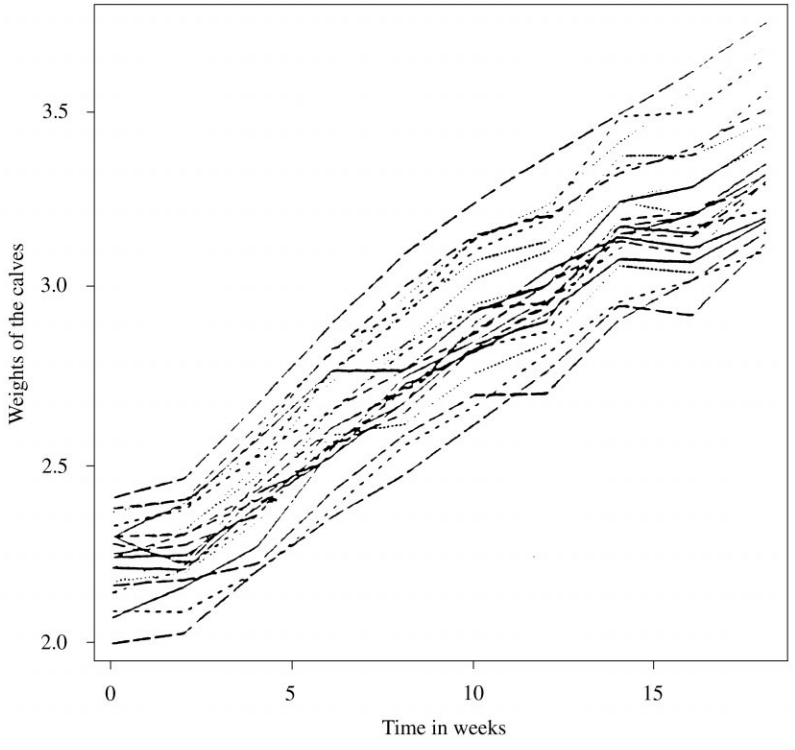


Fig. 1. Weights of calves.

To get the ML, REML and the approximate Bayesian estimates of parameters in the model and to predict future values, the data were divided by 100 to avoid "overflow" or "underflow" problems. They are plotted in Fig. 1.

Given the model and prior 1, in order to predict the values in the 18th week, the MCMC simulated samples of the related parameters were produced by using the data in the first 16 weeks. In order to fit the model, we first examine Fig. 1 which shows clearly that the data have a linear trend and consist of only one group. By this and the fact that the times of measurement are equally spaced, the value of r was taken as $r = 1$, the vector of regression coefficients $\beta_1 = (\beta_{11}, \beta_{12})'$ and the design matrix for each calf is

$$X = \begin{bmatrix} 1, 1, \ldots, 1 \\ 1, 2, \ldots, 9 \end{bmatrix}'.$$

Tables 2 and 3 give the mean, standard deviation, and 2.5%, 5%, 25%, 50%, 75%, 95%, and 97.5% quantiles of the 7000 MCMC simulated samples of $(\beta_1, \sigma_V^2, \phi_1, \phi_2, \phi_3)$ for priors 1 and 2, respectively. We can then obtain point and interval estimates for any parameter of interest. The parameter estimates from ML and REML are shown in Table 4.

Table 2
MCMC simulated posterior distributions using prior 1

	β_{11}	β_{12}	σ_V^2	ϕ_1	ϕ_2	ϕ_3
Mean	2.1039	0.1259	0.0088	1.2051	0.0860	0.8196
S.D.	0.0372	0.0034	0.0011	0.6511	0.0621	0.0543
2.5%	2.0106	0.1195	0.0071	0.3595	0.0059	0.7268
5%	2.0255	0.1205	0.0074	0.4286	0.0083	0.7366
25%	2.0858	0.1237	0.0081	0.6065	0.0269	0.7742
50%	2.1088	0.1258	0.0085	1.1056	0.0786	0.8195
75%	2.1278	0.1282	0.0092	1.6997	0.1347	0.8622
95%	2.1535	0.1321	0.0109	2.4248	0.1905	0.9053
97.5%	2.1666	0.1329	0.0115	2.5738	0.2143	0.9165

Table 3
MCMC simulated posterior distributions using prior 2

	β_{11}	β_{12}	σ_V^2	ϕ_1	ϕ_2	ϕ_3
Mean	2.1041	0.1261	0.0078	0.0623	0.0845	0.8068
S.D.	0.0216	0.0032	0.0012	0.0152	0.0281	0.0506
2.5%	2.0605	0.1199	0.0049	0.0317	0.0430	0.7229
5%	2.0690	0.1209	0.0057	0.0404	0.0481	0.7380
25%	2.0894	0.1239	0.0072	0.0527	0.0641	0.7716
50%	2.1039	0.1261	0.0079	0.0613	0.0795	0.7995
75%	2.1192	0.1282	0.0085	0.0707	0.0992	0.8328
95%	2.1396	0.1314	0.0095	0.0883	0.1392	0.9069
97.5%	2.1462	0.1321	0.0098	0.0952	0.1518	0.9295

Table 4
Comparison of estimates from ML and REML using real data

	β_{11}	β_{12}	σ_V^2	ϕ_1	ϕ_2	ϕ_3
ML	2.0962	0.1277	0.0079	1.5448	0.1911	0.8245
REML	2.0962	0.1276	0.0087	1.4019	0.1753	0.8429

We now compare prediction ability among the eight prediction methods: ML method, REML method, approximate Bayesian method, and the mean and the median of the MCMC samples for both priors. The ML method has been used in prediction by Donnelly et al. (1995). The approximate Bayesian method is to take $\hat{\mu}_y$ in (3.3) for both priors 1 and 2 as the predictor for $y_{i\ell}$. To get $\hat{\mu}_y$ for prior 2, the minimum accumulated prediction error criterion was used to select the inverse gamma hyperparameters η and θ over the rectangle $(1,40]\times(0,10]$. The values of $S(\eta,\theta)$ were calculated on 1950×500 equally spaced grid points in the rectangle. The minimizer of these values was taken as the selected value of (η,θ) for prior 2.

Using the model and the data in the first 16 weeks in Fig. 1, each of the above eight methods was used to predict the weight of each calf in the 18th week. For the weight of the first calf in the 18th week given all the data in the first 16 weeks, using prior 1, Fig. 2 gives the exact predictive density (solid curve), approximate Bayesian predictive density (dashed curve), and MCMC predictive density (dotted curve). Similar

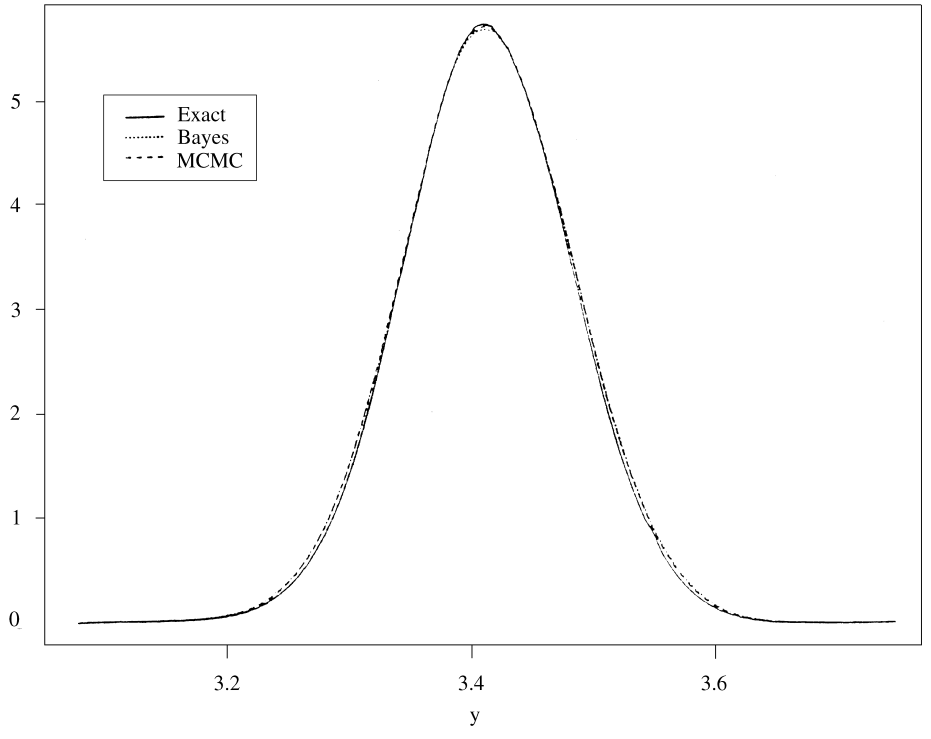


Fig. 2. Comparison of predictive densities $f(y|Y)$ using prior 1.

comparisons are shown in Fig. 3 when using prior 2. Figs. 2 and 3 show that the approximate Bayesian predictive density and the MCMC predictive density are very close to the exact predictive density. This indicates that the predictive density can be approximated closely by both approximate Bayesian and MCMC sampling methods. It is noted that the exact Bayesian predictive densities are obtained by numerical integrations.

Table 5 gives the MAD and MARD for each of the above eight prediction methods. Table 5 shows that REML and the approximate Bayesian method using prior 1 performs about as well as MLE. However, the approximate Bayesian method using prior 2 performs better. The best among the methods compared are MCMC mean and median using either prior. Thus, MCMC methods are quite encouraging for this model.

Fig. 4 gives 95% confidence and posterior regions for β_1 derived respectively by the ML method (solid curve) and the approximate Bayesian method using prior 1 (dashed curve) and using prior 2 (short dashed curve). It also shows the point estimates of β_1 obtained by the three methods denoted by the circle, triangle, and plus signs, respectively. Fig. 4 shows that the point estimates of β_1 derived by the ML method and the approximate Bayesian method using prior 1 are very close. The 95% posterior region derived by the approximate Bayesian method using prior 1 is slightly larger than the confidence region obtained by the ML method. But the approximate Bayesian method

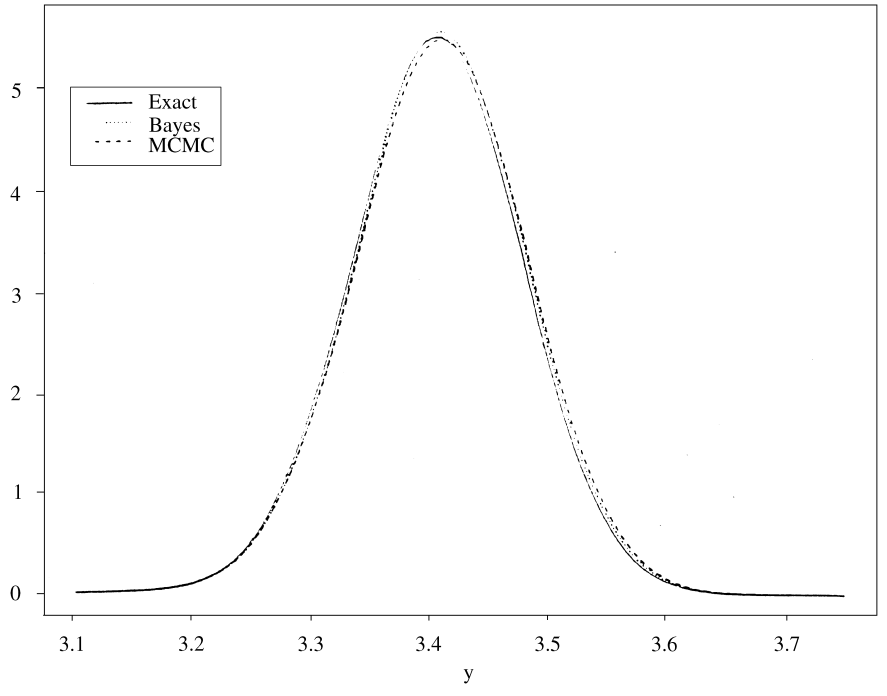


Fig. 3. Comparison of predictive densities $f(y|Y)$ using prior 2.

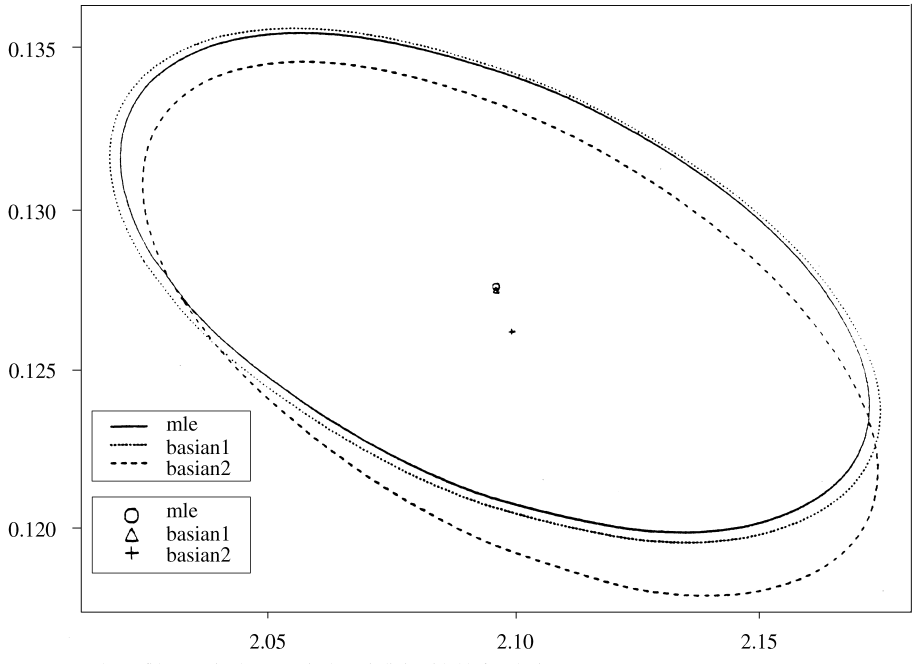
Table 5
Comparison of prediction methods using real data

	MLE	REML	Prior 1	Prior 2	MCMC(Prior 1)		MCMC(Prior 2)	
					Mean	Median	Mean	Median
MAD	0.0412	0.0412	0.0412	0.0398	0.0398	0.0380	0.0386	0.0386
MARD	0.0126	0.0126	0.0126	0.0120	0.0121	0.0116	0.0117	0.0117

using prior 2 has the largest posterior region for β_1. It is noted that the confidence region obtained by the REML method is almost indistinguishable from that of the approximate Bayesian method using prior 1.

5.2. A simulation study

In this section, a simulation study was conducted to compare both the coverage probabilities of confidence and posterior regions of β_1 and the predictive intervals of $y_{i\ell}$ derived by the ML method, the REML method and those obtained by the approximate Bayesian methods using priors 1 and 2. We now describe the simulation settings. The simulated data were independently generated from the model with $r = 1$, $\phi_1 = 0.5$,



* The confidence region by REML is almost indistinguishable from basian 1

Fig. 4. 95% confidence and posterior region of beta.

Table 6
Comparison of coverage probabilities for $\beta_1(1-\alpha=0.95)$

N	p	MLE	REML	Prior 1	Prior 2
5	5	0.829	0.870	0.894	0.915
10	5	0.904	0.921	0.934	0.957
15	5	0.923	0.935	0.936	0.948
20	5	0.933	0.941	0.941	0.953

$\phi_2 = 0.1,\ \phi_3 = 0.8,\ \sigma_V^2 = 0.2,\ \beta_1 = (10,1)'$, and

$$X = \begin{bmatrix} 1,1,\ldots,1 \\ 1,2,\ldots,p \end{bmatrix}'.$$

Given $p=5$, for $N=5,10,15$, and 20, Table 6 gives the coverage probabilities of 95% confidence and posterior regions for β_1. Given the values of N and p, each data set contains N subjects and for each subject p measurements were made. For each combination of N and p, 1000 independent data sets were generated. Table 6 shows that the approximate posterior regions, using both priors 1 and 2, have larger coverage probabilities (and closer to 0.95) than the confidence regions derived by the ML method and the REML method.

Given $p=5$, for $N=5,10,15$, and 20, Table 7 gives the coverage probabilities of 95% predictive intervals for $y_{i\ell}$, the $(p+1)$th measurement for each subject. Given

Table 7
Comparison of coverage probabilities for $y_{i\ell}(1-\alpha=0.95)$

N	p	MLE	REML	Prior 1	Prior 2
5	5	0.8724	0.8798	0.9346	0.9372
10	5	0.9102	0.9156	0.9414	0.9472
15	5	0.9237	0.9273	0.9439	0.9506
20	5	0.9298	0.9313	0.9440	0.9508

Table 8
Comparison of prediction method (MAD)

N	p	MLE		REML		Prior 1		Prior 2	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
5	5	0.2796	0.0940	0.2782	0.0937	0.2783	0.0937	0.2757	0.0946
10	5	0.2716	0.0641	0.2710	0.0640	0.2710	0.0640	0.2681	0.0632
15	5	0.2676	0.0487	0.2673	0.0486	0.2673	0.0485	0.2648	0.0479
20	5	0.2653	0.0453	0.2651	0.0452	0.2651	0.0452	0.2629	0.0448

Table 9
Comparison of prediction methods (MARD)

N	p	MLE		REML		Prior 1		Prior 2	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
5	5	0.0175	0.0059	0.0174	0.0059	0.0174	0.0059	0.0173	0.0059
10	5	0.0170	0.0040	0.0170	0.0040	0.0170	0.0040	0.0168	0.0040
15	5	0.0168	0.0031	0.0167	0.0030	0.0167	0.0030	0.0166	0.0030
20	5	0.0166	0.0028	0.0166	0.0028	0.0166	0.0028	0.0165	0.0028

the values of N and p, each data set contains N subjects and for each subject $p+1$ measurements were made. Among the N subjects, the first p measurements were used to predict $y_{i\ell}$. For each combination of N and p, 1000 independent data sets were generated. Hence there were $N\times 1000$ predicted values to be compared with $N\times 1000$ true values. Table 7 shows that the approximate Bayesian predictive intervals using both priors 1 and 2 have larger coverage probabilities (and closer to 0.95) than those for the intervals derived from the ML method and the REML method.

Given $p=5$, for $N=5, 10, 15$, and 20, Tables 8 and 9 give the prediction comparisons for the various methods in terms of MAD and MARD, respectively. Prediction intervals from the four methods are given in Table 10.

All four predictors perform about equally, although prior 2 is slightly better. Finally, we will show biasedness of the estimate for σ_V^2. Given $p=5$, for $N=5, 10, 15$, and 20 and set $\sigma_V^2=0.2$, Table 11 shows that the mean of the REML estimates is positively biased.

Table 10
Comparison of prediction intervals

True weight	MLE		REML		Prior 1		Prior 2	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
3.4000	3.2809	3.5508	3.2798	3.5507	3.2767	3.5538	3.2652	3.5439
3.6500	3.4854	3.7553	3.4873	3.7582	3.4842	3.7613	3.4956	3.7743
3.5300	3.3676	3.6365	3.3664	3.6372	3.3633	3.6403	3.3519	3.6306
3.4800	3.3827	3.6517	3.3809	3.6517	3.3778	3.6548	3.3644	3.6432
3.3300	3.2109	3.4808	3.2102	3.4810	3.2071	3.4841	3.1957	3.4744
3.1800	3.1526	3.4225	3.1512	3.4221	3.1481	3.4252	3.1263	3.4050
3.4400	3.3702	3.6401	3.3698	3.6406	3.3667	3.6437	3.3544	3.6331
3.6200	3.4653	3.7353	3.4663	3.7372	3.4632	3.7403	3.4610	3.7397
3.2800	3.1405	3.4104	3.1384	3.4092	3.1353	3.4123	3.1090	3.3878
3.7200	3.5519	3.8218	3.5528	3.8236	3.5497	3.8267	3.5488	3.8278
3.1700	3.1092	3.3791	3.1075	3.3784	3.1044	3.3815	3.0845	3.3632
3.1500	3.0794	3.3493	3.0777	3.3486	3.0746	3.3517	3.0568	3.3355
3.2000	3.1756	3.4455	3.1753	3.4461	3.1722	3.4492	3.1666	3.4454
3.2800	3.1671	3.4371	3.1660	3.4368	3.1629	3.4399	3.1469	3.4256
3.1100	2.9660	3.2359	2.9650	3.2359	2.9619	3.2390	2.9491	3.2278
3.3000	3.1753	3.4452	3.1742	3.4450	3.1711	3.4481	3.1559	3.4346
3.3800	3.2845	3.5544	3.2848	3.5556	3.2817	3.5587	3.2767	3.5554
3.0900	3.0374	3.3074	3.0366	3.3075	3.0335	3.3106	3.0261	3.3048
3.3300	3.2197	3.4896	3.2187	3.4896	3.2156	3.4927	3.2041	3.4828
3.1400	3.0299	3.2998	3.0281	3.2989	3.0250	3.3020	3.0163	3.2940
3.3000	3.1942	3.4641	3.1947	3.4655	3.1916	3.4686	3.1913	3.4690
3.3300	3.2277	3.4976	3.2266	3.4975	3.2235	3.5006	3.2063	3.4840
3.2700	3.2147	3.4846	3.2126	3.4835	3.2095	3.4866	3.1938	3.4725

Table 11
REML estimate of σ_V^2 [a]

N	p	Mean	S.D.
5	5	0.2216	0.1546
10	5	0.2270	0.1311
15	5	0.2211	0.1125
20	5	0.2231	0.1043

[a]The true value of σ_V^2 is 0.2.

6. Concluding remarks

In this paper we consider, from a Bayesian viewpoint, estimation of parameters and prediction of future values for the longitudinal model proposed by Diggle (1988). Two different priors are employed in this study. For the noninformative prior, it is found that the approximate Bayesian estimates of the regression coefficients are identical to the REML estimates. For the informative prior, the inverse gamma distribution is assumed for both variance ratios with the same set of hyperparameters which are determined from the sample using the minimum accumulated prediction error criterion.

In real and simulated data, it is found that the proposed methods are quite encouraging. For the prediction of future values, the approximate Bayesian predictor using noninformative prior performs about the same as the REML predictor, but slightly better than the ML predictor, although the approximate Bayesian predictor using informative prior is even better. The best among the predictors compared are the MCMC mean and median using either noninformative or informative priors. For the confidence and posterior regions, it is found that the posterior region based on the approximate posterior distribution of the regression coefficients when noninformative prior is employed has a much more accurate coverage probability than the ML method and the REML method when the sample size is small. Thus we believe that the proposed approximate method is a useful alternative to the ML method and the REML method in dealing with the longitudinal model proposed by Diggle (1988). Of course, better results are those using the MCMC simulation as suggested in the paper, although the implementation is more involved. However, the steps outlined in the paper can be followed rather easily.

References

Casella, G., George, E.I., 1992. Explaining the Gibbs sampler. Amer. Statist. 46, 167–174.

Dawid, A.P., 1984. Present position and potential developments: some personal views. Statistical theory — the prequential approach (with distribution). J. Roy. Statist. Soc. Ser. A 147, 278–292.

Diggle, P.J., 1988. An approach to the analysis of repeated measurements. Biometrics 44, 959–971.

Diggle, P.J., Liang, K.Y., Zeger, S.L., 1994. Analysis of Longitudinal Data. Oxford Statistical Science Series, Vol. 13. Oxford Science Publications, Oxford.

Donnelly, C.A., Laird, N.M., Ware, J.H., 1995. Prediction and creation of smooth curve for temporally correlated longitudinal data. J. Amer. Statist. Assoc. 90, 984–989.

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

Fearn, T., 1975. A Bayesian approach to growth curves. Biometrika 62, 89–100.

Gelfand, A.E., Hills, S.E., Racine-Poon, Smith, A.F.M., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. J. Amer. Statist. Assoc. 85, 972–985.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculating densities. J. Amer. Statist. Assoc. 85, 298–309.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statist. Sci. 7, 457–472.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall, London.

Harville, D., 1974. Bayesian inference for variance components using only error contrasts. Biometrika 61, 383–385.

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

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

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

Lee, J.C., Geisser, S., 1975. Applications of growth curve prediction. Sankhyā A 37, 239–256.

Lee, J.C., Tsao, S.L., 1992. On estimation and prediction procedures for AR(1) models with power transformation. J. Forecast. 12, 499–511.

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

Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554.

Rao, C.R., 1987. Prediction of future observations in growth curve models. Statist. Sci. 2, 434–471.
Rissanen, J., 1986. Order estimation by accumulated prediction errors. In: Gani, J., Priestley, M.B. (Eds.), Essays in Time Series and Allied Processes — Papers in Honor of E.J. Hannan. Applied Probability Trust, Sheffield, England.
Zellner, A., Tiao, G.C., 1964. Bayesian analysis of regression model with autocorrelated errors. J. Amer. Statist. Assoc. 59, 763–778.