
Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

# Bayesian analysis of a growth curve model with power transformation, random effects and AR(1) dependence

Jack C. Lee [a] & W. H. Lien [a]
[a] Institute of Statistics , National Chiao Tung University , Taiwan
Published online: 02 Aug 2010.

*Journal of Applied Statistics, Vol. 28, No. 2, 2001, 223–238*

# Bayesian analysis of a growth curve model with power transformation, random effects and AR(1) dependence

**JACK C. LEE & W. H. LIEN,** *Institute of Statistics, National Chiao Tung University, Taiwan*



ABSTRACT *In this paper we devote ourselves to a general growth curve model with power transformation, random effects and AR(1) dependence via a Bayesian approach. Two priors are proposed and both parameter estimation and prediction of future values are considered. Some numerical results with a set of real data are also given.*

## 1 Introduction

In this paper we consider analysing growth curve data through a Bayesian point of view using a growth curve model with random effects and AR(1) dependence, while applying the Box–Cox transformation on observations.

Since Poffhoff & Roy (1964) the growth curve model has been considered by many authors. Laird & Ware (1982) considered the random effects model with white noise errors. Jennrich & Schluchter (1986) discussed various types of covariance structures, including random effects models and the AR(1) model separately. Chi & Reinsel (1989) considered the ML estimates for the model with both random effects and AR(1) errors by the score method. Lee & Niu (1999) considered the model from a Bayesian point of view and suggested some approximations that are better than the ML approach. The importance of the AR(1) dependence was demonstrated in Lee (1988, 1991) and Keramidas & Lee (1990). In Lee & Lu (1987) tremendous improvement was found in predictive accuracy using the data-based transformation models for technology substitutions. This is primarily due to the fact that the linearity assumption for the growth function can be enhanced significantly with the Box–Cox power transformation, along with incorporating into the model the proper dependence structure among the observations. Enhancement

*Correspondence*: J. C. Lee, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

ISSN 0266-4763 print; 1360-0532 online/01/020223-16 
DOI: 10.1080/02664760020016127

of normality and constancy of variance may have relatively minor roles in the improvement of predictive accuracy.

The model considered here is:

$$\underset{p_{ij}\times 1}{Y_{ij}^{(\lambda_i)}} = \underset{p_{ij}\times m_i}{X_{ij}}\ \underset{m_i\times 1}{\tau_i} + \underset{p_{ij}\times w_i}{D_{ij}}\ \underset{w_i\times 1}{U_{ij}} + \underset{p_{ij}\times 1}{\varepsilon_{ij}}, \tag{1}$$

$$\text{for } i = 1, 2, \ldots, r; j = 1, 2, \ldots, N_i$$

where $Y_{ij}$ are the measurements and not necessarily of equal lengths, $\tau_i$ is an unknown $m_i \times 1$ vector of regression coefficients of group $i$, $X_{ij}$ and $D_{ij}$ are known design matrices, $U_{ij}$ is the random effect error and possesses $w_i$-variate normal distribution with mean vector 0 and covariance matrix $\sigma^2\Gamma_i$, and the disturbance term $\varepsilon_{ij}$ is an independent $p_{ij}$-variate normal distribution with mean 0 and covariance matrix $\sigma^2 C_{ij}$, where $C_{ij} = (\rho^{|a-b|})$, $a, b = 1, 2, \ldots, p_{ij}$; and $\sigma > 0$ and $-1 < \rho < 1$ are unknown. The Box–Cox power transformation is defined as:

$$Y_{ijk}^{(\lambda_i)} = \begin{cases} \dfrac{(Y_{ijk}+\gamma)^{\lambda_i} - 1}{\lambda_i}, & \text{if } \lambda_i \neq 0, \\ \log(Y_{ijk}+\gamma), & \text{if } \lambda_i = 0, \end{cases} \tag{2}$$

for $i = 1, 2, \ldots, r; j = 1, 2, \ldots, N_i; k = 1, 2, \ldots, p_{ij}$

where $Y_{ijk}$ is the $k$th component of $Y_{ij}$, $\gamma$ is a known constant such that $(Y_{ijk} + \gamma) > 0$ for all $i$, $j$, $k$, and $\lambda_i$ is an unknown parameter. Thus, the covariance matrix of $Y_{ij}^{(\lambda_i)}$ can be written as

$$\Sigma_{ij} = \sigma^2(D_{ij}\Gamma_i D_{ij}' + C_{ij}) \equiv \sigma^2\Lambda_{ij} \tag{3}$$

For the choice of priors, there are two possibilities considered for our Bayesian analysis of model (1). In addition to parameter estimation, we also derive two specific types of prediction problems useful in practice. Furthermore, in recent years, statisticians have been increasingly drawn to Markov chain Monte Carlo (MCMC) methods, especially the Metropolis–Hastings (M-H) algorithm (Metropolis *et al.*, 1953; Hastings, 1970) and the Gibbs sampler (Geman & Geman, 1984). Therefore, we will consider Bayesian point and interval estimates for each of the unknown parameters and prediction of future observations via MCMC methods.

In Section 2, two types of priors are introduced. In Sections 3 and 4, Bayesian methods for parameter estimation and prediction of future values are developed. In Section 5, Bayesian inference by means of MCMC methods is studied for model (1). Some numerical results, with real data analysed by the methods developed in this paper are illustrated and compared with the method of maximum likelihood (ML) in Section 6. Finally, some concluding remarks are given in Section 7.

## 2 Choices of priors

For convenience, we shall denote $(\tau_1, \tau_2, \ldots, \tau_r)$, $(\Gamma_1, \Gamma_2, \ldots, \Gamma_r)$ and $(\lambda_1, \lambda_2, \ldots, \lambda_r)$ as $\tau$, $\Gamma$ and $\lambda$, respectively.

To choose prior densities, we will assume $\tau$, $\sigma^2$, $\Gamma$, $\rho$ and $\lambda$ to be independent *a priori*. It is easily seen that the non-informative prior distribution for $\tau$, $\rho$ and $\lambda$ is proportional to a constant since all of the elements of $\tau$ and $\lambda$ are just real and $\rho$ is confined within $(-1, 1)$. For $\sigma^2$, we choose $\sigma^{-2}$ as its prior for $\sigma^2 > 0$. For $\Gamma$, two

possibilities are considered. 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 variance $\sigma^2$, so every element of $\Gamma$ should be within a finite interval such that neither random effects nor the AR(1) errors dominate the others. Utilizing the 'principle of stable estimation' suggested by Edwards *et al.* (1963), a uniform prior is appropriate for $\Gamma$. Therefore, the prior for $\tau$, $\sigma^2$, $\Gamma$, $\rho$ and $\lambda$ can be written as:

$$\pi(\tau, \sigma^2, \Gamma, \rho, \lambda) \propto \pi(\tau)\pi(\sigma^2)\pi(\Gamma)\pi(\rho)\pi(\lambda) \propto \sigma^{-2} \tag{4}$$

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

We shall refer to the priors discussed above as prior 1 and prior 2 in the rest of this paper. In the following Bayesian inference, we shall denote the prior of $\Gamma$, $\rho$ and $\lambda$ as $\pi(\Gamma, \rho, \lambda)$ without specifying which prior is being used.



## 3 Parameter estimation

Combining the likelihood with the prior and integrating w.r.t. $\tau$ and $\sigma^2$, we obtain

$$P(\Gamma, \rho, \lambda \mid Y)$$

$$\propto \pi(\Gamma, \rho, \lambda) \left( \prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Lambda_{ij}|^{-1/2} \right) |\mathcal{J}| \tag{5}$$

$$\left( \prod_{i=1}^{r} \left| \sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} X_{ij} \right|^{-1/2} \right) B^{-(n - \sum_{i=1}^{r} m_i)/2}$$

where

$$Y = (Y_{11}, \ldots, Y_{1N_1}, Y_{21}, \ldots, Y_{2N_2}, \ldots, Y_{r1}, \ldots, Y_{rN_r})$$

$$B = \sum_{i=1}^{r} \sum_{j=1}^{N_i} (Y_{ij}^{(\lambda_i)} - X_{ij}\hat{\tau}_i)' \Lambda_{ij}^{-1} (Y_{ij}^{(\lambda_i)} - X_{ij}\hat{\tau}_i) \tag{6}$$

$$\hat{\tau}_i = \left( \sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} X_{ij} \right)^{-1} \left( \sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} Y_{ij}^{(\lambda_i)} \right), \quad i = 1, 2, \ldots, r$$

$$\mathcal{J} = \begin{cases} \prod_{i=1}^{r} \prod_{j=1}^{N_i} \prod_{k=1}^{P_{ij}} (Y_{ijk} + Y)^{\lambda_i - 1} & \text{if } \lambda_i \neq 0 \\ \prod_{i=1}^{Y} \prod_{j=1}^{N_i} \prod_{k=1}^{P_{ij}} (Y_{ijk} + Y)^{-1} & \text{if } \lambda_i = 0 \end{cases}$$

By applying Ljung & Box (1980), we have the following approximate posterior distribution and region of $\tau_i$ for $i = 1, 2, \ldots, r$:

$$\tau_i | Y \sim T_{m_i}(\hat{\tau}_i^{\star}, \hat{B}(\nu \sum_{j=1}^{N_i} X_{ij}' \hat{\Lambda}_{ij}^{-1} X_{ij})^{-1}, \nu) \tag{7}$$

$$\hat{B}^{-1}(\tau_i - \hat{\tau}_i^{\star})' \left( \sum_{j=1}^{N_i} X_{ij}' \hat{\Lambda}_{ij}^{-1} X_{ij} \right) (\tau_i - \hat{\tau}_i^{\star}) \leqslant \frac{m_i}{v} F_{1-\alpha}(m_i, v) \tag{8}$$

where

$$\hat{\tau}_i^{\star} = \left( \sum_{j=1}^{N_i} X_{ij}' \hat{\Lambda}_{ij}^{-1} X_{ij} \right)^{-1} \left( \sum_{j=1}^{N_i} X_{ij}' \hat{\Lambda}_{ij}^{-1} Y_{ij}^{(\hat{\lambda}_i)} \right)$$

$$\hat{B} = \sum_{i=1}^{r} \sum_{j=1}^{N_i} (Y_{ij}^{(\hat{\lambda}_i)} - X_{ij} \hat{\tau}_i^{\star})' \hat{\Lambda}_{ij}^{-1} (Y_{ij}^{(\hat{\lambda}_i)} - X_{ij} \hat{\tau}_i^{\star}) \tag{9}$$

$$\hat{\Lambda}_{ij} = D_{ij} \hat{\Gamma}_i D_{ij}' + \hat{C}_{ij}, \quad \hat{C}_{ij} = (\hat{\rho}^{|a-b|}), \text{ for } a, b = 1, 2, \ldots, p_{ij}$$

$$\nu = n - \sum_{k=1}^{r} m_k$$

$\hat{\Gamma}$, $\hat{\rho}$ and $\hat{\lambda}$ maximize $P(\Gamma, \rho, \lambda | Y)$, as given in equation (5), and $T_p(\mu, \Sigma, n)$ is the $p$-variate $T$ distribution with mean $\mu$ and covariance matrix $n(n-2)^{-1}\Sigma$.

With similar arguments as above, the posterior distribution of $\sigma^2$ may be approximated as $IG((n - \Sigma_{i=1}^{r} m_i)/2, \hat{B}/2)$, where $\hat{B}$ is given in equation (9) and $IG(\nu_1, \nu_2)$ is the inverse gamma distribution with parameters $\nu_1$ and $\nu_2$.

## 4 Prediction of partially observed future values

The prediction of $y_{lk}$, future $q$-dimensional values of measurement $Y_{lk}$, will be considered in this section. To accomplish this purpose, the covariance structure must be assumed to be extendible to the future values of all observed individuals.

Let $x$ and $d$, respectively, be the $q \times m_i$ and $q \times w_i$ matrices corresponding to $y_{lk}$. Then we have: $E(Y_{lk}^{(\lambda_l)'}, y_{lk}^{(\lambda_l)'})' = (X_{lk}', x')' \tau_l$, and $\mathrm{Cov}(Y_{lk}^{(\lambda_l)'}, y_{lk}^{(\lambda_l)'})' = \sigma^2(D^{\star} \Gamma_l D^{\star\prime} + C^{\star}) \equiv \sigma^2 \Lambda^{\star}$, where $\Lambda^{\star} = (\Lambda_{ij}^{\star})$, $i, j = 1, 2$, $D^{\star} = (D_{lk}', d')'$ and $C^{\star} = (\rho^{|a-b|})$, $a, b = 1, \ldots, p_{lk} + q$. Let $Y^{\star} = (Y_{lk}', y')'$, $X^{\star} = (X_{lk}', x')'$ and $Y^{\star(\lambda_l)} = (Y_{lk}^{(\lambda_l)'}, y^{(\lambda_l)'})'$. The conditional density function of $y_{lk}^{(\lambda_l)}$ given $Y_{lk}$, $\tau_l$, $\sigma^2$, $\Gamma_l$, $\rho$ and $\lambda_l$ is:

$$\begin{aligned} & f(y_{lk}^{(\lambda_l)} | \tau_l, \sigma^2, \Gamma_l, \rho, \lambda_l, Y_{lk}) \\ & \propto (\sigma^2)^{-q/2} |\Lambda_{22.1}^{\star}|^{-1/2} \exp\left[ -\frac{1}{2\sigma^2} (y_{lk}^{(\lambda_l)} - \mu_{2.1})' \Lambda_{22.1}^{\star -1} (y_{lk}^{(\lambda_l)} - \mu_{2.1}) \right] \end{aligned} \tag{10}$$

where

$$\begin{aligned} \Lambda_{22.1}^{\star} &= \Lambda_{22}^{\star} - \Lambda_{21}^{\star} \Lambda_{11}^{\star -1} \Lambda_{12}^{\star} \\ \mu_{2.1} &= x\tau_l + \Lambda_{21}^{\star} \Lambda_{11}^{\star -1} (Y_{lk}^{(\lambda_l)} - X_{lk} \tau_l) \end{aligned} \tag{11}$$

Upon combining (10) with the joint posterior of $\tau$, $\sigma^2$, $\Gamma$, $\rho$, $\lambda$ and integrating w.r.t. $\sigma^2$, $\tau$ and $y_{lk}^{(\lambda_l)}$ we have the posterior density of $\Gamma$, $\rho$, $\lambda$:

$$P(\Gamma, \rho, \lambda | Y)$$

$$\propto \pi(\Gamma, \rho, \lambda) \Lambda_{22.1}^{\star} |^{-1/2} \left( \prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Lambda_{ij}|^{-1/2} \right) |\mathcal{J}| Q |^{-1/2} |G_{22}|^{-1/2} \tag{12}$$

$$\left( \prod_{i \neq k} \left| \sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} X_{ij} \right|^{-1/2} \right) [B_1 + B_2 + B_3]^{-(n - \Sigma_{i=1}^{r} m_i)/2}$$

where

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

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

$$B_3 = (Y_{lk}^{(\lambda_l)} - X_{lk} \hat{\tau}_{l1})' G_{11.2} (Y_{lk}^{(\lambda_l)} - X_{lk} \hat{\tau}_{21})$$

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

$$\hat{\tau}_{l1} = Q_1^{-1} \left( \sum_{j \neq k} X_{lj}' \Lambda_{lj}^{-1} Y_{lj}^{(\lambda_l)} \right)$$

$$Q = Q_1 + Q_2, \quad Q_1 = \sum_{j \neq k} X_{lj}' \Lambda_{lj}^{-1} X_{lj}, \quad Q_2 = X^{\star\prime} \Lambda^{\star -1} X^{\star}$$

$$G = \Lambda^{\star -1} X^{\star} Q_2^{-1} Q_1 Q^{-1} X^{\star\prime} \Lambda^{\star -1} + Z^{\star} (Z^{\star\prime} \Lambda^{\star} Z^{\star})^{-1} Z^{\star\prime} = (G_{ij}), \quad i, j = 1, 2$$

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

$Z^{\star}$ : $(p_{lk} + q) \times (p_{lk} + q - m_l)$ matrix such that $X^{\star\prime} Z^{\star} = 0$

With arguments similar to those in (7), we also have the following approximate predictive density for $y_{lk}^{(\lambda_l)}$:

$$F(y_{lk}^{(\lambda_l)} | Y) \doteq T_q(\hat{\mu}, (\hat{B}_1 + \hat{B}_2 + \hat{B}_3)(\nu \hat{G}_{22})^{-1}, \nu) \tag{13}$$

where $(\hat{\Gamma}, \hat{\rho}, \hat{\lambda})$ is the mode of $P(\Gamma, \rho, \lambda | Y)$, as in (12), and the quantities $\hat{B}_1$, $\hat{B}_2$, $\hat{B}_3$ and $\hat{G}$ are, respectively, $B_1$, $B_2$, $B_3$ and $G$ evaluated at $(\hat{\Gamma}, \hat{\rho}, \hat{\lambda})$, $\nu = n - \Sigma_{i=1}^{r} m_i$ and $\hat{\mu} = x \hat{\tau}_{l1} - \hat{G}_{22}^{-1} \hat{G}_{21} (Y_{lk}^{(\lambda_l)} - X_{lk} \hat{\tau}_{l1})$.

Therefore, we can predict $y_{lk}$ by the following approximate predictor:

$$\hat{y}_{lkh} = \begin{cases} (1 + \hat{\lambda}_l \hat{\mu}_h)^{1/\lambda_l} - \gamma, & \text{if } \hat{\lambda}_l \neq 0, \\ \exp(\hat{\mu}_h) - \gamma, & \text{if } \hat{\lambda}_l = 0, \end{cases} \quad \text{for } h = 1, 2, \ldots, q \tag{14}$$

where $\hat{y}_{lkh}$ is the $h$th component of $\hat{y}_{lk}$ and $\hat{\mu}_h$ is the $h$th component of $\hat{\mu}$.

A $(1 - \alpha)$ predictive region for $y_{lk}$ can be constructed through the following inequality:

$$(y_{lk}^{(\hat{\lambda}_l)} - \hat{\mu})' \hat{G}_{22} (y_{lk}^{(\hat{\lambda}_l)} - \hat{\mu}) \leqslant \text{const} \tag{15}$$

where

$$\text{const} = [\hat{B}_1 + \hat{B}_2 + \hat{B}_3] \frac{q}{n - \sum_{i=1}^{r} m_i} F_{1-\alpha}\left(q, n - \sum_{i=1}^{r} m_i\right)$$

## 5 Bayesian inference via Markov chain Monte Carlo methods

In order to make inference about model parameters and to make predictions, we need to integrate over high-dimensional probability distributions. Markov chain Monte Carlo (MCMC) methods are very helpful for solving our problems. MCMC is essentially Monte Carlo integration using Markov chains. It draws samples from the required distribution by running a cleverly constructed Markov chain for a long time and then forms sample averages to approximate expectations. The Gibbs sampler and the Metropolis–Hastings (M-H) algorithm are well-known among the several ways of constructing those chains. A great advantage of the Gibbs sampler and the M-H algorithm is the ease of implementation which makes heavy use of modern computational capabilities. Excellent references on the methodology have been provided by Gelfand & Smith (1990), Casella & George (1992), Gelman & Rubin (1992), Chib & Greenberg (1995) and Gilks *et al.* (1996). The Gibbs sampler is used in conjunction with the M-H algorithm to make inference and to make predictions. The strategy is described in this section.



### *5.1 Model and algorithm*

The joint posterior density of $\tau$, $\sigma^2$, $\Gamma$, $\rho$ and $\lambda$, given the sample, $Y$, can be obtained easily from combining the likelihood with the prior. The MCMC algorithm proceeds as follows.

*Step 1.* Generate $\tau_i$ given $\sigma^2$, $\Gamma$, $\rho$, $\lambda$ and $Y$ from

$$N_{m_i}\left(\hat{\tau}_i, \sigma^2\left(\sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} X_{ij}\right)^{-1}\right) \tag{16}$$

where $\hat{\tau}_i$ is given in equation (6).

*Step 2.* Generate $\sigma^2$ given $\tau$, $\Gamma$, $\rho$, $\lambda$ and $Y$ from the inverse gamma distribution

$$IG\left(\frac{n}{2}, \frac{S(\tau, \Gamma, \rho, \lambda)}{2}\right) \tag{17}$$

where $S(\tau, \Gamma, \rho, \lambda) = \sum_{i=1}^{r} \sum_{j=1}^{N_i} (Y_{ij}^{(\lambda_i)} - X_{ij}\tau_i)' \Lambda_{ij}^{-1} (Y_{ij}^{(\lambda_i)} - X_{ij}\tau_i)$ and $n = \sum_{i=1}^{r} \sum_{j=1}^{N_i} p_{ij}$.

*Step 3.* Generate $\Gamma$ given $\tau$, $\sigma^2$, $\rho$, $\lambda$ and $Y$ using the M-H algorithm, where

$$f(\Gamma) \propto \pi(\Gamma, \rho, \lambda) \left(\prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Lambda_{ij}(\Gamma, \rho)|^{-1/2}\right) \exp\left[-\frac{1}{2\sigma^2} S(\tau, \Gamma, \rho, \lambda)\right] \tag{18}$$

and $\Lambda_{ij}(\Gamma, \rho) = D_{ij} \Gamma_i D_{ij}' + C_{ij}$.

*Step 4.* Generate $\rho$ given $\tau$, $\sigma^2$, $\Gamma$, $\lambda$ and $Y$ using the M-H algorithm, where

$$f(\rho) \propto \pi(\Gamma, \rho, \lambda) \left( \prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Lambda_{ij}(\Gamma, \rho)|^{-1/2} \right) \exp\left[ -\frac{1}{2\sigma^2} S(\tau, \Gamma, \rho, \lambda) \right] \tag{19}$$

*Step 5.* Generate $\lambda$ given $\tau$, $\sigma^2$, $\Gamma$, $\rho$ and $Y$ using the M-H algorithm, where

$$f(\lambda) \propto \pi(\Gamma, \rho, \lambda) |\mathcal{J}(\lambda)| \exp\left[ -\frac{1}{2\sigma^2} S(\tau, \Gamma, \rho, \lambda) \right] \tag{20}$$

$|\mathcal{J}(\lambda)| = \prod_{i=1}^{r} |\mathcal{J}_i(\lambda_i)|$ and

$$|\mathcal{J}_i(\lambda_i)| = \begin{cases} \prod_{j=1}^{N_i} \prod_{k=1}^{p_{ij}} |Y_{ijk} + Y|^{\lambda_i - 1}, & \text{if } \lambda_i \neq 0 \\ \prod_{j=1}^{N_i} \prod_{k=1}^{p_{ij}} |Y_{ijk} + Y|^{-1}, & \text{if } \lambda_i = 0 \end{cases}, \qquad \text{for } i = 1, 2, \ldots, r$$

To elaborate on Steps 3, 4 and 5 of the above algorithm, we choose the Wishart distribution, $W((\Gamma_i^{(k)}/N_i - 1), N_i - 1)$, as the proposal distribution, $q(\Gamma_i^{(k+1)} | \Gamma_i^{(k)})$, in Step 3. As regards $\rho$, we first transform $\rho$ into $\rho^\star = \log(1 + \rho)/(1 - \rho)$ and then apply the M-H algorithm to the following function:

$$g(\rho^\star | \tau, \sigma^2, \Gamma, \lambda, Y) \propto \pi(\Gamma, \rho^\star, \lambda) \left( \prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Lambda_{ij}(\Gamma, \rho^\star)|^{-1/2} \right) \exp\left[ -\frac{1}{2\sigma^2} S(\tau, \Gamma, \rho^\star, \lambda) \right] |\mathcal{J}(\rho^\star)| \tag{21}$$

where $|\mathcal{J}(\rho^\star)| = (2e^{\rho^\star}/(1 + e^{\rho^\star})^2)$ and $S(\tau, \Gamma, \rho^\star, \lambda)$ is obtained from $S(\tau, \Gamma, \rho, \lambda)$ with $\rho$ replaced by $(e^{\rho^\star} - 1)/(e^{\rho^\star} + 1)$. Two normal distributions with means $\rho^{\star(k)}$, $\lambda_i^{(k)}$ and variances $\sigma^2_{\rho^{\star(k)}}$, $\sigma^2_{\lambda_i^{(k)}}$ are chosen as the proposal distributions, $q(\rho^{\star(k+1)} | \rho^{\star(k)})$ and $q(\lambda_i^{(k+1)} | \lambda_i^{(k)})$, respectively. The values of $\sigma_{\rho^{\star(k)}}$ and $\sigma_{\lambda_i^{(k)}}$ are usually chosen to reflect the conditional standard deviations of $\rho^\star$ given $\tau$, $\sigma^2$, $\Gamma$, $\lambda$ and $Y$, and $\lambda_i$ given $\tau$, $\sigma^2$, $\Gamma$, $\rho^\star$ and $Y$, respectively. Thus, we can estimate the variances $\sigma^2_{\rho^{\star(k)}}$ and $\sigma^2_{\lambda_i^{(k)}}$ by the following method. Let $l(\rho^\star; \tau, \sigma^2, \Gamma, \lambda, Y) = \log(g(\rho^\star | \tau, \sigma^2, \Gamma, \lambda, Y))$. At the $(k+1)$th iteration, the preliminary variance estimate of $\rho^\star$, $\hat{\sigma}^2_{\rho^\star}$, would be the inverted sample information of $l(\rho^\star; \tau, \sigma^2, \Gamma, \lambda, Y)$ given the $\rho^{\star(k)}$ value in the MCMC algorithm and likewise for $\sigma^2_{\lambda_i^{(k)}}$. After obtaining $\rho^\star$, we transform $\rho^\star$ back to $\rho$ by $(e^{\rho^\star} - 1)/(e^{\rho^\star} + 1)$.

It is worth noting that for the fatigue data to be illustrated later, all observations are considered in the same group and to have a random effect in slope, i.e. $r = 1$ and $w = 1$. In Step 3, we first decompose $\Gamma$ into $f^2$ for practical calculation and then apply the M-H algorithm to the following function:

$$g(f | \tau, \sigma^2, \rho^\star, \lambda, Y) \propto \pi(f, \rho^\star, \lambda) \left( \prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Lambda_{ij}(f, \rho^\star)|^{-1/2} \right) \exp\left[ -\frac{1}{2\sigma^2} S(\tau, f, \rho^\star, \lambda) \right] |\mathcal{J}(f)| \tag{22}$$

where $|\mathcal{J}(f)| = |2f|$ and $S(\tau, f, \rho^\star, \lambda)$ is obtained from $S(\tau, \Gamma, \rho, \lambda)$ with $\Gamma$ and $\rho$

replaced by $f^2$ and $(e^{\rho^\star} - 1)/(e^{\rho^\star} + 1)$, respectively. The proposal distribution $q(f^{(k+1)} | f^{(k)})$, is taken as a mixture of two normal distributions, $N(f^{(k)}, \sigma^2_{f^{(k)}})$ and $N(-f^{(k)}, \sigma^2_{f^{(k)}})$. The same operation as above can be applied to specify the values of $\sigma^2_{f^{(k)}}$. We transform $f$ back to $\Gamma$ by $f^2$ after obtaining $f$.

So far, we have considered running only one chain, but multiple chains are permissible. In fact, comparing several seemingly converged chains might reveal genuine differences if the chains have not yet approached stationarity, and so it is reasonable to simulate multiple chains. Gelman & Rubin (1992) suggested using 'overdispersed' starting values in multiple chains to assist in assessing convergence. Overdispersed starting points are an important design feature because starting far apart can make the lack of convergence apparent and, for the purpose of inference, ensure that all major regions of the target distribution are represented in the simulations.

After sufficiently long *burn-in* iterations, we can use the remaining samples to estimate the functional of the parameters in which we are interested.



### *5.2 Forecast*

Now we implement the above algorithm with $R$ chains and $M$ iterations for each chain. Once a $R \times (M - K)$ array of values is obtained from the MCMC runs, dropping the $R \times K$ burn-in samples out, the functional, such as estimators and percentiles of the future values, can be computed.

Assuming that $\gamma = 0$, we have $y_{lk}^{(\lambda_l)} = \mu_{2.1} + \xi$ for given $Y_{lk}$, where $\xi \sim N_q(0, \sigma^2 \Lambda^\star_{22.1})$, and from equation (2),

$$y_{lkh} = \begin{cases} (\lambda_l y_{lkh}^{(\lambda_l)} + 1)^{1/\lambda_l}, & \text{if } \lambda_l \neq 0 \\ \exp(y_{lkh}^{(\lambda_l)}), & \text{if } \lambda_l = 0 \end{cases}, \text{ for } h = 1, 2, \ldots, q \tag{23}$$

where $y_{lkh}$ is the $h$th component of $y_{lk}$. To obtain the prediction of the future values, $y_{lk}$, we consider two cases:

**Case 1.** $\lambda_l \neq 0$,

$$y_{lkh} = [\lambda_l(\mu_{2.1.h} + \xi_h) + 1]^{1/\lambda_l} \tag{24}$$

**Case 2.** $\lambda_l = 0$,

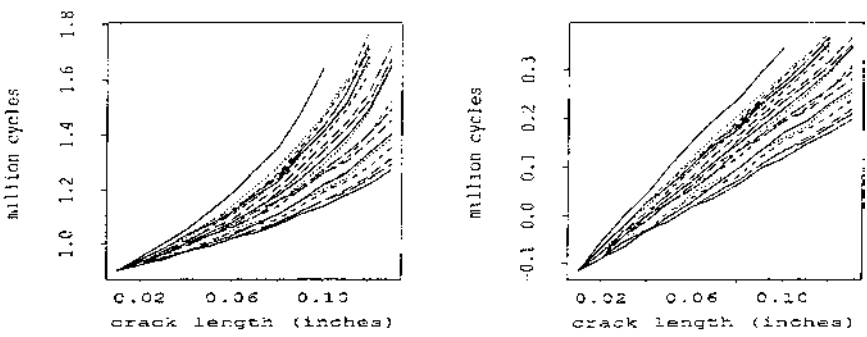
$$y_{lkh} = \exp(\mu_{2.1.h} + \xi_h) \tag{25}$$

where $y_{lkh}$, $\mu_{2.1.h}$ and $\xi_h$ are the $h$th component of $y_{lk}$, $\mu_{2.1}$ and $\xi$, respectively, and $\mu_{2.1}$ is given in equation (11). Let $\mu_{2.1}^{(s,t)}$ and $\Lambda_{22.1}^{\star(s,t)}$ denote the functional (11) evaluated at the $s$th chain and $t$th replication of the MCMC runs and $\xi^{(s,t)}$ is generated from $N_q(0, \sigma^{2(s,t)} \Lambda_{22.1}^{\star(s,t)})$, then we can get the predicted value of $y_{lk}$ by computing the functional (24) and (25) with $\lambda_l$, $\mu_{2.1.h}$ and $\xi_h$ replaced by $\lambda_l^{(s,t)}$, $\mu_{2.1.h}^{(s,t)}$ and $\xi_h^{(s,t)}$. It means that we predict $y_{lk}$ by

$$\hat{y}_{lk} = \frac{1}{R} \sum_{s=1}^{R} y_{lk}^{(s)} \tag{26}$$

TABLE 1. Fatigue crack growth data from Bogdanoff & Kozin (1985) (million cycles)

| Path | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.10 | 0.11 | 0.12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.9 | 0.95 | 1.00 | 1.05 | 1.12 | 1.19 | 1.27 | 1.35 | 1.48 | 1.64 | | | |
| 2 | 0.9 | 0.94 | 0.98 | 1.03 | 1.08 | 1.14 | 1.21 | 1.28 | 1.37 | 1.47 | 1.60 | | |
| 3 | 0.9 | 0.94 | 0.98 | 1.03 | 1.08 | 1.13 | 1.19 | 1.26 | 1.35 | 1.46 | 1.58 | 1.77 | |
| 4 | 0.9 | 0.94 | 0.98 | 1.03 | 1.07 | 1.12 | 1.19 | 1.25 | 1.34 | 1.43 | 1.55 | 1.73 | |
| 5 | 0.9 | 0.94 | 0.98 | 1.03 | 1.07 | 1.12 | 1.19 | 1.24 | 1.34 | 1.43 | 1.55 | 1.71 | |
| 6 | 0.9 | 0.94 | 0.98 | 1.03 | 1.07 | 1.12 | 1.18 | 1.23 | 1.33 | 1.41 | 1.31 | 1.68 | |
| 7 | 0.9 | 0.94 | 0.98 | 1.02 | 1.07 | 1.11 | 1.17 | 1.23 | 1.32 | 1.41 | 1.52 | 1.66 | |
| 8 | 0.9 | 0.93 | 0.97 | 1.00 | 1.06 | 1.11 | 1.17 | 1.23 | 1.30 | 1.39 | 1.49 | 1.62 | |
| 9 | 0.9 | 0.92 | 0.97 | 1.01 | 1.05 | 1.09 | 1.15 | 1.21 | 1.28 | 1.36 | 1.44 | 1.55 | 1.72 |
| 10 | 0.9 | 0.92 | 0.96 | 1.00 | 1.04 | 1.08 | 1.13 | 1,19 | 1.26 | 1.34 | 1.42 | 1.52 | 1.67 |
| 11 | 0.9 | 0.93 | 0.96 | 1.00 | 1.04 | 1.08 | 1.13 | 1.18 | 1.24 | 1.31 | 1.39 | 1.49 | 1.65 |
| 12 | 0.9 | 0.93 | 0.97 | 1.00 | 1.03 | 1.07 | 1.10 | 1.16 | 1.22 | 1.29 | 1.37 | 1.48 | 1.64 |
| 13 | 0.9 | 0.92 | 0.97 | 0.99 | 1.03 | 1.06 | 1.10 | 1.14 | 1.20 | 1.26 | 1.31 | 1.40 | 1.52 |
| 14 | 0.9 | 0.93 | 0.96 | 1.00 | 1.03 | 1.07 | 1.12 | 1.16 | 1.20 | 1.26 | 1.30 | 1.37 | 1.45 |
| 15 | 0.9 | 0.92 | 0.96 | 0.99 | 1.03 | 1.06 | 1.10 | 1.16 | 1.21 | 1.27 | 1.33 | 1.40 | 1.49 |
| 16 | 0.9 | 0.92 | 0.95 | 0.97 | 1.00 | 1.03 | 1.07 | 1.11 | 1.16 | 1.22 | 1.26 | 1.33 | 1.40 |
| 17 | 0.9 | 0.93 | 0.96 | 0.97 | 1.00 | 1.05 | 1.08 | 1.11 | 1.16 | 1.20 | 1.24 | 1.32 | 1.38 |
| 18 | 0.9 | 0.92 | 0.94 | 0.97 | 1.01 | 1.04 | 1.07 | 1.09 | 1.14 | 1.19 | 1.23 | 1.28 | 1.35 |
| 19 | 0.9 | 0.92 | 0.94 | 0.97 | 0.99 | 1.02 | 1.05 | 1.08 | 1.12 | 1.16 | 1.20 | 1.25 | 1.31 |
| 20 | 0.9 | 0.92 | 0.94 | 0.97 | 0.99 | 1.02 | 1.05 | 1.08 | 1.12 | 1.16 | 1.19 | 1.24 | 1.29 |
| 21 | 0.9 | 0.92 | 0.94 | 0.97 | 0.99 | 1.02 | 1.04 | 1.07 | 1.11 | 1.14 | 1.18 | 1.22 | 1.27 |




(a) original fatigue crack data

(b) data after taking power transformation with $\lambda = -1.59$

FIG. 1. Plots of fatigue crack data.

where

$$y_{lk}^{(s)} = \frac{1}{M-K} \sum_{t=K+1}^{M} y_{lk}^{(s,t)} \tag{27}$$

and $y_{lk}^{(s,t)}$ consists of $y_{lkh}^{(s,t)}$, $h = 1, 2, \ldots, q$.

## 6 Numerical illustration

In this section, we apply the results developed in Sections 3 and 4 to the fatigue crack growth data from Bogdanoff & Kozin (1985), which are unbalanced, as listed in Table 1 and plotted in Fig. 1(a). All observations are considered in the same group, i.e. $r = 1$, and to have a random effect on the slope.

When implementing the MCMC, the convergence is monitored by examining the measure $\sqrt{\hat{R}}$ proposed by Gelman & Rubin (1992). Before running the MCMC, we locate the mode of the target distribution first and estimate roughly the scale of the target distribution near the mode. Starting points are chosen from a random perturbation around the mode so that they are more variable than the target distribution. There are seven independently simulated chains running in parallel, simultaneously on two or more workstations in a network with seven different initial values. We ran the simulations until the values of $\sqrt{\hat{R}}$ are all less than 1.1 or 1.2. Once the convergence is judged to be adequate, the functionals of the parameters are summarized based on the last 1000 simulated values of each chain.

### *6.1 Parameter estimation*

The estimates of parameters obtained via the approximate Bayesian approach are listed in Table 2, and those summarized by the MCMC methods, including the sample percentiles, means and standard deviations, are listed in Tables 3 and 4. It can be seen that all the methods considered here give similar estimates for



TABLE 2. Comparison of parameter estimations

| | $\beta_1$ | $\beta_2$ | $\sigma^2$ | $\Gamma$ | $\rho$ | $\lambda$ |
|---|---|---|---|---|---|---|
| prior 1 | −0.15018 | 0.03695 | 0.00004 | 0.96769 | 0.52716 | −1.58983 |
| prior 2 | −0.15019 | 0.03695 | 0.00004 | 0.89508 | 0.53259 | −1.59006 |
| MLE | −0.15014 | 0.03695 | 0.00004 | 0.93614 | 0.51964 | −1.59054 |

TABLE 3. Summaries of MCMC posterior disributions (prior 1)

| | $\beta_1$ | $\beta_2$ | $\sigma^2$ | $\Gamma$ | $\rho$ | $\lambda$ |
|---|---|---|---|---|---|---|
| 2.5% | −0.15251 | 0.03388 | 0.00003 | 0.61185 | 0.35862 | −1.67762 |
| 5% | −0.15215 | 0.03443 | 0.00003 | 0.66837 | 0.38207 | −1.66292 |
| 25% | −0.15083 | 0.03588 | 0.00003 | 0.95903 | 0.44540 | −1.62080 |
| 50% | −0.15001 | 0.03686 | 0.00004 | 1.21613 | 0.48740 | −1.59449 |
| 75% | −0.14916 | 0.03782 | 0.00004 | 1.54509 | 0.53367 | −1.57355 |
| 95% | −0.14799 | 0.03927 | 0.00005 | 2.32800 | 0.59409 | −1.53184 |
| 97.5% | −0.14760 | 0.03986 | 0.00005 | 2.63815 | 0.61459 | −1.52099 |
| Mean | −0.15002 | 0.03685 | 0.00004 | 1.31905 | 0.48851 | −1.59593 |
| SD | 0.00125 | 0.00150 | 0.00000 | 0.52202 | 0.06506 | 0.03928 |

TABLE 4. Summaries of MCMC posterior distributions (prior 2)

| | $\beta_1$ | $\beta_2$ | $\sigma^2$ | $\Gamma$ | $\rho$ | $\lambda$ |
|---|---|---|---|---|---|---|
| 2.5% | −0.15273 | 0.03445 | 0.00003 | 0.46298 | 0.36717 | −1.68643 |
| 5% | −0.15229 | 0.03484 | 0.00003 | 0.50546 | 0.39960 | −1.66881 |
| 25% | −0.15095 | 0.03604 | 0.00004 | 0.69612 | 0.46892 | −1.61635 |
| 50% | −0.15009 | 0.03690 | 0.00004 | 083702 | 0.51382 | −1.59331 |
| 75% | −0.14921 | 0.03777 | 0.00004 | 1.04686 | 0.55954 | −1.56734 |
| 95% | −0.14800 | 0.03901 | 0.00005 | 1.46374 | 0.62202 | −1.52314 |
| 97.5% | −0.14755 | 0.03945 | 0.00006 | 1.62054 | 0.63824 | −1.50796 |
| Mean | −0.15010 | 0.03690 | 0.00004 | 0.89818 | 0.51220 | −1.59367 |
| SD | 0.00131 | 0.00129 | 0.00001 | 0.29941 | 0.06710 | 0.04243 |

TABLE 5. Comparison of the various approximate percentiles of the marginal density of $\beta_1$

| | prior 1 | | prior 2 | |
|---|---|---|---|---|
| | Bayesian | MCMC | Bayesian | MCMC |
| 2.5% | −0.15257 | −0.15251 | −0.15261 | −0.15273 |
| 5% | −0.15218 | −0.15215 | −0.15221 | −0.15229 |
| 25% | −0.15100 | −0.15083 | −0.15102 | −0.15095 |
| 50% | −0.15018 | −0.15001 | −0.15019 | −0.15009 |
| 75% | −0.14936 | −0.14916 | −0.14936 | −0.14921 |
| 95% | −0.14818 | −0.14799 | −0.14816 | −0.14800 |
| 97.5% | −0.14779 | −0.14760 | −0.14777 | −0.14755 |

TABLE 6. Comparison of the various approximate percentiles of the marginal density of $\beta_2$

| | prior 1 | | prior 2 | |
|---|---|---|---|---|
| | Bayesian | MCMC | Bayesian | MCMC |
| 2.5% | 0.03427 | 0.03388 | 0.03436 | 0.03445 |
| 5% | 0.03471 | 0.03443 | 0.03478 | 0.03484 |
| 25% | 0.03603 | 0.03588 | 0.03606 | 0.03604 |
| 50% | 0.03695 | 0.03686 | 0.03695 | 0.03690 |
| 75% | 0.03787 | 0.03782 | 0.03784 | 0.03777 |
| 95% | 0.03920 | 0.03927 | 0.03913 | 0.03901 |
| 97.5% | 0.03963 | 0.03986 | 0.03955 | 0.03945 |



$\tau = (\beta_1, \beta_2)'$, $\sigma^2$, $\rho$, and $\lambda$ but the estimates using different priors are somewhat different for the other components of the covariance structure. Prior 1 gives estimates close to those by the ML method and puts more weight on the variation from random effects than prior 2.

Figure 1(b) is the plot of the data after taking the power transformation with $\lambda$ being $-1.59$, and it reveals that the linearity assumption is satisfied after applying the adequate power transformation on observations. It also exhibits an obvious random effect in slope. For $\tau$, the comparison for the various percentiles of the marginal densities of $\beta_1$ and $\beta_2$, obtained by the approximate Bayesian method, with those estimates from the MCMC methods, is given in Tables 5 and 6, and Fig. 2. They all show that the two different approximate marginal densities of $\beta_1$ and $\beta_2$ obtained via the approximate Bayesian approach and the MCMC methods are nearly the same. The 95% posterior regions for $\tau$ constructed by the Bayesian approach derived in this paper and the corresponding confidence region by the ML method are plotted in Fig. 3. The posterior region constructed using prior 1 is slightly larger than that by the ML method and the region constructed using prior 2 is the largest. Table 7 is a comparison of the coverage probability for $\tau$ in a balanced case in which $r = 1$ and $p_{ij} = 10$. It is clear that the coverage probability by each method considered here becomes closer to 0.95 as the sample size $N$ increases. Generally speaking, prior 1 performs much better than prior 2 and the ML method.

### *6.2 Prediction of future values*

We will consider both balanced and unbalanced cases. Two types of prediction problems are considered here for the balanced subset of fatigue crack data with

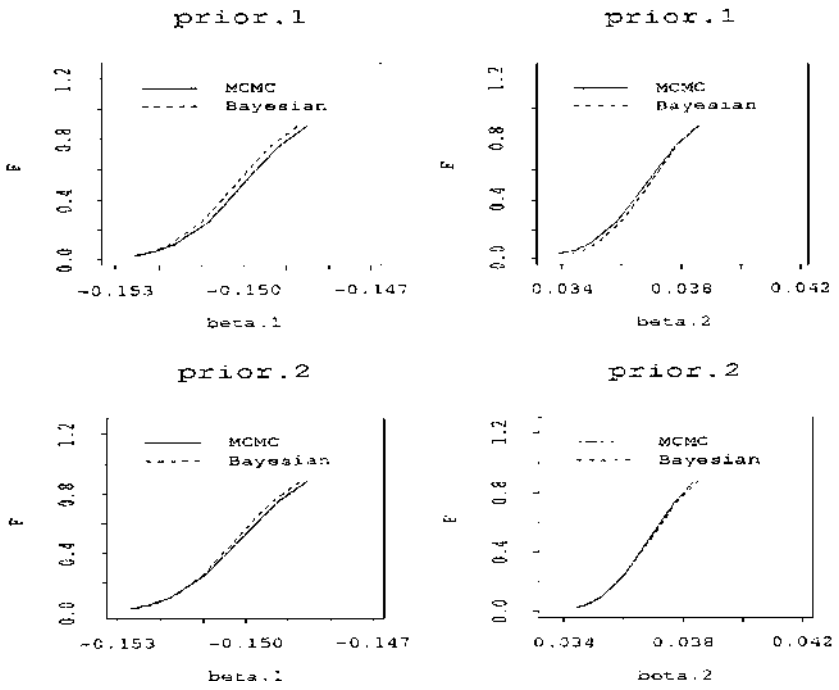



FIG. 2. Graphical comparison of Tables 5 and 6.

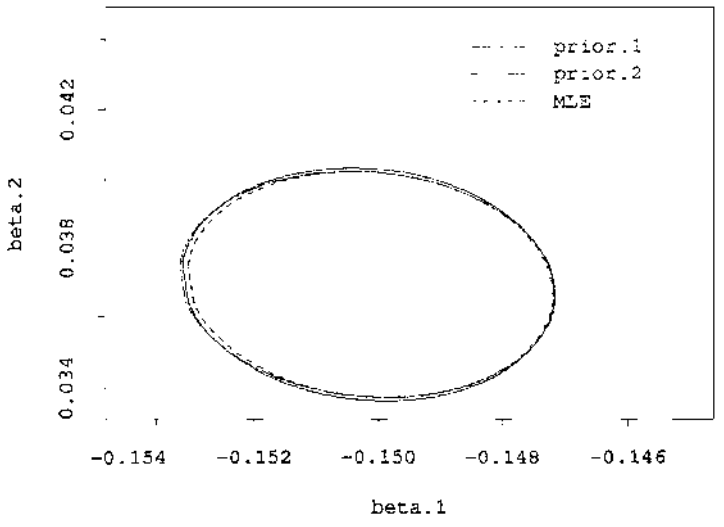



FIG. 3. 95% posterior and confidence regions for $\tau = (\beta_1, \beta_2)'$.

$p_{ij} = 10$: conditional prediction and extended prediction. The goal of both problems is to predict $y_{ip}$, $i \in 1, 2, \ldots, N$, and we use two criteria, the mean absolute deviation (MAD) and the mean absolute relative deviation (MARD) of the predictions from the actuals, as the measure of the predictive accuracy to compare the performances of different prediction methods.

TABLE 7. Comparison of coverage probabilities ($1-\alpha=0.95$)

| | Prior 1 | Prior 2 | MLE |
|---|---|---|---|
| $N=10$ | 0.9165 | 0.8780 | 0.8960 |
| $N=20$ | 0.9275 | 0.9100 | 0.9150 |
| $N=30$ | 0.9315 | 0.9160 | 0.9225 |

The presumed parameters are given as:

$$\tau=\begin{pmatrix}-0.15\\0.0373\end{pmatrix},\quad \sigma^2=0.00003,\quad \Gamma=1.1,\quad \rho=0.4,\quad \lambda=-1.45,$$

$$X=\begin{pmatrix}1&1&1&1&1&1&1&1&1&1\\1&2&3&4&5&6&7&8&9&10\end{pmatrix}'$$

$$D=(1\,2\,3\,4\,5\,6\,7\,8\,9\,10)'$$

and the number of replications is 2000.

In the conditional prediction we use the leave-one-out procedure as described in Lachenbruch (1975) and Rao (1987) or the cross-validation of Stone (1974), and the predictive sample ruse of Geisser (1974, 1975) and Lee & Geisser (1975). It means that the last point of each vector measurement is taken out as the true future value to be predicted each time. Although not shown, all predictors have similar errors.

When comparing the accuracy of extended prediction, we consider the one-step-ahead forecasts using the pseudo-cross-validation approach, which is in the spirit of the cross-validation and predictive sample reuse mentioned above. This process is addressed by using all the observations from all paths before the point being forecast in our samples. Table 8 shows the MAD and MARD of all methods including the MCMC. Generally speaking, they all perform similarly. Moreover, compared with the approximate Bayesian and the ML method, the MCMC methods does not render any significant improvement in predictive accuracy.

For the unbalanced data set, we consider only extended prediction. In addition, the prediction will be made using the most recent eight observations from the point

TABLE 8. Comparison of predictive accuracy in extended prediction I*

| Point being forecast | Point used to forecast | | Approximate Bayesian | | | MCMC | |
|---|---|---|---|---|---|---|---|
| | | | Prior 1 | Prior 2 | MLE | Prior 1 | Prior 2 |
| 6 | 1 ~ 5 | MAD | 0.00681 | 0.00688 | 0.00686 | 0.00702 | 0.00683 |
| 7 | 1 ~ 6 | | 0.00626 | 0.00633 | 0.00628 | 0.00623 | 0.00619 |
| 8 | 1 ~ 7 | | 0.00797 | 0.00795 | 0.00800 | 0.00800 | 0.00793 |
| 9 | 1 ~ 8 | | 0.01156 | 0.01160 | 0.01148 | 0.01131 | 0.01132 |
| 10 | 1 ~ 9 | | 0.00744 | 0.00725 | 0.00747 | 0.00738 | 0.00719 |
| Average | | | 0.00801 | 0.00800 | 0.00802 | 0.00799 | 0.00789 |
| 6 | 1 ~ 5 | MARD | 0.00630 | 0.00637 | 0.00635 | 0.00649 | 0.00631 |
| 7 | 1 ~ 6 | | 0.00555 | 0.00561 | 0.00556 | 0.00552 | 0.00549 |
| 8 | 1 ~ 7 | | 0.00678 | 0.00677 | 0.00681 | 0.00681 | 0.00675 |
| 9 | 1 ~ 8 | | 0.00897 | 0.00900 | 0.00891 | 0.00880 | 0.00879 |
| 10 | 1 ~ 9 | | 0.00557 | 0.00542 | 0.00559 | 0.00553 | 0.00559 |
| Average | | | 0.00663 | 0.00663 | 0.00664 | 0.00663 | 0.00658 |

*for the balanced subset

TABLE 9. Comparison of predictive accuracy in extended prediction II*

| Point being forecast | Point used to forecast | | Approximate Bayesian Prior 1 | Approximate Bayesian Prior 2 | MLE |
|---|---|---|---|---|---|
| 9 | 1 ~ 8 | MAD | 0.00725 | 0.00729 | 0.00762 |
| 10 | 2 ~ 9 | | 0.00741 | 0.00727 | 0.00740 |
| 11 | 3 ~ 10 | | 0.00799 | 0.00796 | 0.00773 |
| 12 | 4 ~ 11 | | 0.01164 | 0.01156 | 0.01198 |
| 13 | 5 ~ 12 | | 0.01684 | 0.01616 | 0.01702 |
| Average | | | 0.01023 | 0.01005 | 0.01035 |
| 9 | 1 ~ 8 | MARD | 0.00565 | 0.00568 | 0.00596 |
| 10 | 2 ~ 9 | | 0.00555 | 0.00543 | 0.00555 |
| 11 | 3 ~ 10 | | 0.00594 | 0.00593 | 0.00572 |
| 12 | 4 ~ 11 | | 0.00774 | 0.00767 | 0.00792 |
| 13 | 5 ~ 12 | | 0.01070 | 0.01026 | 0.01081 |
| Average | | | 0.00712 | 0.00699 | 0.00719 |

*using the most recent 8 observations as samples (unbalanced)



of being forecast as the samples in the pseudo-cross-validation process. Since the approximate Bayesian and the MCMC methods are about the same for predictive accuracy in the balanced case, only the approximate Bayesian approach and the ML method will be compared in this case. Table 9 lists the comparison of predictive accuracy. It shows that all three methods are somewhat comparable, although the approximate Bayesian approach with prior 2 is better.

## 7 Concluding remarks

The Bayesian methods presented in this paper, including the approximate Bayesian and Bayesian via MCMC methods, provide alternative ways for dealing with the growth curve model having random effects and AR(1) dependence while applying the Box–Cox transformation on the observations. From the analysis in Section 6 it is evident that model (1) is useful in forecasting the fatigue crack length growth data.

It is worthwhile noting that the proposed approximate Bayesian method can be quite useful in analysing the growth curve data when the prior is properly chosen. Moreover, the situations in which no transformation is needed and no random effect exists can be treated as special cases of the model (1). Therefore, the results are also useful for those situations. Of course, the MCMC methods outlined in this paper will be useful for data in which the appropriate Bayesian method is not adequate.

Finally, we also analysed the fatigue crack length data using model (1) assuming that all observations have random effects both in intercept and in slope. The model is bad at the prediction accuracy. Thus, we know that proper modelling is important for prediction purposes.

## REFERENCES

BOGDANOFF, J. L. & KOZIN, F. (1985) *Probability Model of Cumulative Damage* (New York, Wiley).

CASELLA, G. & GEORGE, E. I. (1992) Explaining the Gibbs sampler, *American Statistician*, 46, pp. 167–174.

CHI, E. M. & REINSEL, G. C. (1989) Models for longitudinal data with random effects and AR(1) errors, *J. Am. Statist. Assoc.*, 84, pp. 452–459.

CHIB, S. & GREENBERG, E. (1995) Understanding the Metropolis–Hastings algorithm, *American Statistician*, 49, pp. 327–335.

EDWARDS, W., LINDMAN, H. & SAVAGE, L. J. (1963) Bayesian statistical inference for psychological research, *Psychological Review*, 70, pp. 193–242.

GEISSER, S. (1974) A predictive approach to the random effect model, *Biometrika*, 61, pp. 101–107.

GEISSER, S. (1975) The predictive sample reuse method with application, *Journal of the American Statistical Association*, 70, pp. 320–328.

GELFAND, A. E. & SMITH, A. F. M. (1990) Sampling-based approaches to calculating marginal densities, *J. Am. Statist. Assoc.*, 85, pp. 398–409.

GELMAN, A. & RUBIN, D. B. (1992) Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, pp. 457–472.

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

GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. (1996) *Markov Chain Monte Carlo in Practice* (London, Chapman & Hall).

HASTINGS, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, pp. 97–109.

JENNRICH, R. I. & SCHLUCHTER, M. D. (1986) Unbalanced repeated-measures models with structured covariance matrices, *Biometrics*, 42, pp. 805–820.

KERAMIDAS, E. M. & LEE, J. C. (1990) Forecasting technological substitutions with concurrent short time series, *J. Am. Statis. Assoc.*, 85, pp. 625–632.

LACHENBRUCH, P. A. (1975) *Discriminant Analysis* (New York, Hafner).

LAIRD, N. M. & WARE, J. H. (1982) Random effects models for longitudinal data, *Biometrics*, 38, pp. 963–974.

LEE, J. C. (1988) Prediction and estimation of growth curve with special covariance structures, *Journal of the American Statistical Association*, 83, pp. 432–440.

LEE, J. C. (1991) Test and model selection for the general growth curve model, *Biometrics*, 47, pp. 147–159.

LEE, J. C. & GEISSER, S. (1975) Applications of growth curve prediction, *Sankya, Ser. A*, 37, pp. 239–256.

LEE, J. C. & LU, K. W. (1987) On a family of data-based transformed models useful in forecasting technological substitutions, *Technological Forecasting and Social Change*, 31, pp. 61–78.

LEE, J. C. & NIU, W. F. (1999) On an unbalanced growth curve model with random effects and AR(1) errors from a Bayesian and the ML points of view, *Journal of Statistical Planning and Inference*, 76, pp. 41–55.

LJUNG, G. M. & BOX, G. E. P. (1980) Analysis of variance with autocorrelated observations, *Scandinavian Journal of Statistics*, 7, pp. 172–180.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953) Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, 21, pp. 1087–1092.

POTTHOFF, R. F. & ROY, S. N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, 51, pp. 313–326.

RAO, C. R. (1987) Prediction of future observations in growth curve models, *Statistical Science*, 2, pp. 434–471.

STONE, M. (1974) Cross-validatory choice and assessment of statistical prediction (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 36, pp. 111–147.



## Appendix

Some results derived based on the maximum likelihood method.

### *A. Parameter estimation*

Through the fact that

$$\max_{\tau,\sigma^2,\Gamma,\rho,\lambda} L(\tau,\sigma^2,\Gamma,\rho,\lambda) = \max_{\Gamma,\rho,\lambda,\tau} \max_{\sigma^2|\Gamma,\rho,\lambda} \mathrm{L}(\tau,\sigma^2,\Gamma,\rho,\lambda) \tag{A.1}$$

the MLEs of the parameters $\tau$, $\sigma^2$, $\Gamma$, $\rho$ and $\lambda$ for model (1) can be given as:

$$\hat{\tau}_i = \left( \sum_{j=1}^{N_i} X_{ij}' \hat{\Lambda}_{ij}^{-1} X_{ij} \right)^{-1} \left( \sum_{j=1}^{N_i} X_{ij}' \hat{\Lambda}_{ij}^{-1} Y_{ij}^{(\hat{\lambda}_i)} \right) \tag{A.2}$$

$$\hat{\sigma}^2 = \frac{1}{n} \left[ \sum_{i=1}^{r} \sum_{j=1}^{N_i} (Y_{ij}^{(\hat{\lambda}_i)} - X_{ij}\hat{\tau}_i)' \hat{\Lambda}_{ij}^{-1} (Y_{ij}^{(\hat{\lambda}_i)} - X_{ij}\hat{\tau}_i) \right] \tag{A.3}$$

where $n = \Sigma_{i=1}^{r} \Sigma_{j=1}^{N_i} p_{ij}$, $\hat{\Lambda}_{ij} = D_{ij}\hat{\Gamma}_i D_{ij}' + \hat{C}_{ij}$ and $\hat{C}_{ij} = (\hat{\rho}^{|a-b|})$, $a, b = 1, 2, \ldots, p_{ij}$. It is noted that

$$(\hat{\Gamma}, \hat{\rho}, \hat{\lambda}) = \underset{\Gamma, \rho, \lambda}{\operatorname{argmax}} [\hat{\sigma}^2(\Gamma, \rho, \lambda)]^{-1/2} \left( \prod_{i=1}^{r} \prod_{j=1}^{N_i} |\Lambda_{ij}(\Gamma, \rho)|^{-1/2} \right) |J^{(\lambda)}| \tag{A.4}$$

where $|\mathcal{J}(\lambda)|$ is the Jacobian of the power transformation as defined in equation (20), and $\hat{\sigma}^2(\Gamma, \rho, \lambda)$ is the $\hat{\sigma}^2$ given by equation (A.3) with $\hat{\Gamma}$, $\hat{\rho}$ and $\hat{\lambda}$ replaced by $\Gamma$, $\rho$ and $\lambda$, respectively.

Since

$$\hat{\tau}_i | \sigma^2, \Gamma, \rho, \lambda \sim N_{m_i} \left( \tau_i, \sigma^2 \left( \sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} X_{ij} \right)^{-1} \right)$$

a $1 - \alpha$ confidence region for $\tau_i$ is approximately constructed as

$$\frac{1}{\sigma^2} (\tau_i - \hat{\tau}_i)' \left( \sum_{j=1}^{N_i} X_{ij}' \Lambda_{ij}^{-1} X_{ij} \right) (\tau_i - \hat{\tau}_i) \leqslant \chi^2_{m_i}(1 - \alpha)$$

where $\chi^2_{m_i}(1 - \alpha)$ is the $100(1 - \alpha)$th percentile of the $\chi^2$ distribution with $m_i$ degrees of freedom.

*B. Prediction of partially observed future values*

With the same assumption as described in Section 4, the estimators, $\hat{\tau}_l$, $\sigma^2$, $\hat{\Gamma}_l$, $\hat{\rho}_l$ and $\hat{\lambda}_l$ are obtained from equations (A.2), (A.3) and (A.4) with $Y = (Y_1, Y_2, \ldots, Y_N)$. We can predict the $q$-dimensional future value of the $k$th measurement in the $l$th group, $y_{lk}$, as the following:

$$\hat{y}_{lk} = \begin{cases} \{\mathbf{1} + \hat{\lambda}_l [x\hat{\tau}_l + \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}(Y_{lk}^{(\hat{\lambda}_l)} - X_{lk}\hat{\tau}_l)]\}^{1/\hat{\lambda}_l} & \text{where } \hat{\lambda}_l \neq 0 \\ \exp[x\hat{\tau}_l + \hat{\Omega}_{21}\hat{\Omega}_{11}^{-1}(Y_{lk}^{(\hat{\lambda}_l)} - X_{lk}\hat{\tau}_l)] & \text{where } \hat{\lambda}_l = 0 \end{cases} \tag{B.1}$$

where $\Omega = (\Omega_{ij})$, $i, j = 1, 2$, is the same as the $\Lambda^\star$ defined in equation (11) with $\Gamma_l$ and $\lambda_l$ replaced by $\hat{\Gamma}_l$ and $\hat{\lambda}_l$, $\mathbf{1} = (1, 1, \ldots, 1)'$, a $q \times 1$ vector. When $q = 1$, a $1 - \alpha$ predictive interval for $y_{lk}$ is given as $\hat{y}_{lk} \pm z(\alpha/2)\hat{\sigma}_{y_{lk}}$, where $z(\alpha/2)$ is the $100 \times (\alpha/2)\%$ point of the standard normal distribution and

$$\sigma^2_{y_{lk}} = [h'(\mu_{2.1})] \cdot \sigma^2 \cdot [\Omega_{22.1} + HW^{-1}H' - 2\Omega_{21}\Omega_{11}^{-1}X_{lk}W^{-1}H'] \tag{B.2}$$

is the variance of the forecast error, $H = x - \Omega_{21}\Omega_{11}^{-1}X_{lk}$, $W = \Sigma_{j=1}^{N_i} X_{lj}'\Omega_{11}^{-1}X_{lj}$, $\mu_{2.1} = x\hat{\tau} + \Omega_{21}\Omega_{11}^{-1}(Y_{lk} - X_{lk}\hat{\tau}_l)$, and $h'(u) = (1 + \lambda_l u)^{(1-\lambda_l)/\lambda_l}$.