Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 133 (2005) 435–451
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Bayesian analysis of Box–Cox transformed linear mixed models with ARMA(p, q) dependence☆

Jack C. Lee[a,*], Tsung I. Lin[b], Kuo J. Lee[c], Ying L. Hsu[d]

[a] *Institute of Statistics and Graduate Institute of Finance, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu, Taiwan*
[b] *Department of Statistics, Tunghai University, Taichung, Taiwan*
[c] *Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan*
[d] *Department of Applied Mathematics, National Chung Hsing University, Taichung, Taiwan*

Received 11 October 2002; accepted 19 March 2004
Available online 15 June 2004

Abstract

In this paper, we present a Bayesian inference methodology for Box–Cox transformed linear mixed model with ARMA(p, q) errors using approximate Bayesian and Markov chain Monte Carlo methods. Two priors are proposed and put into comparisons in parameter estimation and prediction of future values. The advantages of Bayesian approach over maximum likelihood method are demonstrated by both real and simulated data.

MSC: Primary 62F15; Secondary 62J99

Keywords: Approximate Bayesian; Maximum likelihood estimation; MCMC; Uniform prior; Random effects; Reparameterization

1. Introduction

The main purpose of this paper is to address the problem of analyzing growth curve data from a Bayesian point of view, using an unbalanced linear mixed model with ARMA(p, q)

☆ Research supported in part by NSC grants 92-2118-M-009-008, 92-2118-M-029-008 and 92-2118-M-05-001.

* Corresponding author. Tel.: +1-886-35131568; fax: +1-886-35728745.
E-mail addresses: jclee@stat.nctu.edu.tw (Jack C. Lee), tilin@mail.thu.edu.tw (Tsung I. Lin), u8826508@stat.nctu.edu.tw (Kuo J. Lee), ylhsu@amath.nchu.edu.tw (Ying L. Hsu).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.03.015

dependence, while applying the Box–Cox transformation (Box and Cox, 1964) on the observations.

The normal linear mixed models proposed by Laird and Ware (1982) have been widely applied in dealing with longitudinal data. They assumed the within-subject errors are independent and provided EM algorithms for obtaining the maximum likelihood (ML) estimates and the restricted maximum likelihood (RML) estimates of model parameters. Jennrich and Schluchter (1986) discussed various types of covariance structures, including random effects models and the AR(1) dependence separately. Chi and Reinsel (1989) presented an explicit ML estimation procedure using the scoring method for the model with both random effects and AR(1) errors and remarked that it may be worthwhile to merge higher-order ARMA(p, q) structures in the model. Bayesian analysis for ARMA(p, q) regression error models using the Markov chain Monte Carlo (MCMC) methods has been considered by Chib and Greenberg (1994). Rochon (1992) presented a fixed-effects model with ARMA structures of time heteroscedasticity for analyzing repeated measures experiments. More recently, Chib and Carlin (1999) constructed several partially and fully blocked MCMC algorithms for hierarchical mixed models with white noise errors.

Some transformations on the observations could enhance the justification of assumptions such as normality of the distribution or linearity of the growth function. Lee and Lu (1987) and Keramidas and Lee (1990) showed tremendous improvement in predictive accuracy using the Box–Cox transformation 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 transformation, along with incorporating into the model the proper dependence structure among the observations. Enhancement of normality and constancy of variance could have relatively minor roles in the improvement of predictive accuracy.

The model considered here is

$$\boldsymbol{Y}_i^{(\lambda)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \quad \text{for } i = 1, 2, \ldots, N, \tag{1}$$

where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{it_i})'$ is a $t_i \times 1$ vector of measurements and is independent of $\boldsymbol{Y}_j$ for all $i \neq j$, $\boldsymbol{\beta}$ is an unknown $m_1 \times 1$ vector of regression coefficients, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are known design matrices, $\boldsymbol{b}_i$ is a $m_2 \times 1$ random effects to be sampled from multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\sigma^2\boldsymbol{\Gamma}$, and $\boldsymbol{\varepsilon}_i$ is an independent $t_i \times 1$ vector of within subject errors whose components are assumed to follow the ARMA(p, q) model, i.e.,

$$\varepsilon_{ik} = \sum_{j=1}^{p} \phi_j \varepsilon_{i,k-j} - \sum_{j=1}^{q} \theta_j a_{i,k-j} + a_{ik} \quad \text{for } k = 1, \ldots, t_i,$$

where $\{a_{ik}\}$ is a series of shocks or white noise, which are identically and independently distributed as $\mathrm{N}(0, \sigma_a^2)$. In our study, we assume the observations for each subject are made at equally spaced intervals. Following Box et al. (1994), we write $\phi(B)\varepsilon_{it} = \theta(B)a_{it}$, where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials of B, which is the backshift operator such that $B\varepsilon_{ik} = \varepsilon_{i,k-1}$. For the process to be stationary and invertible so that there will be a unique model corresponding to the likelihood function, the roots of $\phi(B)$ and $\theta(B)$ must lie outside the unit circle, which constrains the parameter vectors $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$ to lie in regions $\mathbb{C}_p$ and $\mathbb{C}_q$, respectively.

For simplifying the estimating procedure, we shall denote $\sigma^2 \boldsymbol{C}_i$ as the covariance matrix of $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{C}_i = [\rho_{|r-s|}]$, where $r, s = 1, 2, \ldots, t_i$. We found that

$$\sigma^2 = (1 - \theta_1\psi_1 - \cdots - \theta_q\psi_q)\sigma_a^2/(1 - \phi_1\rho_1 - \cdots - \phi_p\rho_p),$$

where $\psi_j = \sum_{k=1}^{p} \phi_k \psi_{j-k} - \theta_j$ with $\theta_j = 0$ for $j > q$ and $\psi_j = 0$ for $j < 0$. It is noted that ρ_j's are implicit functions of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

The Box–Cox transformation is defined as

$$Y_{ij}^{(\lambda)} = \begin{cases} \frac{(Y_{ij}+\upsilon)^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0, \\ \log(Y_{ij} + \upsilon) & \text{if } \lambda = 0, \end{cases} \tag{2}$$

where Y_{ij} is the jth component of $\boldsymbol{Y}_i$, υ is a known constant such that $Y_{ij} + \upsilon > 0$, and λ is an unknown parameter. Without loss of generality, we will assume $\upsilon = 0$ for the rest of the paper. The covariance matrix of $\boldsymbol{Y}_i^{(\lambda)}$ can be written as

$$\boldsymbol{\Sigma}_i = \sigma^2(\boldsymbol{Z}_i \boldsymbol{\Gamma} \boldsymbol{Z}_i' + \boldsymbol{C}_i) = \sigma^2 \boldsymbol{\Lambda}_i(\boldsymbol{\Gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}). \tag{3}$$

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 which is useful in practice. Furthermore, in recent years statisticians have been increasingly drawn to MCMC methods, especially the M–H algorithm (Hastings, 1970; Chib and Greenberg, 1995) and the Gibbs sampler (Geman and Geman, 1984). Therefore, we also consider the problem for the prediction of future observations from a Bayesian point of view.

In Section 2, two types of priors are introduced. Approximate Bayesian methods for parameter estimation and prediction of future values are presented in Section 3. In Section 4, we considered the Bayesian inference by means of MCMC methodology. The results developed in this paper are illustrated in Section 5 with real and simulated data. Some concluding remarks are given in Section 6.

2. The priors

For the joint priors of the parameters $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Gamma}, \boldsymbol{\psi}, \lambda$, two categories of priors will be considered in this paper. In practice, we might use a uniform prior distribution if we really have no prior knowledge about the parameters. Following the consideration of Box and Cox (1964), the joint prior distribution is specified by

$$\pi(\boldsymbol{\beta}, \log\, \sigma, \boldsymbol{\Gamma}, \boldsymbol{\psi}, \lambda) = \pi(\boldsymbol{\beta}, \log\, \sigma \mid \lambda)\pi(\boldsymbol{\Gamma})\pi(\boldsymbol{\psi})\pi(\lambda). \tag{4}$$

In (4), for any specified λ,

$$\pi(\boldsymbol{\beta}, \log\, \sigma \mid \lambda) \propto |\, l_\lambda|^{-m_1},$$

where m_1 is the dimension of $\boldsymbol{\beta}$, l_λ denotes the geometric mean of the Jacobian

$$l_\lambda = J_\lambda^{1/n} = \left(\prod_{i=1}^{N} \prod_{j=1}^{t_i} Y_{ij}^{\lambda-1} \right)^{1/n}$$

and $n = \sum_{i=1}^{N} t_i$ denotes the size of observations.

For $\boldsymbol{\psi}$, we apply the Durbin–Levinson recursion (see Monahan, 1984) to reparameterize $\boldsymbol{\psi} = (\boldsymbol{\phi}, \boldsymbol{\theta})$ in terms of $\boldsymbol{\gamma}_{\boldsymbol{\psi}} = (\boldsymbol{\gamma}_{\boldsymbol{\phi}}, \boldsymbol{\gamma}_{\boldsymbol{\theta}})$, which is confined within $\mathbb{R}_1^{p+q}$, $\mathbb{R}_1 = (-1, 1)$, as discussed in Section 3.1. Thus, we allow for a uniform prior on $\boldsymbol{\gamma}_{\boldsymbol{\psi}}$. For σ^2, we choose σ^{-2} as its prior (see Zellner and Tiao, 1964). As for $\boldsymbol{\Gamma}$, two possibilities are considered. We may utilize the "principle of stable estimation" suggested by Edwards et al. (1963), a uniform prior is appropriate for $\boldsymbol{\Gamma}$. Therefore, the joint prior density can be represented as

$$\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Gamma}, \boldsymbol{\gamma}_{\boldsymbol{\psi}}, \lambda) \propto \sigma^{-2} J_\lambda^{-m_1/n} \pi(\lambda). \tag{5}$$

Another prior for $\boldsymbol{\Gamma}$ is to construct an informative prior distribution such as the inverse Wishart distribution $\mathrm{IW}(\boldsymbol{\Omega}^{-1}, \nu)$. The joint prior is then specified by

$$\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Gamma}, \boldsymbol{\gamma}_{\boldsymbol{\psi}}, \lambda) \propto \sigma^{-2} \mathrm{IW}(\boldsymbol{\Omega}^{-1}, \nu) J_\lambda^{-m_1/n} \pi(\lambda). \tag{6}$$

The hyperparameter ν is held fixed as small as possible, say $m_2 + 2$. Meanwhile, $\boldsymbol{\Omega}$ could be set as diagonal with diagonal elements being the sample variance of the corresponding regression coefficients that each $\mathbf{Y}_i^{(\lambda)}$ regresses on the design matrix $\boldsymbol{X}_i$ assuming white noise errors.

We shall refer to the joint priors (5) and (6) as prior 1 and prior 2 for the rest of this paper. In the following Bayesian inferences, we shall denote the prior of $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}_{\boldsymbol{\psi}}$ and λ as $\pi(\boldsymbol{\Gamma}, \boldsymbol{\gamma}_{\boldsymbol{\psi}}, \lambda) J_\lambda^{-m_1/n}$ without specifying which prior being used.

3. Approximate Bayesian inference

A simple reparameterization is given that can implicitly restrict the autoregressive moving average parameters to the stationary and invertible region. We can estimate the mode from bounded constrained regions for the parameters by way of transformation.

3.1. Reparameterization

In order to facilitate the estimating procedure and achieve the objective of ensuring admissibility of $(\boldsymbol{\phi}, \boldsymbol{\theta})$, we need to perform reparameterization on these parameters. Barndorff-Nielsen and Schou (1973) proposed the following one-to-one and onto transformation which reparameterizes $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ in terms of the partial autocorrelations $\boldsymbol{\gamma}_{\boldsymbol{\phi}} = (\gamma_{\phi_1}, \ldots, \gamma_{\phi_p})$ for the AR(p) process:

$$\phi_k^{(k)} = \gamma_{\phi_k}, \quad \phi_i^{(k)} = \phi_i^{(k-1)} - \gamma_{\phi_k} \phi_{k-i}^{(k-1)}, \quad i = 1, 2, \ldots, k-1 \tag{7}$$

and the condition that $\boldsymbol{\phi} \in \mathbb{C}_p$ becomes $| \gamma_{\phi_k} | \leqslant 1, k = 1, \ldots, p$. For the MA($q$) process, the reparameterization scheme is identical to the AR(p) process, with ϕ_k and γ_{ϕ_k} in (7) replaced by θ_k and γ_{θ_k}, as noted by Monahan (1984).

For the general ARMA(p, q) process, a reparameterization of $(\boldsymbol{\phi}, \boldsymbol{\theta})$ in terms of $(\boldsymbol{\gamma_\phi}, \boldsymbol{\gamma_\theta})$ can be obtained by applying (7) on both $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. With this reparameterization, the condition that $\boldsymbol{\psi} \in \mathbb{C}_p \times \mathbb{C}_q$ will be simplified to the condition that $\boldsymbol{\gamma_\psi} \in \mathbb{R}_1^{p+q}$, $\mathbb{R}_1 = (-1, 1)$. Thus, the Bayesian estimation procedure and further Bayesian MCMC generation are done in the space of $\boldsymbol{\gamma_\psi}$ before inverting back to $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

3.2. Parameter estimation

After transforming $\boldsymbol{\psi}$ to $\boldsymbol{\gamma_\psi}$, the covariance matrix of $\boldsymbol{Y}_i^{(\lambda)}$ as given in (3) can be rewritten as $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{\Lambda}_i(\boldsymbol{\Gamma}, \boldsymbol{\gamma_\psi})$. Integrating the joint posterior density w.r.t. σ^2 and $\boldsymbol{\beta}$, we have

$$p(\boldsymbol{\Gamma}, \boldsymbol{\gamma_\psi}, \lambda \,|\, \boldsymbol{Y}) \propto \pi(\boldsymbol{\Gamma}, \boldsymbol{\gamma_\psi}, \lambda) \prod_{i=1}^{N} | \boldsymbol{\Lambda}_i |^{-1/2} \boldsymbol{B}^{-(n-m_1)/2} \times \left| \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Lambda}_i^{-1} \boldsymbol{X}_i \right|^{-1/2} J_\lambda^{(n-m_1)/n}, \tag{8}$$

where

$$\boldsymbol{B} = \sum_{i=1}^{N} (\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \tilde{\boldsymbol{\beta}})' \boldsymbol{\Lambda}_i^{-1} (\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \tilde{\boldsymbol{\beta}}), \tag{9}$$

with

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Lambda}_i^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Lambda}_i^{-1} \boldsymbol{Y}_i^{(\lambda)} \right). \tag{10}$$

Let $\hat{\boldsymbol{\beta}}$ denote the mode of $\boldsymbol{\beta}$ and the mode for σ^2 be $\hat{\boldsymbol{B}}/(n+2)$. Here $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{B}}$ are $\tilde{\boldsymbol{\beta}}_i$ and $\boldsymbol{B}$ with parameters $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma_\psi}$ and λ, respectively, replaced by the modes $\hat{\boldsymbol{\Gamma}}$, $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\psi}}$ and $\hat{\lambda}$, which maximize (8). The maximization can be solved by the "nlminb" function in S-PLUS with bounded constraint of $(-1, 1)$ on γ_{ϕ_k}'s and γ_{θ_k}'s.

3.3. Approximate Bayesian prediction

We consider the prediction of $\boldsymbol{y}_i$, a future w-dimensional values of measurements $\boldsymbol{Y}_i$. Let $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ be, respectively, the $w \times m_1$ and $w \times m_2$ design matrices corresponding to $\boldsymbol{y}_i$ and let $\boldsymbol{Y}_i^* = (\boldsymbol{Y}_i', \boldsymbol{y}_i')'$, $\boldsymbol{X}_i^* = (\boldsymbol{X}_i', \boldsymbol{x}_i')'$ and $\boldsymbol{Z}_i^* = (\boldsymbol{Z}_i', \boldsymbol{z}_i')'$.

We have

$$\mathrm{Var}(\boldsymbol{Y}_i^*) = \sigma^2 (\boldsymbol{Z}_i^* \boldsymbol{\Gamma} \boldsymbol{Z}_i^{*\prime} + \boldsymbol{C}_i^*) = \sigma^2 \boldsymbol{\Lambda}_i^*(\boldsymbol{\Gamma}, \boldsymbol{\gamma_\psi}),$$

where $\boldsymbol{\Lambda}_i^* = [\Lambda_{i,jj'}^*](j, \ \ j' = 1, 2)$ and $\boldsymbol{C}_i^* = [\rho_{|r-s|}]$ for $r, s = 1, \ldots, t_i + w$.

Combining the conditional density function of $\boldsymbol{y}_i^{(\lambda)}$ given $\boldsymbol{Y}_i$, $\boldsymbol{\beta}$, σ^2, $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}_{\boldsymbol{\psi}}$ and λ with the joint posterior density of $\boldsymbol{\beta}$, σ^2, $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}_{\boldsymbol{\psi}}$ and λ, and integrating w.r.t. $\boldsymbol{\beta}$ and σ^2, we obtain, after some algebraic manipulation, the following posterior density:

$$
\begin{aligned}
&p(\boldsymbol{y}_i^{(\lambda)}, \boldsymbol{\Gamma}, \boldsymbol{\gamma}_{\boldsymbol{\psi}}, \lambda \,|\, \boldsymbol{Y}) \\
&\quad \propto \pi(\boldsymbol{\Gamma}, \boldsymbol{\gamma}_{\boldsymbol{\psi}}, \lambda) |\boldsymbol{\Lambda}_{i,22\cdot 1}^{*}|^{-1/2} \prod_{i=1}^{N} |\boldsymbol{\Lambda}_i|^{-1/2} |\boldsymbol{Q}_1 + \boldsymbol{Q}_2|^{-1/2} \left| \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Lambda}_i^{-1} \boldsymbol{X}_i \right|^{-1/2} \\
&\qquad \times \left\{ \boldsymbol{B}_1 + \boldsymbol{B}_2 + (\boldsymbol{y}_i^{(\lambda)} - \boldsymbol{\mu}_i)' \boldsymbol{G}_{22} (\boldsymbol{y}_i^{(\lambda)} - \boldsymbol{\mu}_i) \right\}^{-(n+w-m_1)/2} J_{\lambda}^{(n-m_1)/n}, \qquad (11)
\end{aligned}
$$

where

$$
\boldsymbol{Q}_1 = \sum_{j \neq i} \boldsymbol{X}_j' \boldsymbol{\Lambda}_j^{-1} \boldsymbol{X}_j, \quad \boldsymbol{Q}_2 = \boldsymbol{X}_i^{*\prime} \boldsymbol{\Lambda}_i^{*-1} \boldsymbol{X}_i^{*},
$$

$$
\boldsymbol{\mu}_i = \boldsymbol{x}_i \boldsymbol{\beta}_i^{*} - \boldsymbol{G}_{22}^{-1} \boldsymbol{G}_{21} (\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \boldsymbol{\beta}_i^{*}),
$$

$$
\begin{aligned}
\boldsymbol{G} = &\boldsymbol{\Lambda}_i^{*-1} \boldsymbol{X}_i^{*} \boldsymbol{Q}_2^{-1} \boldsymbol{Q}_1 (\boldsymbol{Q}_1 + \boldsymbol{Q}_2)^{-1} \boldsymbol{X}_i^{*\prime} \boldsymbol{\Lambda}_i^{*-1} \\
&+ \boldsymbol{A}_i^{*} (\boldsymbol{A}_i^{*\prime} \boldsymbol{\Lambda}_i^{*} \boldsymbol{A}_i^{*})^{-1} \boldsymbol{A}_i^{*\prime} = [\boldsymbol{G}_{jj'}] \quad \text{for } j, j' = 1, 2,
\end{aligned}
$$

with $\boldsymbol{A}_i^{*}$ is an $(t_i + w) \times (t_i + w - m_1)$ matrix such that $\boldsymbol{X}_i^{*\prime} \boldsymbol{A}_i^{*} = \boldsymbol{0}$, and

$$
\begin{aligned}
\boldsymbol{B}_1 &= \sum_{j \neq i} (\boldsymbol{Y}_j^{(\lambda)} - \boldsymbol{X}_j \boldsymbol{\beta}_i^{*})' \boldsymbol{\Lambda}_j^{-1} (\boldsymbol{Y}_j^{(\lambda)} - \boldsymbol{X}_j \boldsymbol{\beta}_i^{*}), \\
\boldsymbol{B}_2 &= (\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \boldsymbol{\beta}_i^{*})' \boldsymbol{G}_{11.2} (\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \boldsymbol{\beta}_i^{*}), \\
\boldsymbol{\beta}_i^{*} &= \left(\sum_{j \neq i} \boldsymbol{X}_j' \boldsymbol{\Lambda}_j^{-1} \boldsymbol{X}_j \right)^{-1} \left(\sum_{j \neq i} \boldsymbol{X}_j' \boldsymbol{\Lambda}_j^{-1} \boldsymbol{Y}_j^{(\lambda)} \right), \\
\boldsymbol{G}_{11\cdot 2} &= \boldsymbol{G}_{11} - \boldsymbol{G}_{12} \boldsymbol{G}_{22}^{-1} \boldsymbol{G}_{21}.
\end{aligned}
$$

Following the approximate method of Ljung and Box (1980), the approximate predictive distribution of $\boldsymbol{y}_i^{(\lambda)}$ is

$$
\boldsymbol{y}_i^{(\lambda)} \mid \boldsymbol{Y} \dot{\sim} T_w(\hat{\boldsymbol{\mu}}_i, \ (\hat{\boldsymbol{B}}_1 + \hat{\boldsymbol{B}}_2) \left((n - m_1) \hat{\boldsymbol{G}}_{22} \right)^{-1}, \ n - m_1), \qquad (12)
$$

where $T_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the n-variate t distribution with location vector $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$ and

$$
\hat{\boldsymbol{\mu}}_i = \boldsymbol{x}_i \hat{\boldsymbol{\beta}}_i^{*} - \hat{\boldsymbol{G}}_{22}^{-1} \hat{\boldsymbol{G}}_{21} (\boldsymbol{Y}_i^{(\hat{\lambda})} - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}_i^{*}).
$$

It is noted that the quantities $\hat{\boldsymbol{\beta}}_i^*$, $\hat{\boldsymbol{B}}_1$, $\hat{\boldsymbol{B}}_2$ and $\hat{\boldsymbol{G}}_{22}$ are $\boldsymbol{\beta}_i^*$, $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ and $\boldsymbol{G}_{22}$ with $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}_\psi$ and λ replaced, respectively, by $\hat{\boldsymbol{\Gamma}}$, $\hat{\boldsymbol{\gamma}}_\psi$ and $\hat{\lambda}$ which maximize

$$p(\boldsymbol{\Gamma}, \boldsymbol{\gamma}_\psi, \lambda \mid \boldsymbol{Y}) \propto \pi(\boldsymbol{\Gamma}, \boldsymbol{\gamma}_\psi, \lambda) \mid \boldsymbol{\Lambda}_{i,22\cdot1}^*|^{-1/2} \prod_{i=1}^{N} \mid \boldsymbol{\Lambda}_i|^{-1/2} |\boldsymbol{Q}_1 + \boldsymbol{Q}_2|^{-1/2} \times |\boldsymbol{G}_{22}|^{-1/2} (\boldsymbol{B}_1 + \boldsymbol{B}_2)^{-(n-m_1)/2} J_\lambda^{(n-m_1)/n}. \tag{13}$$

Also, we can predict $\boldsymbol{y}_i$ by the following approximate predictor:

$$\hat{y}_{ih} = \begin{cases} (1 + \hat{\lambda}\hat{\mu}_{ih})^{1/\hat{\lambda}} & \text{if } \hat{\lambda} \neq 0, \\ \exp(\hat{\mu}_{ih}) & \text{if } \hat{\lambda} = 0, \end{cases} \quad \text{for } h = 1, 2, \ldots, w, \tag{14}$$

where $\hat{y}_{ih}$ and $\hat{\mu}_{ih}$ denote the hth component of $\hat{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{\mu}}_i$, respectively.

4. Bayesian inference via MCMC sampling

4.1. The algorithm

The following sampling scheme is used to obtain the posterior distributions of $\boldsymbol{\beta}$, σ^2, $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}_\psi$ and λ. Starting with some initial values of all the unknown parameters, the sampler successively generates a parameter or block of parameters conditional on the observations and the other parameters as described below. The iterations of the sampler are divided into a burn-in period and a sampling period. It is assumed that the sampler has converged to the correct posterior distribution at the end of the burn-in period and estimates of the posterior moments and densities are based on the iterations in the sampling period. The MCMC procedures are outlined as follows:

Step 1: Generate $\boldsymbol{\beta}$ given σ^2, $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}_\psi$, λ, and $\boldsymbol{Y}$ from

$$N_{m_1}\left(\tilde{\boldsymbol{\beta}}, \sigma^2 \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Lambda}_i^{-1}(\boldsymbol{\Gamma}, \boldsymbol{\gamma}_\psi) \boldsymbol{X}_i\right)^{-1}\right),$$

where $\tilde{\boldsymbol{\beta}}$ is given in (10).

Step 2: Generate σ^2 given $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}_\psi$, λ and $\boldsymbol{Y}$ from

$$\mathrm{IG}\left(\frac{n}{2}, \frac{S(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\gamma}_\psi, \lambda)}{2}\right),$$

where $S(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\gamma}_\psi, \lambda) = \sum_{i=1}^{N} (\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \boldsymbol{\beta})' \boldsymbol{\Lambda}_i^{-1}(\boldsymbol{\Gamma}, \boldsymbol{\gamma}_\psi)(\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \boldsymbol{\beta})$.

Step 3: Generate $\boldsymbol{\Gamma}$ via the M–H algorithm from

$$f(\boldsymbol{\Gamma}) \propto \pi(\boldsymbol{\Gamma}) \prod_{i=1}^{N} \mid \boldsymbol{\Lambda}_i(\boldsymbol{\Gamma}, \boldsymbol{\gamma}_\psi)|^{-1/2} \exp\left\{-\frac{S(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\gamma}_\psi, \lambda)}{2\sigma^2}\right\}.$$

Step 4: Generate γ_ψ via the M–H algorithm from

$$f(\gamma_\psi) \propto \pi(\gamma_\psi) \prod_{i=1}^{N} \mid \Lambda_i(\Gamma, \gamma_\psi)|^{-1/2} \exp\left\{-\frac{S(\beta, \Gamma, \gamma_\psi, \lambda)}{2\sigma^2}\right\}.$$

Step 5: Generate λ via M–H algorithm from

$$f(\lambda) \propto \pi(\lambda) \exp\left\{-\frac{S(\beta, \Gamma, \gamma_\psi, \lambda)}{2\sigma^2}\right\} J_\lambda^{(n-m_1)/n}.$$

For implementing the M–H algorithm at the kth iteration in Step 3, we choose Wishart distribution, $W(\Gamma^{(k)}/(N-1), N-1)$, as the proposal distribution $q(\Gamma^{(k+1)}|\Gamma^{(k)})$. As regards $\gamma_\psi = (\gamma_\phi, \gamma_\theta)$, we can transform γ_ψ to $\psi^* = (\phi_1^*, \ldots, \phi_p^*, \theta_1^*, \ldots, \theta_q^*) \in \mathbb{R}^{p+q}$, where $\mathbb{R} = (-\infty, \infty)$, $\phi_i^* = \log((1+\gamma_{\phi_i})/(1-\gamma_{\phi_i}))$ $(i = 1, \ldots, p)$ and $\theta_j^* = \log((1+\gamma_{\theta_j})/(1-\gamma_{\theta_j}))$ $(j = 1, \ldots, q)$. We then apply the M–H algorithm to the following conditional distribution:

$$g(\psi^* \mid \beta, \sigma^2, \Gamma, \lambda, Y) \propto \pi(\Gamma, \gamma_\psi(\psi^*), \lambda) \left(\prod_{i=1}^{N} \mid \Lambda_{ij}(\Gamma, \gamma_\psi(\psi^*))|^{-1/2}\right) \times \exp\left\{-\frac{S(\beta, \Gamma, \gamma_\psi(\psi^*), \lambda)}{2\sigma^2}\right\} J(\gamma_\psi(\psi^*)),$$

where $\gamma_\psi(\psi^*) = (e^{\psi^*} - 1)/(e^{\psi^*} + 1)$, and

$$J(\gamma_\psi(\psi^*)) = \prod_{i=1}^{p} \prod_{j=1}^{q} (2e^{\phi_i^*}/(1+e^{\phi_i^*})^2)(2e^{\theta_j^*}/(1+e^{\theta_j^*})^2).$$

The $p+q$ dimensional multivariate normal distribution with mean $\psi^{*(k)}$ and covariance matrix $c^2\Sigma_{\psi^*}$ are chosen as the proposed distribution, $q(\psi^{*(k+1)} \mid \psi^{*(k)})$, where the scale $c \approx 2.4/\sqrt{d}$ and d is the dimension of the parameters, as suggested in Gelman et al. (1995). The value of Σ_{ψ^*} is usually chosen to reflect the conditional covariance of ψ^* given β, σ^2, Γ, λ and Y. Thus, we can estimate the covariance matrix Σ_{ψ^*} by the following method. At the $(k+1)$th iteration, the preliminary covariance matrix estimate of ψ^*, Σ_{ψ^*}, would be the inverted sample information matrix of ψ^* given $\psi^{*(k)}$ in the MCMC algorithm. Having obtained ψ^* from the M–H algorithm, we transform ψ^* back to ψ. The same operation can also be applied to λ with the prior on λ being uniform over some finite interval, for example, $(-4, 4)$. Repeat Steps 1–5 until the sequences become stable. After sufficiently long burn-in iterations, we can use the remaining samples to estimate the function of the parameters in which we are interested.

4.2. Convergence diagnostics

Before conducting inference using MCMC samples, the output should be analyzed to determine a time point at which the sample has converged to the proper limiting distribution. Our recommended general approach to monitoring convergence is based on comparing

several sequences drawn from different starting points and checking that they are indistinguishable. The approach of Gelman and Rubin (1992) requires running $I \geqslant 2$ parallel chains with over-dispersed starting values. It provides a basis for an estimate of how close the process is to stationarity. However, their approach can only be applied in univariate problems. An alternative method is provided by Brooks and Gelman (1998), which extends the method of Gelman and Rubin (1992) to consider several parameters simultaneously.

Suppose there are I independent parallel chains and the length of each chain is $2T$. Let $\boldsymbol{\alpha}$ denote a vector of parameters and $\boldsymbol{\alpha}_i^{(t)}$ denote the parameter vector in ith chain at time t ($i = 1, ..., I$). Then the posterior variance–covariance matrix of $\boldsymbol{\alpha}$ is estimated by

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}|\boldsymbol{Y}} = \frac{T-1}{T}\boldsymbol{W} + \left(1 + \frac{1}{I}\right)\boldsymbol{B}/T, \tag{15}$$

where

$$\boldsymbol{W} = \frac{1}{I(T-1)} \sum_{i=1}^{I} \sum_{t=T+1}^{2T} (\boldsymbol{\alpha}_i^{(t)} - \bar{\boldsymbol{\alpha}}_{i\cdot})(\boldsymbol{\alpha}_i^{(t)} - \bar{\boldsymbol{\alpha}}_{i\cdot})',$$

$$\boldsymbol{B}/T = \frac{1}{I-1} \sum_{i=1}^{I} (\bar{\boldsymbol{\alpha}}_{i\cdot} - \bar{\boldsymbol{\alpha}}_{\cdot\cdot})(\bar{\boldsymbol{\alpha}}_{i\cdot} - \bar{\boldsymbol{\alpha}}_{\cdot\cdot})'$$

and

$$\bar{\boldsymbol{\alpha}}_{i\cdot} = \frac{1}{T} \sum_{t=T+1}^{2T} \boldsymbol{\alpha}_i^{(t)}, \quad \bar{\boldsymbol{\alpha}}_{\cdot\cdot} = \frac{1}{I} \sum_{i=1}^{I} \boldsymbol{\alpha}_{i\cdot}.$$

Brooks and Gelman (1998) provided the multivariate potential scale reduction factor (MPSRF), $\hat{\mathscr{R}}^p$, which is defined by

$$\hat{\mathscr{R}}^p = \frac{T-1}{T} + \left(1 + \frac{1}{I}\right)\lambda_1, \tag{16}$$

where λ_1 is the largest eigenvalue of $\boldsymbol{W}^{-1}\boldsymbol{B}/T$.

Under the assumption that I chains are equal, $\hat{\mathscr{R}}^p$ will tend to 1 for reasonably large T. The advantage of $\hat{\mathscr{R}}^p$ is inherent in the fact that it reliably summarizes each of the univariate measures in a single value. As the simulation converges, the $\hat{\mathscr{R}}^p$ will decline to 1 and the determinants should stabilize, meaning that the parallel Markov chains are essentially overlapping.

4.3. Forecast using MCMC samples

We have $\boldsymbol{y}_i^{(\lambda)} \mid \boldsymbol{\alpha}, \boldsymbol{Y} \sim \mathscr{N}_w(\boldsymbol{\mu}_i, \sigma^2 \boldsymbol{\Lambda}^*_{i,22\cdot1})$, where $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\Gamma}, \sigma^2, \boldsymbol{\gamma}_{\boldsymbol{\psi}}, \lambda)$, $\boldsymbol{\mu}_i = \boldsymbol{x}_i\boldsymbol{\beta} - \boldsymbol{\Lambda}^*_{i,21}\boldsymbol{\Lambda}^{*-1}_{i,11}(\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i\boldsymbol{\beta})$ and $\boldsymbol{\Lambda}^*_{i,22\cdot1} = \boldsymbol{\Lambda}^*_{i,22} - \boldsymbol{\Lambda}^*_{i,21}\boldsymbol{\Lambda}^{*-1}_{i,11}\boldsymbol{\Lambda}^*_{i,12}$. Hence, we can generate $\boldsymbol{y}_i^{(\lambda)(s,t)}$ from $f(\boldsymbol{y}_i^{(\lambda)} \mid \boldsymbol{\alpha}^{(s,t)}, \boldsymbol{Y})$, where $\boldsymbol{\alpha}^{(s,t)}$ is the values of the sth chain and the tth iteration of the

MCMC sampler of $\boldsymbol{\alpha}$. We can transform $\boldsymbol{y}_i^{(\lambda)(s,t)}$ to $\boldsymbol{y}_i^{(s,t)}$ by inverting formula of (2). For each chain, we can predict $\boldsymbol{y}_i$ from the MCMC runs by

$$\hat{\boldsymbol{y}}_i = \frac{1}{IT} \sum_{s=1}^{I} \sum_{t=T+1}^{2T} \boldsymbol{y}_i^{(s,t)}. \tag{17}$$

Prediction interval and quantiles of the functional $\boldsymbol{y}_i$ can be computed from the transformed simulation sample. Theoretically, the posterior probability of the case $\lambda = 0$ is zero. However, we had to program the two cases ($\lambda = 0$ and $\lambda \neq 0$) of the forecast rule separately to avoid the overflow program.

5. Numerical illustration

We present examples with one real data and a simulation study to illustrate our methodology in this section.

5.1. A real example

We apply the results obtained in sections 2–4 to the fatigue crack growth data from Bogdanoff and Kozin (1985) as plotted in Fig. 1(a). The data set was also analyzed by Lu and Meeker (1993), Robinson and Crowder (2000), Lee and Lien (2001) and Lin and Lee (2003). This data set is balanced for the first 10 time points. Fig. 1(b) is the plot of the data

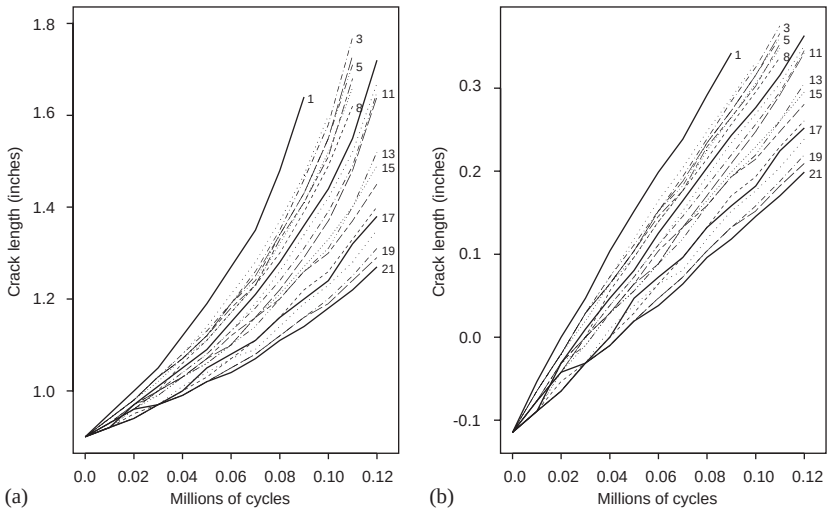

Fig. 1. Fatigue crack growth data over 21 subjects: (a) Original data; (b) data after taking Box–Cox transformation with $\lambda = -1.58$.

Table 1
Comparison of parameter estimations

Time points		β_1	β_2	$\sigma^2(\times 10^{-5})$	Γ	ϕ	θ	λ
	MLE	−0.1507	0.03735	3.3617	1.1204	0.5982	0.2113	−1.4421
$t=10$	Prior 1	−0.1507	0.03736	3.3820	1.1500	0.6196	0.2243	−1.4400
	Prior 2	−0.1508	0.03737	3.5958	0.9207	0.6569	0.2446	−1.4389
	MLE	−0.1506	0.03737	3.7376	1.0616	0.7341	0.3183	−1.4074
$t=11$	Prior 1	−0.1507	0.03738	3.8024	1.0810	0.7523	0.3293	−1.4049
	Prior 2	−0.1507	0.03390	4.1487	0.8497	0.7872	0.3487	−1.4020
	MLE	−0.1506	0.03718	3.7282	1.0310	0.7072	0.2758	−1.5043
$t=12$	Prior 1	−0.1506	0.03719	3.7722	1.0550	0.7220	0.2838	−1.5027
	Prior 2	−0.1507	0.03720	4.0381	0.8413	0.7495	0.2962	−1.5011
	MLE	−0.1506	0.03704	4.2433	0.8713	0.7071	0.2185	−1.5777
$t=13$	Prior 1	−0.1506	0.03705	4.2907	0.8925	0.7191	0.2242	−1.5768
	Prior 2	−0.1507	0.03706	4.5938	0.7100	0.7435	0.2331	−1.5760

after taking the Box–Cox transformation with λ being -1.58, and it reveals that the linearity assumption is satisfied after applying the adequate Box–Cox transformation on observations.

The fitted model is considered by

$$\boldsymbol{Y}_i^{(\lambda)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i b_i + \boldsymbol{\varepsilon}_i, \quad b_i \sim \mathrm{N}(0, \sigma^2\Gamma), \quad \boldsymbol{\varepsilon}_i \sim N_{t_i}(\boldsymbol{0}, \sigma^2\boldsymbol{C}_i), \tag{18}$$

where the design matrices for $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ have the forms

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & t_i \end{bmatrix}, \quad \boldsymbol{Z}_i = [1 \quad 2 \quad \cdots \quad t_i]', \quad i = 1, 2, \ldots, 21.$$

In our preliminary analysis, we fitted Model (18) with selected ARMA(p, q) covariance structures for which $p+q \leqslant 3$ based on the ML approach. We found that ARMA(1,1) model has the best predictive accuracy and is far better than AR(1). The detailed results are not shown in the paper. To further compare the ML approach with the approximate Bayesian and MCMC methods, we will focus on the ARMA(1,1) model only. It would not be necessary to transform (ϕ, θ).

Table 1 lists the estimates of parameters obtained via the ML and the approximate Bayesian methods for the fatigue crack data using the first t measurements of each subject as the sample ($t = 10, 11, 12, 13$). We run seven parallel chains and choose the starting values dispersed around the posterior modes of the parameters.

The convergence of MCMC samplers is monitored by examining the $\hat{\mathscr{R}}^p$ as discussed in Section 4.2. The monitored values of $\hat{\mathscr{R}}^p$ and the determinants of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}|\boldsymbol{Y}}$ and $\boldsymbol{W}$ are plotted in Figs. 2(a) and (b), respectively. The convergence occurs around 2000 iterations, so the determination of burn-in is about 2000. Table 2 gives the 2.5%, 25%, 50%, 75%, and 97.5% posterior quantiles together with the mean and standard deviation using 5000 converged

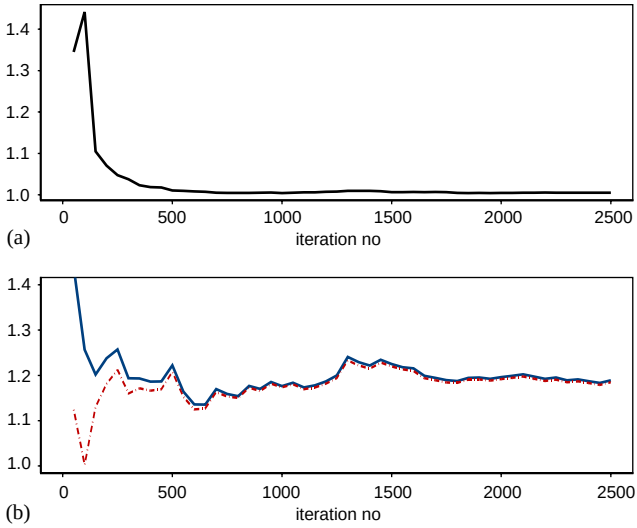

Fig. 2. (a) Plot of MPRSF, $\hat{\mathscr{R}}^p$; (b) plot of the determinants ($\times 10^{27}$) of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}|\boldsymbol{Y}}$ (solid) and $\boldsymbol{W}$ (dotted).

Table 2
Summaries of MCMC posterior distributions for the whole sample

Parameter	Prior	2.5%	25%	median	75%	97.5%	mean	s.d.
β_1	1	−0.1530	−0.1511	−0.1503	−0.1494	−0.1478	−0.1503	0.0013
	2	−0.1532	−0.1512	−0.1503	−0.1494	−0.1477	−0.1503	0.0014
β_2	1	0.0340	0.0360	0.0369	0.0380	0.0397	0.0370	0.0015
	2	0.0342	0.0360	0.0370	0.0378	0.0396	0.0370	0.0013
$\sigma^2(\times 10^{-5})$	1	2.8900	3.3777	3.6778	4.0813	5.1269	3.7600	0.5742
	2	2.9393	3.5183	3.8588	4.2781	5.6619	3.9638	0.6755
Γ	1	0.6208	0.9510	1.2051	1.5327	2.4299	1.2904	0.4883
	2	0.4930	0.7556	0.9306	1.1760	1.9093	0.9979	0.3515
ϕ	1	0.3940	0.5471	0.6306	0.7049	0.8167	0.6228	0.1106
	2	0.4143	0.5738	0.6461	0.7121	0.8361	0.6339	0.1062
θ	1	−0.0569	0.0918	0.1825	0.2602	0.4101	0.1782	0.1150
	2	−0.0424	0.0914	0.1766	0.2531	0.4028	0.1741	0.1173
λ	1	−1.6622	−1.6111	−1.5817	−1.5526	−1.4972	−1.5821	0.0430
	2	−1.6641	−1.6124	−1.5830	−1.5513	−1.4959	−1.5822	0.0436

MCMC simulated samples. It can be seen that all the methods considered here give similar estimates for $\boldsymbol{\beta} = (\beta_1, \beta_2)$ and λ but estimates using different priors are somewhat different for the components of the covariance structure.

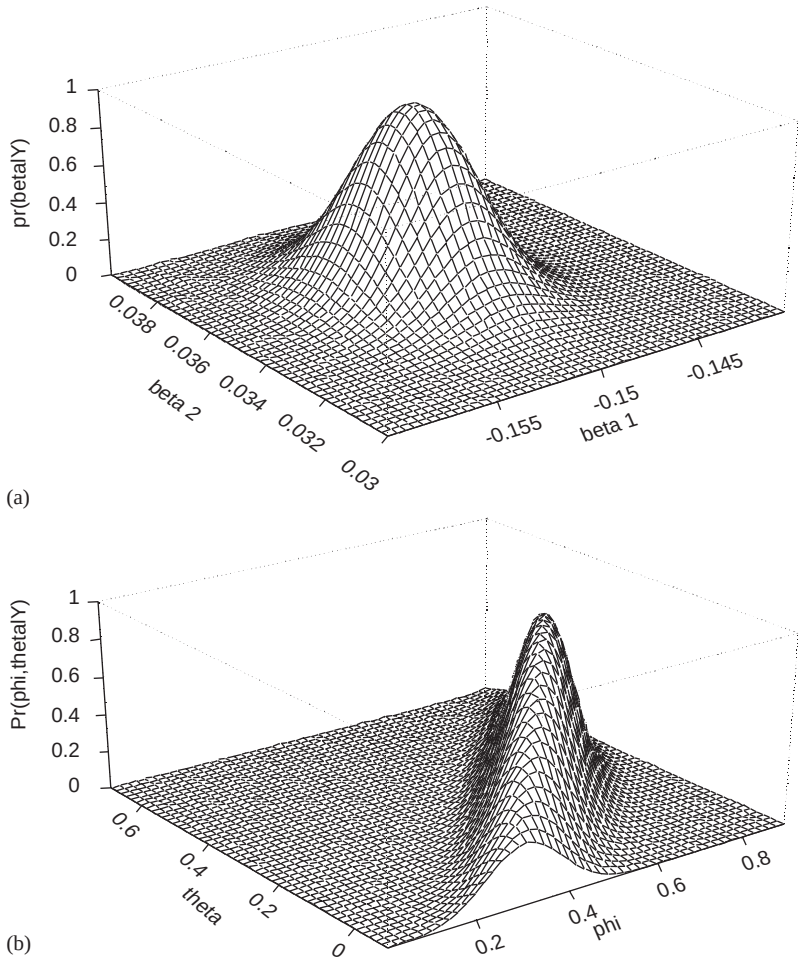

Fig. 3. (a) Marginal posterior distribution of $\boldsymbol{\beta} = (\beta_1, \beta_2)$; (b) marginal posterior distribution of $\boldsymbol{\psi} = (\phi, \theta)$.

By the simulated MCMC samples, we can also obtain the marginal posterior densities of parameters of interest using the well-known Rao-Blackwellization approximation. Fig. 3(a) exhibits the marginal density of $\boldsymbol{\beta} = (\beta_1, \beta_2)$, which indicates that an approximate bivariate normal distribution for $\boldsymbol{\beta}$ is quite reasonable. In particular, the marginal posterior density of $\boldsymbol{\psi} = (\phi, \theta)$ plotted in Fig. 3(b) is well concentrated and nearly symmetric. For the first subject in the last measurement given the entire data, Fig. 4 shows the approximate Bayesian predictive density (dashed curve) of (12), Rao-Blackwellization approximation (solid curve) and MCMC samples (histogram) using the prior 1. In the figure, Rao-Blackwellization provides a good approximation of MCMC samples.

Now we compare the prediction ability among the ML method (similar in form as (B.1) of Lee and Lien, 2001), approximate Bayesian prediction for (14) and MCMC prediction for (17) using both priors. We consider the extended prediction, which is usually of practical

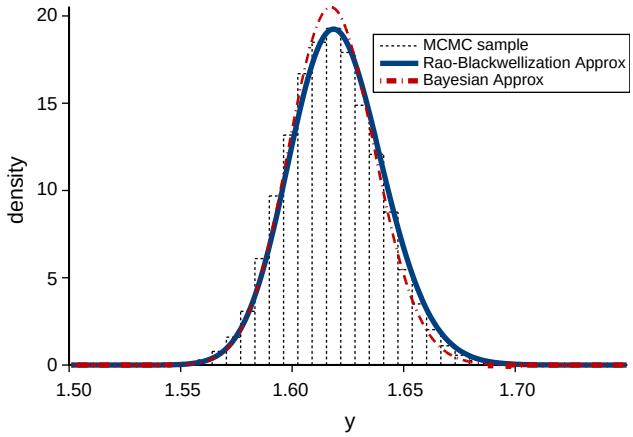

Fig. 4. Comparison of predictive densities $f(y \mid \mathbf{Y})$ using the prior 1.

Table 3
Comparison of predictive accuracy in terms of MAD and MARD (values$\times 10^{-2}$)

Time being forecast (T)		Approximate Bayesian		MLE	MCMC	
		Prior 1	Prior 2		Prior 1	Prior 2
10		0.7814	0.7844	0.7831	0.7735	0.7646
11		0.7992	0.7948	0.8146	0.7899	0.7937
12	MAD	1.6929	1.6850	1.7020	1.6774	1.6541
13		2.0604	2.0403	2.0740	2.0460	2.0336
Average		1.3335	1.3261	1.3434	1.3217	1.3115
10		0.5830	0.5847	0.5841	0.5752	0.5707
11		0.5955	0.5922	0.6071	0.5903	0.5919
12	MARD	1.0793	1.0736	1.0851	1.0703	1.0546
13		1.2919	1.2786	1.3004	1.2818	1.2844
Average		0.8874	0.8823	0.8942	0.8794	0.8754

interest. The extended prediction is addressed by using all the observations from all subjects before the point being forecast T in our sample. Table 3 gives the prediction comparisons for the various methods in terms of mean absolute deviation (MAD) and mean absolute relative deviation (MARD) of the predictions from the actual observations. Among the N subjects, the most recent $T-1$ observations were used to predict y_{iT} for $T=10,11,12,13$. In the table, the approximate Bayesian method using both priors perform better than the ML method. However, the best among the methods are MCMC using the prior 2, followed by MCMC using prior 1. Thus, MCMC method is quite encouraging for this model.

Table 4
Comparison of predictive accuracy for simulated samples (values $\times 10^{-2}$)

N		Approximate Bayesian			MCMC	
		Prior 1	Prior 2	MLE	Prior 1	Prior 2
10		1.0094	1.0083	1.0170	1.0056	1.0044
20	MAD	0.9810	0.9799	0.9871	0.9758	0.9697
30		0.9684	0.9673	0.9733	0.9645	0.9627
10		0.7521	0.7502	0.7568	0.7482	0.7471
20	MARD	0.7309	0.7302	0.7354	0.7274	0.7236
30		0.7224	0.7215	0.7260	0.7190	0.7176

Table 5
Comparison of coverage probabilities for $\boldsymbol{\beta} = (\beta_1, \beta_2)$ $(1 - \alpha = 0.95)$

N	Approximate Bayesian			MCMC	
	Prior 1	Prior 2	MLE	Prior 1	Prior 2
10	0.9031	0.9107	0.8886	0.9226	0.9264
20	0.9207	0.9244	0.9110	0.9310	0.9356
30	0.9330	0.9336	0.9302	0.9450	0.9462

5.2. A simulation study

In this section, a simulation study was conducted to compare the predictive ability and the coverage probability of the ML method and those obtained by the approximate Bayesian methods and MCMC methods using priors 1 and 2. The presumed simulation settings are given as the MLE's at $t = 10$ in Table 1. We generate 2000 replicates for sample size $N = 10, 20, 30$. Among the N subjects, the first $t - 1$ measurements were used as the sample to predict y_{it}. Given the values of N and t, each data set contains N subjects and for each subject t measurements were made. For each N, 2000 independent data sets were generated. Hence there were $N \times 2000$ predicted values to be compared with $N \times 2000$ true values. Table 4 gives the prediction comparisons for the various methods in terms of MAD and MARD, respectively. Better results are obtained by using MCMC samples. As for the comparison of coverage probabilities for the fixed effects $\boldsymbol{\beta}$, the results are given in Table 5. It is clear that both approximate Bayesian and MCMC methods yield much better coverage probabilities than the ML method. The best is still the MCMC among the three methods compared.

6. Concluding remarks

The Bayesian methods presented in this paper, including simple approximate Bayesian and Bayesian via MCMC sampling, provide alternative ways of dealing with the general

growth curve data when the ARMA(p, q) dependence covariance structure with random effects holds, while applying the Box–Cox transformation on the observations. The situation in which no transformation is needed and no random effect exists can be treated as special cases of model (1).

It is worth noting that the Bayesian approach is quite useful and easy to implement in analyzing the growth data when the prior is properly chosen. The forecast accuracy for future values via the simple approximate Bayesian method is better than the ML method. More accurate approximation can be obtained from MCMC samples.

Finally, in the modeling of fatigue crack growth data we favor consideration and inclusion of the time series specification, in addition to possible random effects and Box–Cox transformation, because it may lead to a more parsimonious correlated model and has the potential to provide more accurate representation for the dependence structure among repeated measurements.

References

Barndorff-Nielsen, O.E., Schou, G., 1973. On the reparameterisation of autoregressive models by partial autocorrelations. J. Multivariate Anal. 3, 408–419.

Bogdanoff, J.L., Kozin, F., 1985. Probability Model of Cumulative Damage. Wiley, New York.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations (with discussion). J. Roy. Statist. Soc. Ser. B 26, 211–252.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time Series Analysis Forecasting and Control. 3rd Edition. Holden-Day, San Francisco.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Statist. 7, 434–455.

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.

Chib, S., Carlin, E., 1999. On MCMC sampling in hierarchical models. Statist. Comput. 9, 17–26.

Chib, S., Greenberg, E., 1994. Bayes inference in regression with ARMA(p, q) errors. J. Econometrics 64, 183–206.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. Amer. Statist. 49, 327–335.

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

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

Gelman, A., Roberts, G., Gilks, W., 1995. Efficient metropolis jumping rules. (Eds.), in: .in: J.M.,Bernardo,J.O.,Berger,A.P.,Dawid,A.F.M.,Smith,(Eds.), Bayesian Statistics, Vol. 5. Oxford University Press, New York, pp. 1–5.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6, 721–741.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Jennrich, R.I., Schluchter, 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.

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

Lee, J.C., Lien, W.H., 2001. Bayesian analysis of a growth curve model with power transformation, random effects and AR(1) dependence. J. Appl. Statist. 28, 223–238.

Lee, J.C., Lu, K.W., 1987. On a family of data-based transformed models useful in forecasting technological substitutions. Technol. Forecasting Social Change 31, 61–78.

Lin, T.I., Lee, J.C., 2003. On modelling data from degradation sample paths over time. Aust. N.Z. J. Stat. 45, 257–270.

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

Lu, C.J., Meeker, W.Q., 1993. Using degradation measures to estimate a time-to-failure distribution. Technometrics 35, 161–174.

Monahan, J.F., 1984. A note on enforcing stationarity in autoregressive moving average models. Biometrika 71, 403–404.

Robinson, M.E., Crowder, M.J., 2000. Bayesian methods for a growth-curve degradation model with repeated measures. Lifetime Data Anal. 6, 357–374.

Rochon, J., 1992. ARMA covariance structures with time heteroscedasticity for repeated measures experiments. J. Amer. Statist. Assoc. 87, 777–784.

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