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

Bayesian analysis of a general growth curve model with predictions using power transformations and AR(1) autoregressive dependence

Jack C. Lee & Kuo-Ching Liu
Published online: 02 Aug 2010.

To cite this article: Jack C. Lee & Kuo-Ching Liu (2000) Bayesian analysis of a general growth curve model with predictions using power transformations and AR(1) autoregressive dependence, Journal of Applied Statistics, 27:3, 321-336, DOI: 10.1080/02664760021637

To link to this article: http://dx.doi.org/10.1080/02664760021637

Journal of Applied Statistics, Vol. 27, No. 3, 2000, 321–336

Bayesian analysis of a general growth curve model with predictions using power transformations and AR(1) autoregressive dependence

JACK C. LEE & KUO-CHING LIU, *Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan*

ABSTRACT *In this paper, we consider a Bayesian analysis of the unbalanced (general) growth curve model with AR(1) autoregressive dependence, while applying the Box–Cox power transformations. We propose exact, simple and Markov chain Monte Carlo approximate parameter estimation and prediction of future values. Numerical results are illustrated with real and simulated data.*

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 growth curve model with AR(1) autoregressive dependence and with the Box–Cox power transformations applied to the observations. It has been observed in the literature, such as Lee (1988), Lee and Lu (1987) and Keramidas and Lee (1990), that the data of biological growth and technological substitutions tend to exhibit a strong correlation across different time periods. In Lee and Lu (1987), tremendous improvement was found in predictive accuracy using the data-based transformation models for technology substitutions. This is primarily because of 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. Keramidas and Lee (1990) combined the concepts of power transformation and a generalized growth curve model with serial structure for forecasting technological substitutions based on the maximum-likelihood (ML) method when repeated measurements of short time series are available. Instead of the balanced case, we will address the problem of a growth

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

ISSN 0266-4763 print; ISSN 1360-0532 online/00/030321-16

curve when unbalanced repeated-measures data are available. This is apparently more general and of more practical importance.

The unbalanced repeated-measures growth curve model with the Box–Cox power transformations is defined as

$$\mathbf{Y}_i^{(\lambda_a)} = \mathbf{X}_i\boldsymbol{\tau}_a + \boldsymbol{\varepsilon}_i, \qquad i = 1, 2, \ldots, N_a^{\star}; N_a^{\star} = N_1 + \ldots + N_a; a = 1, \ldots, r \tag{1}$$

where τ_a is an unknown $m \times 1$ vector of regression coefficients of group a, and $\mathbf{X}_i$ is a known design matrix of rank m, where $2 \leqslant m \leqslant p_i$, and $N = N_1 + \ldots + N_r$. The disturbance terms ε_i are independent p_i-variate normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$. In general, p_i is the number of time (or spatial) points observed on measurement i; m and r, which specify the degree of polynomial in time (or space) and the number of distinct groups, respectively, are assumed known. The power transformations of Box and Cox (1964) are defined as

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

where υ_{ij} is a known constant such that $Y_{ij} + \upsilon_{ij} > 0$ for all i, j and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip_i})'$. In practice, $\upsilon_{ij} = 0$ if all Y_{ij} terms are positive. Without loss of generality, we will assume $\upsilon_{ij} = 0$ for the rest of the paper, because υ_{ij} is assumed known. The convention $\mathbf{Y}_i^{(\lambda)} = (Y_{i1}^{(\lambda)}, \ldots, Y_{ip}^{(\lambda)})'$ is adopted throughout the paper. Also, λ_a is unknown.

If all the p_i are equal, i.e. balanced repeated measurements, then equation (1) can be written as the generalized growth curve model with Box–Cox power transformations. The generalized growth curve model, when no power transformations are applied, is defined as

$$\underset{p\times N}{\mathbf{Y}} = \underset{p\times m}{\mathbf{X}}\ \underset{m\times r}{\boldsymbol{\tau}}\ \underset{r\times N}{\mathbf{A}} + \underset{p\times N}{\boldsymbol{\varepsilon}} \tag{3}$$

where $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_r)$ is unknown, $\mathbf{X}$ and $\mathbf{A}$ are known design matrices of ranks $m < p$ and $r < N$ respectively. Furthermore, the columns of ε are independent p-variate normal with mean vector $\mathbf{0}$ and common covariance matrix $\boldsymbol{\Sigma}$. The model was first proposed by Potthoff and Roy (1964), and subsequently considered by many authors, including Rao (1967, 1987), Khatri (1966), Grizzle and Allen (1969), Geisser (1970, 1980), Lee and Geisser (1972, 1975), Fearn (1975), Lee (1988), Jennrich and Schluchter (1986) and Von Rosen (1991). Lee (1988) studied equation (3) when $\boldsymbol{\Sigma}$ has AR(1) dependence, while Keramidas and Lee (1990) incorporated into the model the Box–Cox power transformations in the forecast of technology substitutions.

Lee (1988), Lee and Lu (1987) and Keramidas and Lee (1990) demonstrated repeatedly the importance of AR(1) dependence, or serial covariance structure, for the covariance matrix for equation (3). Applying the AR(1) dependence to the covariance matrix $\boldsymbol{\Sigma}_i$ for equation (1), we have

$$\boldsymbol{\Sigma}_i = \sigma^2\mathbf{C}_i \tag{4}$$

where $\mathbf{C}_i = (\rho^{|d-d'|})$; $d, d' = 1, 2, \ldots, p_i$; and $\sigma^2 > 0$ and $-1 < \rho < 1$ are unknown.

The estimation of parameters and prediction of future values for equations (3) and (4) have been studied using the ML method by Lee (1988). The purpose of this paper is to consider equation (1) from a Bayesian point of view. The results using the ML method are also included in the appendices.

We will consider two types of prediction problem for unbalanced growth curve models with power transformations as specified by equations (1), (2) and (4). For the first type of prediction, let $\mathbf{V}$ be a set of $p \times 1$ future measurements of group a, assumed to have been drawn from equation (1); i.e., the set of future values of group a is such that, given the parameters τ_a and Σ, we have

$$E(\mathbf{V}^{(\lambda_a)}) = \mathbf{X}\tau_a \tag{5}$$

where E( ) denotes the expected value, $\mathbf{X}$ is a known $p \times m$ matrix and $\mathbf{V}^{(\lambda_a)}$ is a multivariate normal with the covariance matrix Σ. We address the prediction of $\mathbf{V}$ given $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$, the unbalanced repeated-measures data, as the sample. It is noted that $\mathbf{Y}$ is no longer $p \times N$, because $\mathbf{Y}_i$ is $p_i \times 1$.

The second type of prediction problem is concerned with predicting the future values of the observed sample. Let $\mathbf{y}_j$ be a future q-dimensional vectoral measurement of group a. We are interested in predicting $\mathbf{y}_j$ given $\mathbf{Y}$. This is a time series prediction, so is important in practice. This type of prediction is called extended prediction of $\mathbf{y}_j$, because the prediction is made beyond the observed time range of the sample. It is noted that extended prediction is identical to conditional prediction as studied in Lee and Geisser (1972, 1975), Fearn (1975), and Lee (1988) when using the model specified by equations (1) and (4).

In this paper, Bayesian inference by means of numerical integration, the Markov chain Monte Carlo (MCMC) method, and some simple approximations are studied for equation (1) with AR(1) dependence. In recent years, statisticians have been increasingly drawn to MCMC methods, especially the Metropolis–Hastings (MH) algorithm (Metropolis *et al.*, 1953; Hastings, 1970) and the Gibbs sampler (Geman & Geman, 1984; Gelfand & Smith, 1990). They have emerged as extremely popular tools for the analysis of complex statistical models. While they have been most widely used in Bayesian analysis, they have also been employed by frequentists in missing and dependent data settings, where the likelihood itself involves complicated high-dimensional integrals (see, for example, Gelfand & Carlin, 1993). Excellent tutorials on the methodology have been provided recently by Casella and George (1992) and Gilks *et al.* (1996).

In Section 2, Bayesian estimation of parameters is considered for the model. In Section 3, two types of prediction problem are presented. In Section 4, Bayesian inference by means of MCMC methods is studied with a non-informative joint prior for the parameters. The results developed in this paper are illustrated in Section 5, with real and simulated data. Finally, some concluding remarks are presented in Section 6.

2 Bayesian estimation of parameters

Combining the likelihood function of σ^2, $\tau_1, \ldots, \tau_r$, $\lambda_1, \ldots, \lambda_r$ and ρ with the non-informative prior (Zellner & Tiao, 1964), i.e.

$$\prod(\tau, \sigma^2, \rho, \lambda) \propto \sigma^{-2} \tag{6}$$

and integrating with respect to σ^2 and $\tau_1, \ldots, \tau_r$, we have

$$P(\rho, \lambda | \mathbf{Y}) \propto (1-\rho^2)^{-(n-N)/2} |\mathcal{J}| B^{-(n-mr)/2} \prod_{a=1}^{r} \left| \sum_{j=N_{a-1}^{\star}+1}^{N_a^{\star}} \mathbf{X}_i' \mathbf{C}_{p_i}^{-1} \mathbf{X}_i \right|^{-1/2} \tag{7}$$

where

$$B = \sum_{a=1}^{r} \sum_{i=N_{a-1}^{\star}+1}^{N_a^{\star}} (\mathbf{Y}_i^{(\lambda_a)} - \mathbf{X}_i \hat{\tau}_a)' \mathbf{C}_i^{-1} (\mathbf{Y}_i^{\lambda_a)} - \mathbf{X}_i \hat{\tau}_a), \qquad \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)'$$

$$\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_{N_1}, \mathbf{Y}_{N_1+1}, \ldots, \mathbf{Y}_{N_r^{\star}}), \qquad N_0^{\star} = 1$$

$$\mathcal{J} = \prod_{a=1}^{r} \prod_{i=N_{a-1}^{\star}+1}^{N_a} \prod_{j=1}^{p_i} Y_{ij}^{\lambda_a - 1}, \qquad n = \sum_{i=1}^{N} p_i, N = \sum_{a=1}^{r} N_a^{\star}$$

Also, a simple approximation is

$$P(\tau_a | \mathbf{Y}) \doteq P(\tau_a | \hat{\rho}, \hat{\lambda}, \mathbf{Y}) \tag{8}$$

where $(\hat{\rho}, \hat{\lambda})$ is the mode of $P(\rho, \lambda | \mathbf{Y})$, if $P(\rho, \lambda | \mathbf{Y})$ is concentrated and nearly symmetric, as pointed out by Ljung and Box (1980). Thus, we have the following posterior distribution of τ_a, for $a = 1, \ldots, r$:

$$\tau_a | \mathbf{Y} \sim T_m(\hat{\tau}_a, [\hat{B}/(n - mr)]^{-1} \sum_{i=N_{a-1}^{\star}+1}^{N_a^{\star}} \mathbf{X}_i' \hat{\mathbf{C}}_i^{-1} \mathbf{X}_i, n - mr) \tag{9}$$

where

$$\hat{\tau}_a = \left(\sum_{i=N_{a-1}^{\star}+1}^{N_a^{\star}} \mathbf{X}_i' \hat{\mathbf{C}}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=N_{a-1}^{\star}+1}^{N_a^{\star}} \mathbf{X}_i' \hat{\mathbf{C}}_i^{-1} \mathbf{Y}_i^{(\hat{\lambda}_a)}$$

$$\hat{\mathbf{C}}_i = (\hat{\rho}^{|d-d'|}), \qquad d, d' = 1, \ldots, p_i$$

$\hat{B}$ is the value of B evaluated at $\lambda = \hat{\lambda}$ and $\rho = \hat{\rho}$, with $\hat{\rho}$, $\hat{\lambda}$ maximize $P(\rho, \lambda | \mathbf{Y})$, as given in equation (7).

It is noted that, for a $p \times 1$ vector $\mathbf{y}$, we say that $\mathbf{y} \sim T_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, the multivariate T-distribution, if its density is

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

where

$$K(\nu, p) = \Gamma\left(\frac{p + \nu}{2}\right) \Big/ \Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{p/2}$$

Similar to equation (8), we have

$$\sigma^2 | \mathbf{Y} \sim \mathrm{IG}\left(\frac{n - mr}{2}, \frac{\hat{B}}{2}\right)$$

where $\mathrm{IG}(\nu_1, \nu_2)$ is the inverse gamma distribution with parameters ν_1 and ν_2.

To compute the posterior region for τ_a of group a, for $a = 1, \ldots, r$, let $F_{1-a}(\nu_1, \nu_2)$ be the upper 100α per cent point of the F distribution. Then, we have that

$$\hat{B}^{-1} (\tau_a - \hat{\tau}_a)' \left(\sum_{i=N_{a-1}^{\star}+1}^{N_a^{\star}} \mathbf{X}_i \hat{\mathbf{C}}_i^{-1} \mathbf{X}_i \right) (\tau_a - \hat{\tau}_a) \leqslant \frac{m}{n - mr} F_{1-\alpha}(m, n - mr) \tag{10}$$

which will provide a $1 - \alpha$ posterior region for τ_a.

3 Prediction

Two types of prediction for the model specified by equations (1), (2) and (4) will be considered in this section.

3.1 Prediction of the future values $\mathbf{V}$ *of group a*

The prediction of the future vectoral measurement $\mathbf{V} = (V_1, \ldots, V_p)'$ given the sample $\mathbf{Y}$ is considered here. The density function of $\mathbf{V}^{(\lambda_a)}$ given τ_a, σ^2, ρ and λ_a is

$$f(\mathbf{V}^{(\lambda_a)} | \tau_a, \sigma^2, \rho, \lambda_a) \propto |\sigma^2 \mathbf{C}_p|^{-1/2} \exp\left[\frac{-1}{2\sigma^2}(\mathbf{V}^{(\lambda_a)} - \mathbf{X}\tau_a)' \mathbf{C}_p^{-1} (\mathbf{V}^{(\lambda_a)} - \mathbf{X}\tau_a)\right] \quad (11)$$

Upon combining equation (11) with the joint posterior density of τ, σ^2, ρ and λ given $\mathbf{Y}$ and integrating out σ^2, $\tau_1, \ldots, \tau_r$ and $\mathbf{V}^{(\lambda_a)}$, we obtain

$$P_1(\rho, \lambda | \mathbf{Y}) \propto (1-\rho^2)^{-[n+p-(N+1)]/2} |\mathcal{J}| |\mathbf{Q}^{-1/2}| \mathbf{M}|^{-1/2} \prod_{k \neq a}^{r} \left| \sum_{i=N_{k-1}^{*}+1}^{N_k^{*}} \mathbf{X}_i' \mathbf{C}_i^{-1} \mathbf{X}_j \right|^{-1/2} B^{-(n-mr)/2} \quad (12)$$

where $\mathcal{J}$ and B are defined in equation (7), and

$$\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2, \qquad \mathbf{Q}_1 = \sum_{i=N_{a-1}^{*}+1}^{N_a^{*}} \mathbf{X}_i' \mathbf{C}_i^{-1} \mathbf{X}_i, \qquad \mathbf{Q}_2 = \mathbf{X}' \mathbf{C}_p^{-1} \mathbf{X}$$

$$\mathbf{M} = \mathbf{C}_p^{-1} \mathbf{X} (\mathbf{X}' \mathbf{C}_p^{-1} \mathbf{X})^{-1} \mathbf{Q}_1 \mathbf{Q}^{-1} \mathbf{Q}_2 (\mathbf{X}' \mathbf{C}_p^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{C}_p^{-1} + \mathbf{Z}(\mathbf{Z}' \mathbf{C}_p \mathbf{Z})^{-1} \mathbf{Z}'$$

With arguments similar to equation (9), we obtain the following approximation for the predictive distribution of $\mathbf{V}^{(\lambda_a)}$:

$$\mathbf{V}^{(\lambda_a)} | \mathbf{Y} \sim T_p(\mathbf{X}\hat{\tau}_{a1}, [\hat{B}/(n-mr)]^{-1} \hat{\mathbf{M}}, n-mr) \quad (13)$$

where

$$\hat{\tau}_{a1} = \left(\sum_{i=N_{a-1}^{*}+1}^{N_a^{*}} \mathbf{X}_j' \hat{\mathbf{C}}_j^{-1} \mathbf{X}_j \right)^{-1} \sum_{i=N_{a-1}^{*}+1}^{N_a^{*}} \mathbf{X}_j' \hat{\mathbf{C}}_j^{-1} \mathbf{Y}_j^{(\hat{\lambda}_a)}$$

$\hat{B}$, $\hat{\mathbf{M}}$, $\hat{\mathbf{Q}}$ and $\hat{\mathbf{C}}_i$ are B, $\mathbf{M}$, $\mathbf{Q}$ and $\mathbf{C}_i$ evaluated at $\rho = \hat{\rho}$ and $\lambda = \hat{\lambda}$, with $\hat{\rho}$, $\hat{\lambda}$ maximize $P_1(\rho, \lambda | \mathbf{Y})$, as given in equation (12).

Next, from equation (13) and from the definition of the power transformation in equation (2), we can obtain the following approximate predictive density of $\mathbf{V}$, when $\hat{\lambda}_a \neq 0$:

$$P(\mathbf{V} | \mathbf{Y}) \doteq \mathbf{C}_{\mathbf{v}} \left[\hat{B} + \left(\frac{\mathbf{V}^{\hat{\lambda}_a} - 1}{\hat{\lambda}_a} - \mathbf{X}\hat{\tau}_{a1} \right)' \hat{\mathbf{C}}_p^{-1} \left(\frac{\mathbf{V}^{\hat{\lambda}_a} - 1}{\hat{\lambda}_a} - \mathbf{X}\hat{\tau}_{a1} \right) \right]^{-(n+p-mr)/2} |\mathcal{J}_{\mathbf{v}}| \quad (14)$$

where

$$\mathbf{C}_{\mathbf{v}} = \{ [\int [\hat{B} + (\mathbf{V}^{(\lambda_a)} - \mathbf{X}\hat{\tau}_{a1})' \hat{\mathbf{C}}_p^{-1} (\mathbf{V}^{(\lambda_a)} - \mathbf{X}\hat{\tau}_{a1})]^{-(n+p-mr)/2} \mathrm{d}\mathbf{V}^{(\lambda_a)} \}^{-1}$$

$$|\mathcal{J}_{\mathbf{v}}| = \prod_{k=1}^{p} V_k^{\hat{\lambda}_a - 1}$$

Therefore, when $\hat{\lambda}_a \neq 0$, we can predict $\mathbf{V}$ by the approximate predictor

$$\hat{\mathbf{V}} = (1 + \hat{\lambda}_a \mathbf{X}\hat{\tau}_{a1})^{1/\hat{\lambda}_a} \tag{15}$$

Similar to equation (10), when $\hat{\lambda}_a \neq 0$, we have a $1-\alpha$ prediction region for $\mathbf{V}$ by the inequality

$$\hat{B}^{-1}\left(\frac{\mathbf{V}^{\hat{\lambda}_a}-\mathbf{1}}{\hat{\lambda}_a} - \mathbf{X}\hat{\tau}_{a1}\right)'\hat{\mathbf{C}}_p^{-1}\left(\frac{\mathbf{V}^{\hat{\lambda}_a}-\mathbf{1}}{\hat{\lambda}_a} - \mathbf{X}\hat{\tau}_{a1}\right) \leqslant \frac{p}{n-mr}F_{1-\alpha}(p, n-mr) \tag{16}$$

For the case when $\hat{\lambda}_a = 0$, the results are similar, so are omitted here.

3.2 Prediction of future observations when their past is available

We next consider prediction of $\mathbf{y}_j$, a future q-dimensional value of $\mathbf{Y}_j$, given the observed unbalanced repeated measurements $\mathbf{Y} = (\mathbf{Y}_j, \mathbf{Y}_{(j)})$, where $\mathbf{Y}_{(j)} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_{j-1}, \mathbf{Y}_{j+1}, \ldots, \mathbf{Y}_N)$ and $j = 1, \ldots, N_a^\star; a = 1, \ldots, r$. This is a time series prediction which is of practical interest for many types of growth curve data. To make this type of prediction, the covariance structure generally has to be extendable to the future values of the individuals observed. The AR(1) dependence satisfies this requirement.

Let $\mathbf{x}_j, q \times m$, be a design matrix corresponding to $\mathbf{y}_j$; $\mathbf{Y}_j$ is in group a. Also, $E(\mathbf{Y}_j^{(\lambda_a)\prime}, \mathbf{y}_j^{(\lambda_a)\prime})' = (\mathbf{X}_j', \mathbf{x}_j')'\tau_a$ and $\mathbf{\Sigma} = \mathrm{Cov}(\mathbf{Y}_j^{(\lambda_a)\prime}, \mathbf{y}_j^{(\lambda_a)\prime})' = \sigma^2\mathbf{C} = \sigma^2(\mathbf{C}_{ij}), i, j = 1,2$, where $\mathbf{C} = (\rho^{|d-d'|})$, $d, d' = 1, \ldots, (p_j + q)$, $\mathbf{C}_{11}$ is $p_j \times p_j$, $\mathbf{C}_{12}$ is $p_j \times q$, $\mathbf{C}_{22}$ is $q \times q$, and $\mathbf{C}_{21} = \mathbf{C}_{12}'$.

Let $\mathbf{Y}_j^{(\lambda_a)^\star} = (\mathbf{Y}_j^{(\lambda_a)\prime}, \mathbf{y}_j^{(\lambda_a)\prime})', \mathbf{X}^\star = (\mathbf{X}_j', \mathbf{x}_j')'$. Arguments similar to those in Section 3.1 lead to the following approximate predictive density of $\mathbf{y}_j$:

$$P(\mathbf{y}_j|\mathbf{Y}) \doteq C_y\left[\hat{B}_1 + \left(\frac{\mathbf{y}_j^{\hat{\lambda}_a}-\mathbf{1}}{\hat{\lambda}_a} - \boldsymbol{\mu}_y^{(\hat{\lambda}_a)}\right)'\hat{\mathbf{G}}_{22}\left(\frac{\mathbf{y}_j^{\hat{\lambda}_a}-\mathbf{1}}{\hat{\lambda}_a} - \boldsymbol{\mu}_y^{(\hat{\lambda}_a)}\right)\right]^{-(n+q-mr)/2}|\mathcal{J}_\mathbf{y}| \tag{17}$$

where

$$\mathbf{C}_\mathbf{y} = \left\{\int [\hat{B}_1 + (\mathbf{y}_j^{(\lambda_a)} - \boldsymbol{\mu}_y^{(\hat{\lambda}_a)})'\hat{\mathbf{G}}_{22}(\mathbf{y}_j^{(\lambda a)} - \boldsymbol{\mu}_y^{(\hat{\lambda}_a)})]^{-(n+q-mr)/2}\, d\mathbf{y}_j^{(\lambda a)}\right\}^{-1}$$

$$|\mathcal{J}_\mathbf{y}| = \prod_{k=1}^{q} y_{jk}^{\hat{\lambda}_a - 1}$$

$$\hat{\mathbf{G}} = \hat{\mathbf{C}}^{-1}\mathbf{X}^\star(\mathbf{X}^{\star\prime}\hat{\mathbf{C}}^{-1}\mathbf{X}^\star)^{-1}\hat{\mathbf{Q}}_1^\star\hat{\mathbf{Q}}^{\star-1}\hat{\mathbf{Q}}_2^\star(\mathbf{X}^{\star\prime}\hat{\mathbf{C}}^{-1}\mathbf{X}^\star)^{-1}\mathbf{X}^{\star\prime}\hat{\mathbf{C}}^{-1} + \mathbf{Z}^\star(\mathbf{Z}^{\star\prime}\hat{\mathbf{C}}\mathbf{Z}^\star)^{-1}\mathbf{Z}^{\star\prime}$$

$$= \begin{pmatrix} \hat{\mathbf{G}}_{11} & \hat{\mathbf{G}}_{12} \\ \hat{\mathbf{G}}_{21} & \hat{\mathbf{G}}_{22} \end{pmatrix}$$

$$\mathbf{Q}_1^\star = \sum_{\substack{k=N_{a-1}^\star+1 \\ k\neq j}}^{N_a^\star} \mathbf{X}_k'\mathbf{C}_k^{-1}\mathbf{X}_k, \qquad \mathbf{Q}_2^\star = \mathbf{X}^{\star\prime}\mathbf{C}^{-1}\mathbf{X}^\star, \qquad \mathbf{Q}^\star = \mathbf{Q}_1^\star + \mathbf{Q}_2^\star$$

$$\hat{\mathbf{G}}_{11.2} = \hat{\mathbf{G}}_{11} - \hat{\mathbf{G}}_{12}\hat{\mathbf{G}}_{22}^{-1}\hat{\mathbf{G}}_{21}, \qquad \hat{\boldsymbol{\mu}}_y^{(\hat{\lambda}_a)} = \mathbf{x}\hat{\tau}_{a1}^\star - \hat{\mathbf{G}}_{22}^{-1}\hat{\mathbf{G}}_{21}(\mathbf{Y}_j^{(\hat{\lambda}_a)} - \mathbf{X}_j\hat{\tau}_{a1}^\star)$$

$$\hat{B}_1 = \hat{B} + (\mathbf{Y}_j^{(\hat{\lambda}_a)} - \mathbf{X}_j\hat{\tau}_{a1}^\star)'\hat{\mathbf{G}}_{11.2}(\mathbf{Y}_j^{(\hat{\lambda}_{al})} - \mathbf{X}_j\hat{\tau}_{a1}^\star)$$

$$\hat{\tau}_{a1}^{\star} = \left(\sum_{\substack{i=N_{a-1}^{\star}+1 \\ i \neq j}}^{N_a^{\star}} \mathbf{X}_i' \hat{\mathbf{C}}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{\substack{i=N_{a-1}^{\star}+1 \\ i \neq j}}^{N_a^{\star}} \mathbf{X}_i' \hat{\mathbf{C}}_i^{-1} \mathbf{Y}_i^{(\hat{\lambda}_a)}$$

$\mathbf{Z}^{\star}$, $(p_j + q) \times ((p_j + q) - m)$, such that $\mathbf{X}^{\star}\mathbf{Z}^{\star} = \mathbf{0}$, and $\hat{\rho}$, $\hat{\lambda}$ maximize

$$P_2(\rho, \lambda | \mathbf{Y}) \propto (1 - \rho^2)^{-(n+q-N)/2} |\mathfrak{J}| |\mathbf{Q}^{\star}|^{-1/2} |\mathbf{G}_{22}|^{-1/2} \prod_{i \neq a}^{r} \left| \sum_{j=1}^{N_i} \mathbf{X}_j' \mathbf{C}_{p_j}^{-1} \mathbf{X}_j \right|^{-1/2} B_1^{-(n-mr)/2} \tag{18}$$

Therefore, we can predict $\mathbf{y}_j$ by the approximate predictor

$$\hat{\mathbf{y}}_j = (1 + \hat{\lambda}_a \boldsymbol{\mu}_{\mathbf{y}}^{(\hat{\lambda}_a)})^{1/\hat{\lambda}_a} \tag{19}$$

and a $1 - \alpha$ predictive region for $\mathbf{y}_j$ from the inequality

$$\hat{B}_1^{-1} \left(\frac{\mathbf{y}_j^{\hat{\lambda}_a} - \mathbf{1}}{\hat{\lambda}_a} - \boldsymbol{\mu}_{\mathbf{y}}^{(\hat{\lambda}_a)} \right)' \hat{\mathbf{G}}_{22} \left(\frac{\mathbf{y}_j^{\hat{\lambda}_a} - \mathbf{1}}{\hat{\lambda}_a} - \boldsymbol{\mu}_{\mathbf{y}}^{(\hat{\lambda}_a)} \right) \leqslant \frac{q}{n - mr} F_{1-\alpha}(q, n - mr) \tag{20}$$

Another approximate posterior density is the Rao–Blackwellization approximation

$$P(\mathbf{y}_j | \mathbf{Y}) \cong \frac{1}{L} \sum_{s=1}^{L} P(\mathbf{y}_j | \mathbf{Y}, \rho^{(s)}, \lambda^{(s)}) \tag{21}$$

where $\rho^{(s)}, \lambda^{(s)}$ are drawn from $P_2(\rho, \lambda | \mathbf{Y})$. Using this method, we can obtain the approximate predictor and predictive region of $\mathbf{y}_j$ as well.

We can also obtain the following joint density of ρ, λ and $\mathbf{y}_j$:

$$\begin{aligned} P(\mathbf{y}_j, \rho, \lambda | \mathbf{Y}) &= C_{\mathbf{y}}^{\star} (1 - \rho^2)^{-(n+q-N)/2} |\mathfrak{J}| |\mathbf{Q}^{\star}|^{-1/2} \\ &\prod_{k \neq a}^{r} \left| \sum_{j=N_{k-1}^{\star}+1}^{N_k^{\star}} \mathbf{X}_j' \mathbf{C}_j^{-1} \mathbf{X}_j \right|^{-1/2} B_2^{-(n+q-mr)/2} |\mathfrak{J}_{\mathbf{y}}| \end{aligned} \tag{22}$$

where

$$B_2 = B + (\mathbf{Y}_j^{(\lambda_a)} - \mathbf{X}_j \tau_{a1}^{\star})' \mathbf{G}_{11.2} (\mathbf{Y}_j^{(\lambda_a)} - \mathbf{X}_j \tau_{a1}^{\star}) + \left(\frac{\mathbf{y}_j^{\lambda_a} - \mathbf{1}}{\lambda_a} - \boldsymbol{\mu}_{\mathbf{y}}^{(\lambda_a)} \right)' \mathbf{G}_{22} \left(\frac{\mathbf{y}_j^{\lambda_a} - \mathbf{1}}{\lambda_a} - \boldsymbol{\mu}_{\mathbf{y}}^{(\lambda_a)} \right)$$

$\boldsymbol{\mu}_{\mathbf{y}}^{(\lambda_a)}$ is $\boldsymbol{\mu}_{\mathbf{y}}^{(\hat{\lambda}_a)}$ with $\hat{\lambda}_a$ replaced by λ_a, and $\tau_{a1}^{\star}$ is $\hat{\tau}_{a1}^{\star}$ with $\hat{\lambda}_a$ and $\hat{\rho}$ replaced by λ_a and ρ, respectively, and $C_{\mathbf{y}}^{\star}$ is a normalized constant.

From equation (22), we can obtain the following exact predictive density of $\mathbf{y}_j$:

$$P(\mathbf{y}_j | \mathbf{Y}) = \int\int P(\mathbf{y}_j, \rho, \lambda | \mathbf{Y}) \, \mathrm{d}\rho \, \mathrm{d}\lambda \tag{23}$$

4 Bayesian inference via MCMC sampling

The MH algorithm is an extremely powerful method that can be used in conjunction with the Gibbs sampler to extract marginal distributions of interest. Some details can be found in Chib and Greenberg (1995).

4.1 Model and algorithm

In this section, we will outline an MCMC sampling procedure with equation (1). From the joint posterior density of τ, σ^2, ρ, and λ given the sample $\mathbf{Y}$, the MCMC sampling proceeds as follows

(1) Generate τ_a given σ^2, ρ, λ_a and $\mathbf{Y}$ from

$$N_m\left(\hat{\tau}_a, \sigma^2\left(\sum_{j=N^*_{a-1}+1}^{N^*_a} \mathbf{X}_j' \mathbf{C}_j^{-1} \mathbf{X}_j\right)^{-1}\right) \tag{24}$$

(2) Generate σ^2 given τ, ρ, λ and $\mathbf{Y}$ from the inverse gamma distribution

$$\mathrm{IG}\left(\frac{n}{2}, \frac{S(\tau, \rho, \lambda, \mathbf{Y})}{2}\right) \tag{25}$$

where

$$S(\tau, \rho, \lambda, \mathbf{Y}) = \sum_{a=1}^{r} \sum_{i=N^*_{a-1}+1}^{N^*_a} (\mathbf{Y}_i^{(\lambda_a)} - \mathbf{X}_i \tau_a)' \mathbf{C}_i^{-1} (\mathbf{Y}_i^{(\lambda_a)} - \mathbf{X}_i \tau_a)$$

(3) Generate ρ given τ, σ^2, λ and $\mathbf{Y}$ using the MH algorithm from

$$f(\rho) \propto (1-\rho^2)^{-(n-N)/2} \exp\left[-\frac{S(\tau, \rho, \lambda, \mathbf{Y})}{2\sigma^2}\right] \tag{26}$$

(4) Generate λ_a given τ_a, ρ, σ^2 and $\mathbf{Y}$ using the MH algorithm, where

$$f(\lambda_a) \propto |\mathcal{J}_a| \exp\left[-\frac{1}{2\sigma^2} \sum_{j=N^*_a+1}^{N^*_a} (\mathbf{Y}_j^{(\lambda_a)} - \mathbf{X}_j \tau_a)' \mathbf{C}_j^{-1} (\mathbf{Y}_j^{(\lambda_a)} - \mathbf{X}_j \tau_a)\right] \tag{27}$$

with

$$\mathcal{J}_a = \prod_{j=N^*_a+1}^{N^*_a} \prod_{k=1}^{p_j} Y_{jk}^{\lambda_a - 1} \tag{28}$$

To elaborate on the MH algorithm in step (3), let us assume that the prior on ρ is uniform over $(-1, 1)$. We can transform ρ to $\rho' \in (-\infty, \infty)$ by

$$\rho' = \log \frac{1+\rho}{1-\rho}$$

Then, we apply the MH algorithm to the function

$$g(\rho') \propto \frac{\exp[\rho'(n+2-N)/2]}{(1+\exp(\rho'))^{n+2-N}} \exp\left[-\frac{S'(\tau, \rho', \lambda, \mathbf{Y})}{2\sigma^2}\right] \tag{29}$$

where $S'(\tau, \rho', \lambda, \mathbf{Y})$ is obtained from $S(\tau, \rho, \lambda, \mathbf{Y})$ with ρ replaced by $[\exp(\rho') - 1]/[\exp(\rho') + 1]$. We also need to specify $\sigma_{\rho'}$ in the transition kernel. The quantity $\sigma_{\rho'}$ is usually chosen to reflect the conditional standard deviation of ρ' given $\tau, \sigma^2, \lambda, \mathbf{Y}$. We use the following method to estimate the variance $\sigma^2_{\rho'}$.

From equation (29), let $l(\rho'; \tau, \sigma^2, \lambda, \mathbf{Y}) = \log[g(\rho'; \tau, \sigma^2, \lambda, \mathbf{Y})]$. Then, from this,

invert the sample information given the ρ' value in the MCMC sampling to obtain the preliminary variance estimates of ρ', and put it into the MCMC procedure. Having obtained ρ' from the MH algorithm, we transform ρ' back to ρ by $[\exp(\rho') - 1]/[\exp(\rho') + 1]$. The same operation can also be applied to the variate λ_a with the prior on λ_a being uniform over $(-4, 4)$.

4.2 Forecast

Having obtained the posterior distribution of the unknown parameters from the above MCMC samples, we can use it to predict the future values $\mathbf{V}$ of group a and extended future q-dimensional values $\mathbf{y}_j$ of group a. We will illustrate the process of finding the functional, such as predictive density, estimator and intervals and quantiles of the future values $\mathbf{y}_j$.

From Section 3.2, we obtain

$$f(\mathbf{y}_j^{(\lambda_a)} | \theta, \mathbf{Y}_j) = \text{Const} \exp\left[\frac{-1}{2\sigma^2}\left(\frac{\mathbf{y}_j^{(\lambda_a)} - \mathbf{1}}{\lambda_a} - \boldsymbol{\mu}_{2.1}\right)' \mathbf{C}_{22.1}^{-1}\left(\frac{\mathbf{y}_j^{(\lambda_a)} - \mathbf{1}}{\lambda_a} - \boldsymbol{\mu}_{2.1}\right)\right] \quad (30)$$

where $\theta = (\tau_a, \sigma^2, \rho, \lambda_a)$, $\boldsymbol{\mu}_{2.1} = \mathbf{x}\tau_a + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{Y}_j^{(\lambda_a)} - \mathbf{X}_j\tau_a)$, $\mathbf{C}_{22.1} = \mathbf{C}_{22} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$, $|\mathcal{J}_{\mathbf{y}_j}| = \boldsymbol{\Pi}_{k=1}^{q} y_{jk}^{\lambda_a - 1}$.

Prediction for the functional $\mathbf{y}_j$ follows from the predictive density

$$f(\mathbf{y}_j | \mathbf{Y}_j) = \int f(\mathbf{y}_j | \theta, \mathbf{Y}_j)\pi(\theta | \mathbf{Y})\, \mathrm{d}\theta \quad (31)$$

This density can be approximated by Monte Carlo integration from the MCMC samplers

$$\hat{f}(\mathbf{y}_j | \mathbf{Y}_j) \doteq \frac{1}{r}\sum_{s=1}^{r} f(\mathbf{y}_j | \theta^{(k,s)}, \mathbf{Y}_j) \quad (32)$$

where $\theta^{(k,s)}$ is the variate of θ drawn in the kth iteration and sth replication of the MCMC sampler. The mean of this predictive distribution is computed from

$$E(\mathbf{y}_j | \mathbf{Y}) = E(E(\mathbf{y}_j | \theta, \mathbf{Y}_j) | \mathbf{Y}) \quad (33)$$

To evaluate the inner expectation in equation (33), let us consider two cases. Case 1 is $\lambda_a \neq 0$.

It follows from equations (2) and (29) that

$$\mathbf{y}_j = (\mathbf{y}_j^{(\lambda_a)}\lambda_a + \mathbf{1})^{1/\lambda_a} \quad (34)$$

where $\mathbf{y}_j^{(\lambda_a)} = \boldsymbol{\mu}_{2.1} + \xi$, $\xi \sim N_q(\mathbf{0}, \sigma^2\mathbf{C}_{22.1})$. Therefore, we have

$$\mathbf{y}_j = [(\boldsymbol{\mu}_{2.1} + \xi)\lambda_a + \mathbf{1}]^{1/\lambda_a} \quad (35)$$

Case 2 is $\lambda_a = 0$.

We can derive the following equation by a similar method:

$$\mathbf{y}_j = \exp(\boldsymbol{\mu}_{2.1} + \xi) \quad (36)$$

where $\boldsymbol{\mu}_{2.1} = \mathbf{x}\boldsymbol{\tau}_a + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}[\log(\mathbf{Y}_j) - \mathbf{X}_j\boldsymbol{\tau}_l]$ and $\exp(\mathbf{a}) = (\exp(a_1), \ldots, \exp(a_q))'$ for $\mathbf{a} = (a_1, \ldots, a_q)'$.

Let $\mathbf{y}_j^{(k,s)}$ denote the functional in equation (35) evaluated at the kth iteration and sth replication of the samplers, noting that $\xi^{(k,s)}$ is generated from the $N_q(\mathbf{0}, \sigma^{2(k,s)}\mathbf{C}_{22.1}^{(k,s)})$ distribution in each of the MCMC samples. Therefore,

when $\lambda_a \neq 0$, we predict $\mathbf{y}_j$ by

$$\hat{\mathbf{y}}_j = \frac{1}{r}\sum_{s=1}^{r} \mathbf{y}_j^{(k,s)} \tag{37}$$

Alternatively, we can also predict $\mathbf{y}_j$ using the median of the MCMC samplers

$$\tilde{\mathbf{y}}_j = \text{median}(\{\mathbf{y}_j^{(k,s)}\}_{s=1}^{r}) \tag{38}$$

Prediction intervals and quantiles of the functional $\mathbf{y}_j$ can be computed similarly from the samplers $\mathbf{y}_j^{(k,s)}$, $s = 1, \ldots, r$. Likewise, when $\lambda_a = 0$, we can use the same method to predict $\mathbf{y}_j$.

Theoretically, the posterior probability of the case $\lambda_a = 0$ is 0. However, we had to program the two cases ($\lambda_a = 0$ and $\lambda_a \neq 0$) of the forecast rule separately to avoid an overflow problem.

The forecast of the future values $\mathbf{V}$ can be done quite similarly, so is omitted.

5 Numerical illustration

5.1 Applications of technological substitutions with concurrent short time series

In this section, we will apply equation (1), as specified in Section 1, to a set (Region B) of telephone switching data studied by Keramidas and Lee (1990). These data are obtained from a region in the USA consisting of five states. Since, in the area of technological substitutions, the logistic growth curve is most popular, we will restrict our attention to this particular model in which the variable Y_{jt} is defined as

$$Y_{jt} = F_j(t)/[1 - F_j(t)] \tag{39}$$

where $F_j(t)$ denotes the new technology penetration at time t of measurement j. The technology penetration is the fraction of the total number of new technology users divided by the total number of new and old technology users. Therefore, the general growth curve model with power transformation as described in equations (1), (2) and (4) can be applied for technological substitutions.

It is noted that this data set has been studied carefully by Keramidas and Lee (1990). For this paper, we will restrict our attention to the special situation in which $r = 1$, i.e. there is only one group, and apply the transformation in equation (2) to Y_{jt} defined in equation (39). Next, as $q = 1$, we show in Fig. 1 a comparison of approximate and exact predictive densities of $\mathbf{y}$, the last year in state 5, given the entire data except the last year in that state as our sample $\mathbf{Y}$. Note that the exact predictive density of $\mathbf{y}$ is obtained by integrating out w.r.t. ρ and λ via numerical integration as given in equation (23). Figure 2 exhibits that the joint posterior density of ρ and λ given $\mathbf{Y}$ is well concentrated and nearly symmetric. Hence, the approximation given by equation (17) for the predictive distribution $\mathbf{y}$ should be quite adequate.

For the results of Bayesian inference via MCMC sampling, Table 1 lists the estimates of the standard deviation, and the percentiles for each of the parameters for the entire data with Y_{jt} given in equation (39) except year 12 of all states. The Bayes estimates are computed from the MCMC samplers with 50 iterations and 522 replications. Moreover, 50 loops are carried out in each MH algorithm. The starting points for the replications for each parameter are chosen from random perturbations around the maximum likelihood estimates as developed in Appendix

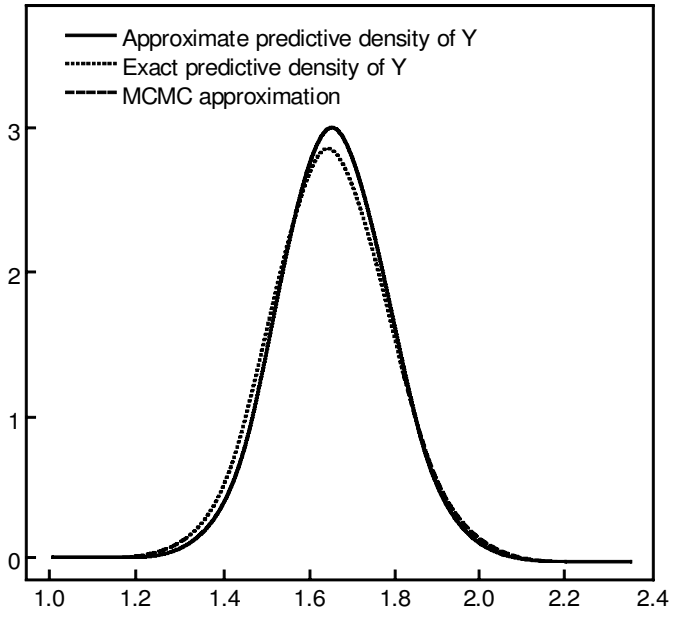

FIG. 1. Comparison of exact and approximate predictive densities of **y** given **Y**. Here, **Y** is the year 12 in state 5 given the entire data set except the last year in that state as the sample.

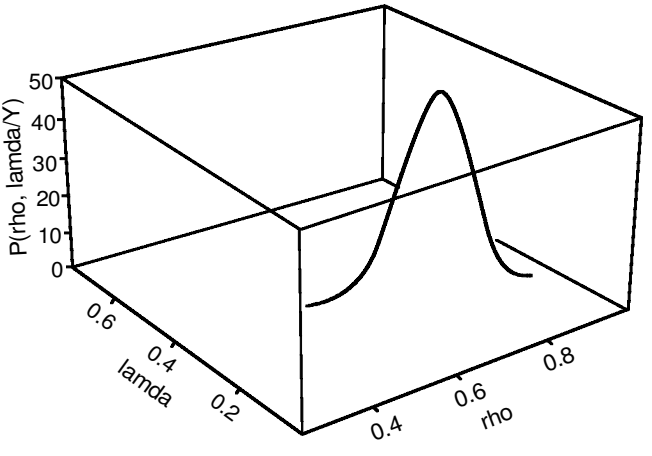

FIG. 2. Posterior of ρ and λ given **Y**. The sample is the same as in Fig. 1.

TABLE 1. MCMC approximations for switching data

	Mean	SD	2.5%	5%	25%	50%	75%	95%	97.5%
τ	−2.146, 0.210	0.143, 0.015	−2.419, 0.183	−2.370, 0.187	−2.242, 0.199	−2.142, 0.210	−2.054, 0.219	−1.926, 0.234	−1.863, 0.239
ρ	0.785	0.083	0.619	0.644	0.731	0.791	0.845	0.915	0.928
λ	0.356	0.048	0.257	0.275	0.324	0.357	0.386	0.441	0.461
σ^2	0.028	0.016	0.013	0.014	0.019	0.024	0.031	0.057	0.069

TABLE 2. MCMC approximate predictive means, the standard deviations, and the percentiles for five states when forecasting the penetration of year 12

State	Mean	SD	2.5%	50%	97.5%
1	0.5735	0.0225	0.5268	0.5734	0.6150
2	0.6043	0.0198	0.5635	0.6054	0.6412
3	0.5872	0.0215	0.5428	0.5885	0.6273
4	0.5687	0.0217	0.5244	0.5685	0.6076
5	0.6229	0.0193	0.5838	0.6219	0.6607

A. The convergence of the MCMC samplers is monitored by examining their empirical quantiles and the measure, $\sqrt{\hat{R}}$, proposed by Gelman and Rubin (1992). It is noted that the choice of starting points can reflect the speed of convergence for the parameters. Therefore, adequate starting values will accelerate the rate of convergence for the estimation of parameters. The MCMC posterior intervals can easily be obtained from Table 1. For example, the 95% intervals are read from the 2.5% and 97.5% columns.

By the MCMC samplers mentioned above, we can also obtain the approximate conditional predictive density of $\mathbf{y}$ as developed in Section 4. Figure 1 shows a comparison of exact density, MCMC approximation and the simple Bayesian approximation given by equation (17) for the predictive density of $\mathbf{y}$. It is clear that the approximation via MCMC sampling is better than the simple Bayesian approximation given by equation (17), although equation (17) performs quite well. The Rao–Blackwellization approximate predictive density, which is given by equation (11), is expected to give excellent approximation as well.

Table 2 lists the MCMC approximate predictive means, the standard deviation and the percentiles of the penetration for each of the five states when year 12 is being forecast. Figure 3 shows a comparison of 95% predictive intervals for the penetration of year 12 by three Bayesian methods: simple approximate, MCMC and exact, when the data consist of the entire data set except year 12 of all states. It is clear that the intervals via the simple approximate Bayesian method have a slight upward-shift from those produced by MCMC sampling. Meanwhile, the intervals via MCMC sampling are almost the same as the exact.

5.2 Simulations

In this subsection, we will present some simulation results using cross-validation (predictive sample reuse or 'leave-one-out'), via the simple approximate Bayesian and ML methods, for the comparison of 95% predictive intervals of y given $\mathbf{Y}$, the simulated data except the measurement of the last value being forecast. Here, we set $p = 6$, $q = 1$, $1 - \alpha = 0.95$, $\tau = (-1.2, 0.2)'$, $\sigma^2 = 0.02$, $\rho = 0.85$, $\lambda = 0.7$, $N = 5$, 10, 20, and the number of replication $g = 50$. This gives N predicted intervals for the last N observed values in each data set. Overall, there are $50 \times N$ predicted intervals to be compared with $50 \times N$ actual observations for each method. Table 3 lists a comparison of coverage probability for $N = 5$, 10 and 20. It is clear that the simple approximate Bayesian method is much better than the ML method, because the percentage of the Bayesian intervals covering the true values is closer to 0.95 than ML intervals for each of the three situations. It is noted that the coverage probability by the ML method will be closer to 0.95 as N increases. Therefore, the simple approximate Bayesian method tends to provide more reliable

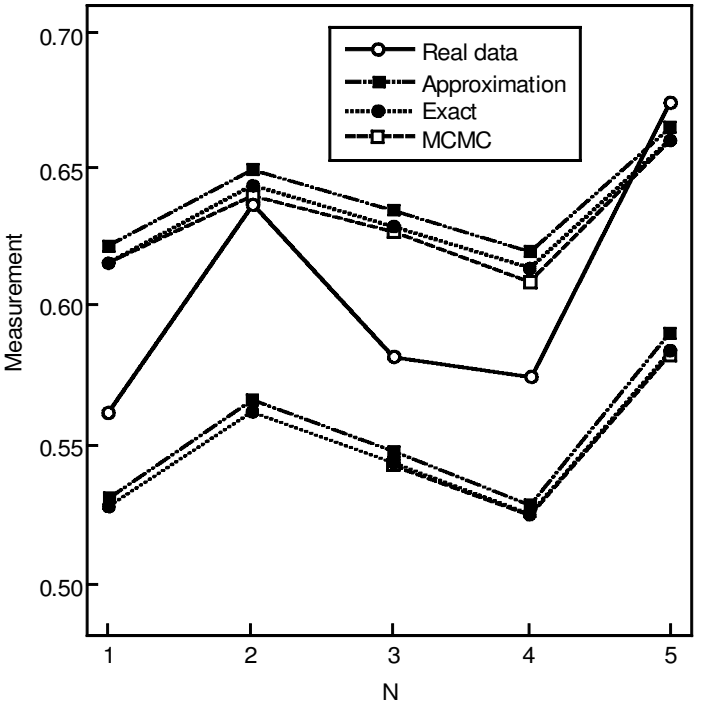

FIG. 3. Comparison of 95% predictive intervals for the penetration of year 12. The data consist of the entire data set except year 12 of all states.

TABLE 3. Comparison of coverage probabilities for the prediction of y

	Coverage probability	
	Bayesian	ML
N = 5	0.932	0.888
N = 10	0.932	0.914
N = 20	0.946	0.935

Here, $p = 6, 1 - \alpha = 0.95$, $\tau = (-1, 2, 0.2)'$, $\sigma^2 = 0.02$, $\rho = 0.85$, $\lambda = 0.7$, $q = 1$ and no. of replications = 50.

predictive intervals than the ML method when the sample size is moderate or small. Better results can be obtained by using MCMC samplers.

6 Concluding remarks

The Bayesian methods presented in this paper, including simple approximate Bayesian and Bayesian via MCMC sampling, provides alternative ways of dealing with the general (balanced or unbalanced) growth curve data when the serial covariance structure holds, while applying the Box–Cox power transformation on the observations. The model as specified in equations (1), (2) and (4) can be very useful in forecasting technological substitutions for concurrent short time series.

The serial covariance structure (AR(1) dependence) is definitely an important dependence structure for the unbalanced repeated measures.

It is noted that the Bayesian methods presented in this paper provide superior ways of constructing more reliable predictive intervals and regions for the future values. Meanwhile, the forecast accuracy for the future values via the simple approximate Bayesian method is better than the ML method. More accurate approximation can be obtained from MCMC samplers. Furthermore, the computations involved are not difficult and do not take much running time. It is noted that all computing results in this paper are conducted in the S-plus environment.

Finally, it is fair to say that the proposed approximate Bayesian methods with power transformation and AR(1) dependence covariance structure should be quite useful for practitioners in dealing with forecasting technological substitutions with concurrent short time series, as well as with other data in which the transformation is helpful. The results are useful for the situation in which no transformation is needed as well, because it will simply be a special case.

REFERENCES

Box, G. E. P. & Cox, D. R. (1964) An analysis of transformation (with discussion), *Journal of the Royal Statistical Society, Series A*, 26, pp. 211–252.

Casella, G. & George, E. I. (1992) Explaining the Gibbs Sampler, *American Statistician*, 46, pp. 167–174.

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

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

Geisser, S. (1970) Bayesian analysis of growth curves, *Sankhya, Series A*, 32, pp. 53–64.

Geisser, S. (1980) Growth curve analysis. *In:* P. R. Krishnaiah (Ed.), *Handbook of Statistics*, Vol. 1, pp. 89–115 (Amsterdam, North-Holland).

Gelfand, A. E. & Carlin, B. P. (1993) Maximum likelihood estimation for constrained or missing data models, *Canadian Journal of Statistics*, 21, pp. 303–312.

Gelfand, A. E. & Smith, A. F. M. (1990) Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, pp. 398–409.

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. T. (Eds) (1996) *Markov Chain Monte Carlo in Practice* (London, Chapman and Hall).

Grizzle, J. E. & Allen, D. M. (1969) Analysis of growth and dose response curves, *Biometrics*, 25, pp. 357–381.

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, *Journal of the American Statistical Association*, 85, pp. 625–632.

Khatri, C. G. (1966) A note on a MANOVA model applied to problems in growth curves, *Annals of the Institute of Statistical Mathematics*, 18, pp. 75–86.

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. & Geisser, S. (1972) Growth curve prediction, *Sankhya, Series A*, 34, pp. 393–412.

Lee, J. C. & Geisser, S. (1975) Applications of growth curve prediction, *Sankhya, Series 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.

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. (1967) Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *In:* L. M. LECAM & J. NEYMAN (Eds), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1), pp. 355–372 (Berkeley, University of California Press).

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

VON ROSEN, D. (1991) The growth curve model: a review, *Comm. Stat.*, 20, pp. 2791–2822.

ZELLNER, A. & TIAO, G. C. (1964) Bayesian analysis of regression model with autocorrelated errors, *Journal of the American Statistical Association*, 59, pp. 763–778.

Appendix A: The ML estimates of the parameters

The ML estimates of parameters τ, σ^2, ρ and λ denoted as $\hat{\tau}$, $\hat{\sigma}^2$, $\hat{\rho}$ and $\hat{\lambda}$ are

$$\hat{\tau}_a = \left(\sum_{j=N_{a-1}^{*}+1}^{N_a^{*}} \mathbf{X}_j' \hat{\mathbf{C}}_j^{-1} \mathbf{X}_j \right)^{-1} \sum_{j=N_{a-1}^{*}+1}^{N_a^{*}} \mathbf{X}_j' \hat{\mathbf{C}}_j^{-1} \mathbf{Y}_j^{(\hat{\lambda}_a)} \tag{A1}$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \left[\sum_{a=1}^{r} \sum_{j=N_{a-1}^{*}+1}^{N_a^{*}} (\mathbf{Y}_j^{(\hat{\lambda}_a)} - \mathbf{X}_j \hat{\tau}_a)' \hat{\mathbf{C}}_j^{-1} (\mathbf{Y}_j^{(\hat{\lambda}_a)} - \mathbf{X}_j \hat{\tau}_a) \right] \tag{A2}$$

with n defined in equation (7), $\hat{\mathbf{C}}_j = (\hat{\rho}^{|a-b|})$, $a, b = 1, \ldots, p_j$ and $\hat{\rho}, \hat{\lambda}_1, \ldots, \hat{\lambda}_r$ maximize the profile likelihood function

$$L_{\max}(\rho, \lambda_1, \ldots, \lambda_r) = [\hat{\sigma}^2(\rho, \lambda_1, \ldots, \lambda_r)]^{-n/2} (1-\rho^2)^{-(n-N)/2} |\mathcal{J}| \tag{A3}$$

where $\mathcal{J}$, the Jacobian of the power transformation, is defined in equation (7), and $\hat{\sigma}^2(\rho, \lambda_1, \ldots, \lambda_r)$ is the $\hat{\sigma}^2$ given in equation (A2) with $\hat{\rho}, \hat{\lambda}_1, \ldots, \hat{\lambda}_r$ replaced by $\rho, \lambda_1, \ldots, \lambda_r$ respectively.

Appendix B: Prediction of $\mathbf{y}_j$ given Y with the ML method

With the same assumption as described in Section 3.2, we have as our predictor for $\mathbf{y}_j$, given $\mathbf{Y}$

$$\begin{aligned} \hat{\mathbf{y}}_j &= \{\mathbf{1} + \hat{\lambda}_a [\mathbf{x}\hat{\tau}_a + \hat{\mathbf{C}}_{21} \hat{\mathbf{C}}_{11}^{-1} (\mathbf{Y}_j^{(\hat{\lambda}_a)} - \mathbf{X}_j \hat{\tau}_a)]\}^{1/\hat{\lambda}_a}, \qquad \hat{\lambda}_a \neq 0 \\ &= \exp\{\mathbf{x}\hat{\tau}_a + \hat{\mathbf{C}}_{21} \hat{\mathbf{C}}_{11}^{-1} [\log(\mathbf{Y}_j) - \mathbf{X}_j \hat{\tau}_a]\}, \qquad \hat{\lambda}_a = 0 \end{aligned} \tag{B1}$$

where $\hat{\tau}_a$ is given in equation (A1), $\mathbf{1} = (1, \ldots, 1)'$, $q \times 1$. In equation (B1), we use the convention that $\mathbf{b}^d = (b_1^d, \ldots, b_p^d)'$.

In addition to the point prediction, we can obtain interval prediction for $\mathbf{y}_j$ given $\mathbf{Y}$. For $q = 1$, we have the approximate predictive interval

$$\hat{y}_j \pm z_{a/2} \hat{\sigma}_y \tag{B2}$$

where $z_{a/2}$ is the $100\alpha/2$ per cent point of the standard normal distribution

$$\sigma_y^2 = [h'(\mu_{2.1})]^2\sigma^2(\mathbf{C}_{22.1} + \mathbf{A}\mathbf{W}^{-1}\mathbf{A}' - 2\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{X}_j\mathbf{W}^{-1}\mathbf{A}')$$

$$\mathbf{A} = \mathbf{x} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{X}_j, \qquad \mathbf{W} = \sum_{j=N_a^*-1}^{N_a^*} \mathbf{X}_j'\mathbf{C}_j^{-1}\mathbf{X}_j$$

$$\mu_{2.1} = \mathbf{x}\hat{\tau}_a + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{Y}_j^{(\hat{\lambda}_a)} - \mathbf{X}_j\hat{\tau}_a), \qquad h'(X) = (1 + \lambda_a X)^{(1-\lambda_a)/\lambda_a}$$

It is noted that $\sigma_{\mathbf{y}}^2$ is the variance of the forecast error for $\mathbf{y}_j$ when the parameters ρ, λ and σ^2 are assumed known, and $\hat{\sigma}_{\mathbf{y}}^2$ is its estimate obtained by substituting the ML estimates for the unknown parameters.

Appendix C: Prediction of future penetration with the ML method

From equations (39) and (B1), we have our predictor for $F_j(t)$ the future penetrations, given by

$$\hat{F}_j(t) = \hat{y}_{jt}/(1 + \hat{y}_{jt}) \tag{C1}$$

For $q = 1$, we have the approximate predictive interval for $F_j(t)$

$$\hat{F}_j(t) \pm z_{a/2}\hat{\sigma}_f \tag{C2}$$

where $z_{a/2}$ is the $100\alpha/2$ per cent point of the standard normal distribution

$$\sigma_f^2 = [g'(\mu_{2.1}^{(\lambda_a)})]^2\sigma^2(\mathbf{C}_{22.1} + \mathbf{A}\mathbf{W}^{-1}\mathbf{A}' - 2\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{X}_j\mathbf{W}^{-1}\mathbf{A}')$$

$$\mathbf{A} = \mathbf{x} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{X}_j, \qquad \mathbf{W} = \sum_{j=N_a^*-1}^{N_a^*} \mathbf{X}_j'\mathbf{C}_j^{-1}\mathbf{X}_j$$

$$\mu_{2.1}^{(\lambda_a)} = \mathbf{x}\hat{\tau}_a + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{Y}_j^{(\hat{\lambda}_a)} - \mathbf{X}_j\hat{\tau}_a)$$

$$g'(X) = (1 + \lambda_a X)^{-(1/\lambda_a + 1)}[1 + (1 + \lambda_a X)^{1/\lambda_a}]^{-2}$$

It is noted that σ_f^2 is obtained in a manner similar to $\sigma_{\mathbf{y}}^2$ discussed in Appendix B.