*Journal of Forecasting, Vol. 16, 65–82 (1997)*

# Bayes Inference for Technological Substitution Data with Data-based Transformation

LYNN KUO
*University of Connecticut, Storrs, USA*

JACK LEE
*National Chiao Tung University, Hsinchu, Taiwan*

PETER CHENG
*Ming Chuan University, Taipei, Taiwan*

JEFFREY PAI
*University of Manitoba, Winnipeg, Canada*

ABSTRACT

Bayesian inference via Gibbs sampling is studied for forecasting technological substitutions. The Box–Cox transformation is applied to the time series AR(1) data to enhance the linear model fit. We compute Bayes point and interval estimates for each of the parameters from the Gibbs sampler. The unknown parameters are the regression coefficients, the power in the Box–Cox transformation, the serial correlation coefficient, and the variance of the disturbance terms. In addition, we forecast the future technological substitution rate and its interval. Model validation and model choice issues are also addressed. Two numerical examples with real data sets are given.

KEY WORDS AR(1); Box–Cox transformation; Metropolis-within-Gibbs sampling; model choice; prediction

## INTRODUCTION

As technology advances, new products replace the old ones. For example, colour televisions replace black-and-white televisions, digital telephone switching systems replace analog switching systems, etc. The rate of change, called *technology* penetration, is defined as the ratio of the number of new products to the combined total of new and old products. A typical growth curve such as Gompertz, logistic, normal, or Weibull is not quite adequate in fitting and forecasting the penetration data. For example, let $F_t$ be the technology penetration at time $t$. Then under a logistic growth curve assumption, we expect to see a linear growth function for $\log F_t/(1 - F_t)$. However, when dealing with real data, Lee and Lu (1987, 1989) and Keramidas and Lee (1990)

CCC 0277–6693/97/020065–18

*Received August 1995*
*Revised September 1996*

found that a more flexible transformation of $y_t = F_t/(1 - F_t)$ gave much superior predictive accuracies. This is primarily due to the violations of the linearity assumption for the growth function and of the independence assumption for the disturbance terms. Lee and Lu found that the residuals of a Box–Cox (1964) transformation of $y_t$ fitted with a linear growth function exhibit strong positive first order autoregressive dependence (AR(1) dependence). They incorporate the Box–Cox transformation in their frequentist's analysis to enhance the linearity of the growth function and to model the proper dependence structure among the observations. In this paper, we pursue a Bayesian analysis along the same line. For a penetration data set with colour television, we show our forecast improves upon the predictor given in Lee and Lu.

A family of data-based transformed models for forecasting technological substitutions has been empirically shown in Lee and Lu (1987, 1989) to be quite useful for short-term forecasts. These models are more general than the four well-known S-shaped growth curve models: logistic, normal, Weibull, and Gompertz. Basically, we first select a transformation from the following choices:

$$\begin{aligned} &(M1) \quad y_t = \mathrm{F}_t/(1 - F_t) \\ &(M2) \quad y_t = \exp\{\Phi^{-1}(F_t)\}, \\ &(M3) \quad y_t = -\log(1 - F_t), \text{ and} \\ &(M4) \quad y_t = -1/\log F_t \end{aligned} \tag{1}$$

Then, we apply the widely used Box–Cox transformation to the AR(1)-dependent data. More specifically, let $y_1, \ldots, y_n$ be a set of first-stage transformed $n$ observations. The second-stage transformation is defined by

$$y_t^{(\lambda)} = \begin{cases} [(y_t + v)^{\lambda} - 1]/\lambda & \text{when } \lambda \neq 0 \\ \log(y_t + v) & \text{when } \lambda = 0 \end{cases} \tag{2}$$

where $v$ is a known constant such that $y_t + v > 0$ for all $t$. In practice $v$ is set to 0 if all $y_t$'s are positive. Furthermore, the $y_t^{(\lambda)}$, $t = 1, \ldots, n$, are assumed to be normally distributed with the mean linear in the regression parameters and with the covariance matrix $\sigma^2\Sigma$, where $\Sigma = (1 - \rho^2)^{-1}V$, $V = (c_{ab})$, $c_{ab} = \rho^{|a-b|}$, for $a, b = 1, \ldots, n$, and $-1 < \rho < 1$. In other words,

$$\mathbf{y}^{(\lambda)} = (y_1^{(\lambda)}, \ldots, y_n^{(\lambda)})' = \mathbf{x}\boldsymbol{\beta} + \mathbf{a} \tag{3}$$

where $\mathbf{x}\boldsymbol{\beta}$ is the growth function, $\boldsymbol{\beta} = (\alpha, \beta)'$, and $\mathbf{x}$ is the design matrix with 1's in the first column and $t_1$ to $t_n$ or $\log(t_1)$ to $\log(t_n)$ in the second column, depending on the growth curve model being considered (see Lee and Lu, 1987). The vector of disturbance terms $\mathbf{a}$ is assumed to be distributed as a multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2\Sigma$. In case a higher degree polynomial is assumed for the growth function, the dimensions of $\mathbf{x}$ and $\boldsymbol{\beta}$ can be adjusted easily. For example, if a second-degree polynomial in time is assumed for the growth function, then the third column of $\mathbf{x}$ is $t_1^2$ to $t_n^2$ or $[\log(t_1)]^2$ to $[\log(t_n)]^2$.

The above families of transformations include the four commonly used link functions: the logit, the probit, the complementary log–log and log–log functions as defined in McCullagh and Nelder (1989, p. 108). These four link functions are obtained by applying the transformations in equations (1) and (2) with $\lambda = 0$.

Lee and Lu (1987, 1989) derive the maximum likelihood estimates of the model in equation (3). This paper studies the problem from a Bayesian point of view. In addition to Bayesian inference

on the parameters, we address prediction for future technology penetrations and predictive intervals. Instead of the plug-in method given in Lee and Lu, we obtain a Bayesian predictive distribution. A Gibbs sampler is used to approximate this predictive distribution and its functionals. This is a Markov chain Monte Carlo algorithm that simulates variates idealistically from the posterior distribution of $\boldsymbol{\beta}$, $\rho$, $\lambda$, and $\sigma$ given the data. The transitional matrix of the Markov chain is written as a product of full conditional densities. The stationary distribution of this Markov chain is the desired posterior density. In addition to Bayesian inference and prediction, we address the issue of model adequacy and model selection using prequential predictive densities. Model selection addresses the question of which of the four first stage transformations together with the Box–Cox transformation is the best. Model adequacy checks whether any of the models are appropriate in forecasting.

We re-analyse two real data sets. One is the technology penetration data for colour televisions provided by Nielson (1985). The other is the penetration data for electronic telephone switching systems of a telephone company. The first data set has been analysed by Lee and Lu (1987, 1989). We show our Bayes solutions to all four models in equation (1) together with the Box–Cox transformation are adequate in fitting the data. Moreover, we show our forecast intervals improve upon the frequentist's ones given in Lee and Lu with better accuracy and shorter lengths.

We seek non-informative priors, so the analysis is more focused on the likelihood. The parameters $\boldsymbol{\beta}$, $\rho$, $\lambda$, and $\sigma$ are assumed to have independent prior distributions. The prior on $\boldsymbol{\beta}$ is assumed to be a bivariate normal with mean $(\mu_\alpha, \mu_\beta)'$ and covariance matrix

$$\Sigma_\pi = \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{pmatrix} \tag{4}$$

The prior on $\lambda$ and $\rho$ can be quite arbitrary, because the Metropolis (1953) algorithm is used to generate these variates. In fact, we have chosen the prior on $\rho$ uniform on $(-1, 1)$, and the prior on $\lambda$ to be flat (improper) on $(-\infty, \infty)$. We have also explored the proper prior on $\lambda$ that is uniform over $(-4, 4)$, and found no appreciable difference. The prior density on $\sigma$ is chosen from a conjugate family

$$\pi(\sigma) \propto \frac{1}{\sigma^{\gamma+1}} e^{-\eta/2\sigma^2} \qquad \gamma > 0, \eta > 0$$

We will denote this modified inverse gamma density by $IG'(\gamma, \eta)$. Note this prior is equivalent to choosing $\sigma^2$ with the inverse gamma density $IG(\gamma/2, 2/\eta)$ as defined by Berger (1985, p. 561). If the non-informative prior $\pi(\sigma) = 1/\sigma$ is desired, we can let $\gamma \to 0$ and $\eta \to 0$. When $\sigma_\alpha \to \infty$ and $\sigma_\beta \to \infty$ as being done in our numerical examples, then our prior on $\boldsymbol{\beta}$, $\sigma$, and $\rho$ is equivalent to the one given by Zellner and Tiao (1964), where Bayesian analysis of the model without the Box–Cox transformation was studied.

When $n$ is small, the parameters are hard to estimate. However, with several independent concurrent short series sharing a common AR(1) covariance structure, we will then have a general growth curve model with AR(1) covariance structure. Keramidas and Lee (1990) apply the maximum likelihood method. Lee and Liu (1996) extend the algorithm here to Bayesian predictive inference for the general growth curve model with AR(1) dependence.

Gibbs sampling has been applied extensively to various problems. We mention only the related work in time series: Albert and Chib (1993); Carlin, Polson, and Stoffer (1992); Chib (1993); Chib and Greenberg (1994); Marriott *et al.* (1995); and McCulloch and Tsay (1994).

The next section develops the conditional densities used in the Gibbs sampler. A detailed discussion of the Gibbs sampler and the Metropolis algorithm is given in Appendix 2. Then we develop the predictive distribution and the one-step forecast. We discuss model adequacy and model selection. Finally, two numerical examples based on real penetration data are given.

## MODEL AND ALGORITHM

In this section, we describe the general model and specify the conditional densities used in the Gibbs sampler. Let us first transform $F_t$ to $y_t$ via equation (1). Then we transform $y_t$ to $y_t^{(\lambda)}$ by applying the Box–Cox transformation (2). Let us note that because of the special structure for $\Sigma$, its inverse is

$$\Sigma^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}$$

Given the model $y_t^{(\lambda)} = \alpha + \beta t + a_t$, where $a_t$ follows the AR(1) model described in the Introduction, we can write our likelihood (cf. Ljung and Box, 1980; Lee and Lu, 1987) as

$$L(\boldsymbol{\beta}, \sigma, \rho, \lambda; \mathbf{y}) \propto \sigma^{-n}(1-\rho^2)^{1/2} \exp\left\{-\frac{1}{2\sigma^2} S(\mathbf{y}, \boldsymbol{\beta}, \rho, \lambda)\right\} \prod_{i=1}^{n} y_i^{\lambda-1} \tag{5}$$

where

$$S(\mathbf{y}, \boldsymbol{\beta}, \rho, \lambda) = (\mathbf{y}^{(\lambda)} - \mathbf{x}\boldsymbol{\beta})' \Sigma^{-1} (\mathbf{y}^{(\lambda)} - \mathbf{x}\boldsymbol{\beta})$$

with

$$\mathbf{x} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 3 & \cdots & n \end{pmatrix}' = (\mathbf{1}\ \mathbf{t}).$$

It is worth noting that the above models with $\lambda = 0$ in the Box–Cox transformation reduce to the four existing growth curve models used in technology substitution data. They are: (1) Logistic (Fisher and Pry, 1971): $\log[F_t/(1-F_t)] = \alpha + \beta t + a_t$; (2) Normal (Stapleton, 1976): $\Phi^{-1}(F_t) = \alpha + \beta t + a_t$; (3) Weibull (Sharif and Islam, 1980): $\log[-\log(1-F_t)] = \alpha + \beta \log t + a_t$; and (4) Gompertz (1825): $-\log(-\log F_t) = \alpha + \beta t + a_t$. For the Weibull model, $\log t$ is used for $t$ in defining $\mathbf{x}$. We can also write the likelihood in equation (5) equivalently in terms of $\mathbf{F} = (F_1, \ldots, F_n)'$ by a change of variable technique. Then we obtain four expressions corresponding to the four transformations in equation (1) respectively. We can pursue inference and prediction using either the likelihood in $\mathbf{y}$ or the likelihood in $\mathbf{F}$. Most of the subsequent developments are derived from equation (5) mostly because it explains the AR(1) process more directly.

Combining the likelihood function (5) with the prior $\pi(\boldsymbol{\beta}, \sigma, \rho, \lambda)$ discussed in the Introduction, we obtain the posterior density

$$\pi(\boldsymbol{\beta}, \sigma, \rho, \lambda \mid \mathbf{y}) \propto L(\boldsymbol{\beta}, \sigma, \rho, \lambda; \mathbf{y}) \pi(\boldsymbol{\beta}, \sigma, \rho, \lambda).$$

Then the Gibbs algorithm proceeds as follows:

(1) Generate $\alpha$ given $\beta, \rho, \lambda, \sigma, \mathbf{y}$ from the normal distribution

$$N\left(\frac{\mu_\alpha\sigma^2 + b\sigma_\alpha^2}{\sigma^2 + a\sigma_\alpha^2}, \frac{\sigma^2\sigma_\alpha^2}{\sigma^2 + a\sigma_\alpha^2}\right),$$

where

$$a = \mathbf{1}'\Sigma^{-1}\mathbf{1} \quad \text{and} \quad b = \mathbf{1}'\Sigma^{-1}(\mathbf{y}^{(\lambda)} - \beta\mathbf{t}).$$

Consequently, if $\sigma_\alpha \to \infty$ in the prior, we generate $\alpha$ from the conditional distribution $N(b/a, \sigma^2/a)$.

(2) Generate $\beta$ given $\alpha, \rho, \lambda, \sigma, \mathbf{y}$ from the normal distribution:

$$N\left(\frac{\mu_\beta\sigma^2 + d\sigma_\beta^2}{\sigma^2 + c\sigma_\beta^2}, \frac{\sigma^2\sigma_\beta^2}{\sigma^2 + c\sigma_\beta^2}\right),$$

where

$$c = \mathbf{t}'\Sigma^{-1}\mathbf{t} \quad \text{and} \quad d = \mathbf{t}'\Sigma^{-1}(\mathbf{y}^{(\lambda)} - \alpha\mathbf{1}).$$

Consequently, if $\sigma_\beta \to \infty$ in the prior, we generate $\beta$ from the conditional distribution $N(d/c, \sigma^2/c)$.

(3) Generate $\rho$ given $\alpha, \beta, \lambda, \sigma, \mathbf{y}$ using the Metropolis (1953) method, where

$$f(\rho) \propto (1 - \rho^2)^{1/2}\ \mathrm{e}^{-(1/2\sigma^2)S(\mathbf{y},\boldsymbol{\beta},\rho,\lambda)}\pi(\rho).$$

(4) Generate $\lambda$ given $\alpha, \beta, \rho, \sigma, \mathbf{y}$ using the Metropolis method, where

$$f(\lambda) \propto \left\{\mathrm{e}^{-(1/2\sigma^2)S(\mathbf{y},\boldsymbol{\beta},\rho,\lambda)}\prod_{i=1}^{n} y_i^{\lambda-1}\right\}\pi(\lambda).$$

(5) Generate $\sigma$ given $\alpha, \beta, \rho, \lambda, \mathbf{y}$ from the modified inverse gamma distribution

$$IG'(\gamma + n, \eta + S(\mathbf{y}, \boldsymbol{\beta}, \rho, \lambda)).$$

The Metropolis algorithm used in steps (3) and (4) is described in Appendix 2. An alternative blocking algorithm that combines steps (1) and (2) into one step should be more efficient especially when $\alpha$ and $\beta$ are highly correlated. Let $\Sigma^* = \mathbf{x}'\Sigma_{-1}\mathbf{x}/\sigma^2 + \Sigma_\pi^{-1}$. That is, we generate the vector $\boldsymbol{\beta}$ given $\sigma, \rho, \lambda$ from the following bivariate normal density:

$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix} \sim N\left(\Sigma^{*-1}\left(\frac{\mathbf{x}'\Sigma^{-1}\mathbf{y}^{(\lambda)}}{\sigma^2} + \Sigma_\pi^{-1}\begin{pmatrix}\mu_\alpha\\ \mu_\beta\end{pmatrix}\right), \Sigma^{*-1}\right)$$

instead of generating $\alpha$ conditioning on $\beta$ and the rest of the parameters and similarly for $\beta$ in two steps. We implemented the latter without blocking because the correlation coefficient between $\alpha$ and $\beta$ is not high in our numerical examples.

## FORECAST

Having obtained the posterior distribution of the parameters, we can use it to predict the future values of $y$. We will illustrate this by the one-step forecast. Generalization to forecast the $l$th future period is straightforward and discussed at the end of this section. Suppose we are at the $n$th period. Let $D_n$ denote the data set $\{y_1, \ldots, y_n\}$. Let $\mathbf{x_i} = (1, i)$. We suppress $\mathbf{x_1}, \ldots, \mathbf{x_n}, \mathbf{x_{n+1}}$ in the notation for convenience, because they are the known covariates. Let $\boldsymbol{\theta}$ denote the parameters $(\boldsymbol{\beta}, \rho, \lambda, \sigma)$. Let $\boldsymbol{\theta}^{(k,s)}$ denote the variate of $\boldsymbol{\theta}$ drawn in the $k$th iteration and $s$th replication of the Gibbs sampler given the data set $D_n$. The same superscript notation will be applied to subsets of $\boldsymbol{\theta}$ and their functionals. We usually take $k$ sufficiently large. Prediction for the $(n+1)$th period follows from the predictive density

$$f(Y_{n+1} \mid D_n) = \int f(Y_{n+1} \mid D_n, \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid D_n)\, \mathrm{d}\boldsymbol{\theta}$$

where $Y_{n+1}$ denotes the random future observation at period $n+1$. This density can be approximated by Monte Carlo integration from the Gibbs sample:

$$\hat{f}(Y_{n+1} \mid D_n) = \frac{1}{r} \sum_{s=1}^{r} f(Y_{n+1} \mid D_n, \boldsymbol{\theta}^{(k,s)}) \tag{6}$$

This predictive density can be obtained by simply drawing $y_{n+1}^{(s)}$ for $s = 1, \ldots, r$ from $f(Y_{n+1} \mid D_n, \boldsymbol{\theta}^{(k,s)})$ for each $\boldsymbol{\theta}^{(k,s)}$ in the replications. The mean of this predictive distribution is computed from

$$E(Y_{n+1} \mid D_n) = E(E(Y_{n+1} \mid D_n, \boldsymbol{\theta}) \mid D_n) \tag{7}$$

To evaluate the inner expectation, let us consider two cases. We assume $v = 0$ in equation (2) for simplicity.

Case (1): $\lambda \neq 0$.
It follows from equation (2) that

$$y_{n+1} = (y_{n+1}^{(\lambda)} \lambda + 1)^{1/\lambda} \tag{8}$$

where

$$y_{n+1}^{(\lambda)} = \mathbf{x_{n+1}} \boldsymbol{\beta} + a_{n+1}$$

and

$$a_{n+1} = \rho a_n + \varepsilon_{n+1}.$$

Note

$$a_n = y_n^{(\lambda)} - \mathbf{x_n} \boldsymbol{\beta} = \frac{y_n^{\lambda} - 1}{\lambda} - \mathbf{x_n} \boldsymbol{\beta}.$$

Therefore,

$$y_{n+1}^{(\lambda)} = \mathbf{x_{n+1}} \boldsymbol{\beta} + \rho \left( \frac{y_n^{\lambda} - 1}{\lambda} - \mathbf{x_n} \boldsymbol{\beta} \right) + \varepsilon_{n+1}.$$

Therefore, from equation (8), we have

$$y_{n+1} = \left\{ \left[ \mathbf{x_{n+1}}\boldsymbol{\beta} + \rho\left( \frac{y_n^{\lambda} - 1}{\lambda} - \mathbf{x_n}\boldsymbol{\beta} \right) + \varepsilon_{n+1} \right] \lambda + 1 \right\}^{1/\lambda}. \tag{9}$$

Case (2): $\lambda = 0$.
We can derive the following equation by a similar method:

$$y_{n+1} = \exp[\mathbf{x_{n+1}}\boldsymbol{\beta} + \rho(\log\, y_n - \mathbf{x_n}\boldsymbol{\beta}) + \varepsilon_{n+1}]. \tag{10}$$

Let $y_{n+1}^{(k,s)}$ denote the functional (9) evaluated at the $k$th iteration and $s$th replication of the sampler, i.e.

$$y_{n+1}^{(k,s)} = \left\{ \left[ \mathbf{x_{n+1}}\boldsymbol{\beta}^{(k,s)} + \rho^{(k,s)}\left( \frac{y_n^{\lambda(k,s)} - 1}{\lambda^{(k,s)}} - \mathbf{x_n}\boldsymbol{\beta}^{(k,s)} \right) + \varepsilon_{n+1}^{(k,s)} \right] \lambda^{(k,s)} + 1 \right\}^{1/\lambda^{(k,s)}} \tag{11}$$

Note that $\varepsilon_{n+1}^{(k,s)}$ is generated from the $N(0, (\sigma^{(k,s)})^2)$ distribution in the $k$th iteration and $s$th replication of the Gibbs sampler.

Therefore, combining equations (7), (9), and (6), when $\lambda \neq 0$, we predict $Y_{n+1}$ by

$$\hat{y}_{n+1} = \frac{1}{r} \sum_{s=1}^{r} y_{n+1}^{(k,s)}. \tag{12}$$

Alternatively, we can predict $Y_{n+1}$ using the median of the Gibbs sample,

$$\tilde{y}_{n+1} = \text{median}\left( \left\{ y_{n+1}^{(k,s)} \right\}_{s=1}^{r} \right). \tag{13}$$

Prediction intervals and quantiles of the functional $Y_{n+1}$ can be computed similarly from the sample $y_{n+1}^{(k,s)}$, $s = 1, \ldots, r$.

Similarly, when $\lambda = 0$, we can use the following variates to predict $Y_{n+1}$:

$$y_{n+1}^{(k,s)} = \exp[\mathbf{x_{n+1}}\boldsymbol{\beta}^{(k,s)} + \rho^{(k,s)}(\log\, y_n - \mathbf{x_n}\boldsymbol{\beta}^{(k,s)}) + \varepsilon_{n+1}^{(k,s)}]. \tag{14}$$

Then we use the mean or the median of the sample to predict $y_{n+1}$ and construct predictive intervals from the Gibbs sample (14).

Theoretically, the posterior probability of the case $\lambda = 0$ is 0 when the prior on $\lambda$ is a continuous measure. However, we had to program the two cases ($\lambda = 0$ and $\lambda \neq 0$) of the forecast rule separately to avoid an overflow problem.

Having obtained $\hat{y}_{n+1}$, we apply the inverse transformations of equation (1) to predict $F_{n+1}$, i.e. (1) $\hat{F}_{n+1} = \hat{y}_{n+1}/(\hat{y}_{n+1} + 1)$; (2) $\hat{F}_{n+1} = \Phi(\log\, \hat{y}_{n+1})$; (3) $\hat{F}_{n+1} = 1 - \exp(-\hat{y}_{n+1})$; and (4) $\hat{F}_{n+1} = \exp\{-1/\hat{y}_{n+1}\}$. The Bayesian predictive intervals for $F_{n+1}$ are calculated from the Gibbs sample of $y_{n+1}$ in equations (11), (14), and the above transformations.

In addition to the naive inverse transformation, we can add a bias-corrected term which involves only the second derivative of the inverse transformation in the Taylor series expansion. However, as noted in Keramidas and Lee (1990, p. 628), the bias-correction does not guarantee to produce better forecasts as the first derivative vanishes. Granger and Newbold (1976) studied

other forecast rules for the transformed variables using Hermite polynomials. We have chosen only the naive inverse transformation in this paper for his simplicity. Moreover, the naive ones have demonstrated desirable forecast performance as in the section of numerical examples.

Now we briefly discuss how we generalize the above one-step prediction rule to multi-step prediction. Let $\mathbf{Y_f} = (Y_{n+1}, \ldots, Y_{n+l})$ denote the vector of future observations. Then the joint density of $\mathbf{Y_f}$ given $D_n$ can be written as

$$f(\mathbf{Y_f} \mid D_n) = f(Y_{n+1} \mid D_n, \boldsymbol{\theta}) f(Y_{n+2} \mid Y_{n+1}, D_n, \boldsymbol{\theta}) \cdots f(Y_{n+l} \mid Y_{n+1}, \ldots, Y_{n+l-1}, D_n, \boldsymbol{\theta})$$

Therefore, the samples from the joint future density can be drawn sequentially with each step similar to the single step in equation (6); a generated new future observation in each step is incorporated as data into the next step of the sequential procedure.

## MODEL VALIDATION AND MODEL CHOICE

Both model adequacy and model selection issues are discussed in this section. Model adequacy is checked by comparing the observed $y_{n+1}$ to its 95% predictive interval developed in the previous section for each $n$ in a series. A model is judged to be adequate if about 95% of the intervals contain the observed future values.

For model selection, we consider three criteria: mean squared error (MSE), mean absolute relative deviation (MARD), and the prequential pseudo-Bayes factor (PPBF).

To define the PPBF, let us define the prequential conditional predictive ordinate (PCPO), which is the predictive density of $Y_{n+1}$ evaluated at the future observed value $y_{n+1}$. Therefore, we have from equation (6),

$$\begin{aligned} d_{n+1} &= \hat{f}(y_{n+1} \mid D_n) \\ &= \frac{1}{r} \sum_{s=1}^{r} \frac{1}{\sqrt{2\pi}\sigma^{(k,s)}} \exp\left\{ - \frac{\left( \frac{y_{n+1}^{\lambda^{(k,s)}} - 1}{\lambda^{(k,s)}} - \mathbf{x}_{n+1}\boldsymbol{\beta}^{(k,s)} - \rho^{(k,s)} a_n^{(k,s)} \right)^2}{2(\sigma^{(k,s)})^2} \right\} y_{n+1}^{\lambda^{(k,s)}-1} \end{aligned} \tag{15}$$

where

$$a_n^{(k,s)} = y_n^{(\lambda^{(k,s)})} - \mathbf{x}\boldsymbol{\beta}^{(k,s)}.$$

Consequently, the PCPO for $F_{n+1}$ is computed by

$$c_{n+1} = \hat{f}(F_{n+1} \mid D_n) = \hat{f}(y_{n+1} \mid D_n) \left| \frac{\mathrm{d}y}{\mathrm{d}F} \right|_{y = y_{n+1}} \tag{16}$$

where $\mathrm{d}y/\mathrm{d}F$ is computed from equation (1).

Let $N$ denote the total number of periods in the data set. We would like to monitor data for a reasonable time period, say for $t = 1, \ldots, I$, before considering model selection. Then we evaluate the following three prediction measurements for each of the models for the periods $I + 1$ to $N$. All the prediction rules are computed prequentially as described in the previous section, i.e. $\hat{F}_t = E(F_t \mid D_{t-1})$. Instead of the mean, we can also consider the median. For each of the models

in equation (1) together with the AR(1) model for the Box–Cox transformation, we can evaluate the following three measurements:

$$\text{MSE} = \frac{1}{N-I}\sum_{t=I+1}^{N}(\hat{F}_t - F_t)^2 \tag{17}$$

$$\text{MARD} = \frac{1}{N-I}\sum_{t=I+1}^{N}\frac{|\hat{F}_t - F_t|}{F_t} \tag{18}$$

$$\text{PPBF} = \prod_{t=I+1}^{N} c_t \tag{19}$$

The best model is the one with smallest MSE, smallest MARD, or the largest PPBF.

Note the PPBF is different from the pseudo Bayes factor in Geisser and Eddy (1979) where the cross-validation idea is used. In fact, the PPBF evaluates the conditional joint predictive density of the data $y_{I+1}, \ldots, y_N$ given $y_1, \ldots, y_I$.

To compute the PPBF in equation (19), we need to run each of the $N - I$ Gibbs samplers from scratch for each prequential conditional predictive ordinate. The sequential imputation algorithm developed by Kong, Liu, and Wong (1994) and Liu and Chen (1995) can be modified to suit for this situation. The sequential imputation algorithm provides a desirable alternative approach to this problem. With the new data $y_{I+1}, \ldots, y_N$ coming in sequentially, the algorithm sequentially impute the unknown parameters using the importance sampling weights based on the same old data batch $y_1, \ldots, y_I$. As long as the new data are not wildly different from the old batch, the importance sampling weights should be well behaved.

## NUMERICAL EXAMPLES

Analysis based on two real-life data sets is summarized here. The first data set, listed in Table I, initially provided by Nielson Inc., is the penetration data for colour television for the period 1956–85. The second data set, listed in Table IV, is the penetrations of electronic telephone switching systems for the years from 1967 to 1984 for a telephone company. The numbers $F_t$ listed are the fractions of the total number of new technology users (users of colour television, electronic switching systems, etc.) divided by the total number of new and old technology users (users of colour and black and white televisions, subscribers for electronic (new) and electro-mechanical (old) switching systems).

Table I. Technology penetration data for colour TV

| Year | Penetrations | Year | Penetrations | Year | Penetrations |
|---|---|---|---|---|---|
| 1956 | 0.00052 | 1966 | 0.09694 | 1976 | 0.73606 |
| 1957 | 0.00219 | 1967 | 0.16325 | 1977 | 0.77065 |
| 1958 | 0.00394 | 1968 | 0.24175 | 1978 | 0.77984 |
| 1959 | 0.00569 | 1969 | 0.32034 | 1979 | 0.80926 |
| 1960 | 0.00743 | 1970 | 0.35744 | 1980 | 0.83028 |
| 1961 | 0.00943 | 1971 | 0.41015 | 1981 | 0.82916 |
| 1962 | 0.01208 | 1972 | 0.48631 | 1982 | 0.87595 |
| 1963 | 0.01889 | 1973 | 0.55355 | 1983 | 0.88703 |
| 1964 | 0.03120 | 1974 | 0.62251 | 1984 | 0.90465 |
| 1965 | 0.05332 | 1975 | 0.68394 | 1985 | 0.91519 |

Table II. Gibbs approximation to the estimates for colour TV data

| | Mean | S.D. | 2.5% | 5% | 25% | 50% | 75% | 95% | 97.5% |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\alpha}$ | −2.38 | 0.30 | −2.94 | −2.79 | −2.54 | −2.38 | −2.25 | −1.88 | −1.76 |
| $\hat{\beta}$ | 0.15 | 0.01 | 0.13 | 0.14 | 0.15 | 0.16 | 0.16 | 0.17 | 0.17 |
| $\hat{\rho}$ | 0.90 | 0.07 | 0.75 | 0.77 | 0.87 | 0.92 | 0.96 | 0.99 | 0.99 |
| $\hat{\lambda}$ | −0.08 | 0.06 | −0.20 | −0.18 | −0.12 | −0.08 | −0.04 | 0.02 | 0.04 |
| $\hat{\sigma}$ | 0.08 | 0.01 | 0.06 | 0.06 | 0.07 | 0.08 | 0.09 | 0.10 | 0.10 |

Diffuse priors on the parameters are chosen for both data sets. The parameters in equation (4) are chosen such that $\sigma_\alpha \to \infty$ and $\sigma_\beta \to \infty$. Given this limiting case, our posterior does not depend on $\mu_\alpha$ or $\mu_\beta$. The prior on $\rho$ is uniform $(-1, 1)$ to reflect the equal likelihood of $\rho$. The prior on $\lambda$ is chosen to be the improper flat prior, i.e. $\pi(\lambda) = 1$ for all $\lambda$. The prior on $\sigma$ is also improper, namely, $\pi(\sigma) = 1/\sigma$.

Table II lists the point estimates, estimates of the standard deviations, and the percentiles for each of the parameters for the complete colour television data set (1956 to 1985) and the model with the reciprocal link transformation ($M4$). The Bayes estimates for the parameters are not significantly different among the four link transformations; therefore we omit the results for the other link transformations. Initially, the Bayes estimates are computed from the Gibbs sampler with 53 iterations and 500 replications and with 50 loops in each Metropolis algorithm to generate $\rho$ and $\lambda$ for each iteration and replication of the Gibbs sampler. The starting points for the replications for each parameter are chosen from random perturbations around the maximum likelihood estimates. The convergence of the Gibbs sampler is monitored by examining their empirical quantiles. This leads us to iterate the above Gibbs sampler ten more times to yield the results in Tables II and III. The Bayesian predictive intervals can easily be obtained from Table II. For example, the 95% intervals are read from the 2.5% and 97.5% columns.

For the first data set, we produce the one-step-ahead prediction rule for the year 1966 using both the mean and the median of the Gibbs sample based on the data set for the year 1956 to 1965. Then we predict the subsequent years prequentially, i.e. using the data from 1956 to 1966 to predict the penetration for 1967, etc. In Figure 1, the one-step forecasts for the years 1966 to 1985 using the medians predictors are plotted for each of the four link transformations. They are also compared with the actually observed data. This figure shows that the logistic transformation ($M1$) produces the best forecasts except for the years 1974–1976 and 1982.

Figure 2 provides model adequacy checks in which the 95% Bayesian credible intervals are plotted pointwise for each of the one-step-ahead prediction rules. The figure shows that all four models are adequate because the future observed values are always contained in the intervals. The confidence bands become narrower as time increases as expected because more data are available. It can be observed that the four models differ less on the left tails of the distributions than on the right tails.

The 90% and 95% frequentist's predictive intervals can be computed from a Student-$t$ distribution. Let $\hat{y}^*$ denote the frequentist's forecast of $y_n$ given in the Appendix 1. Then $\hat{y}^{*(\lambda)} \pm t_{\alpha/2(n-2)}\hat{\sigma}$ will yield a $100(1-\alpha)\%$ predictive interval for the time $t = n$, where $\hat{\sigma}$ is computed similarly to equations (A15) and (A17) of Lee and Lu (1987, p. 76). Hence approximately $100(1-\alpha)\%$ predictive intervals for $y_n$ can be obtained from $[\hat{y}^{*\hat{\lambda}} - \hat{\lambda}t_{\alpha/2(n-2)}\hat{\sigma}]^{1/\hat{\lambda}} < y_n < [\hat{y}^{*\hat{\lambda}} + \hat{\lambda}t_{\alpha/2(n-2)}\hat{\sigma}]^{1/\hat{\lambda}}$. The frequentist's predictive rules are transformed back to obtain the predictive rules $\hat{F}$. In Figure 3, these frequentist's 95% intervals are plotted against the Bayesian intervals for the years 1966 to 1985, both with the logistic link. It can be seen that the Bayesian

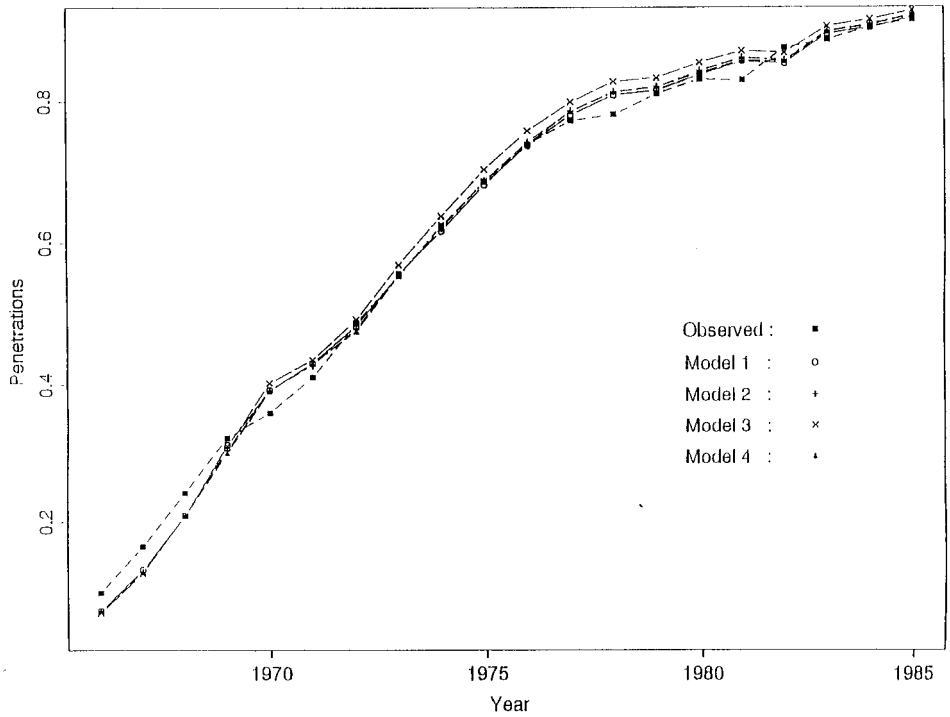


Figure 1. Plot of the Bayesian one-step forecasts for each of the four links versus years

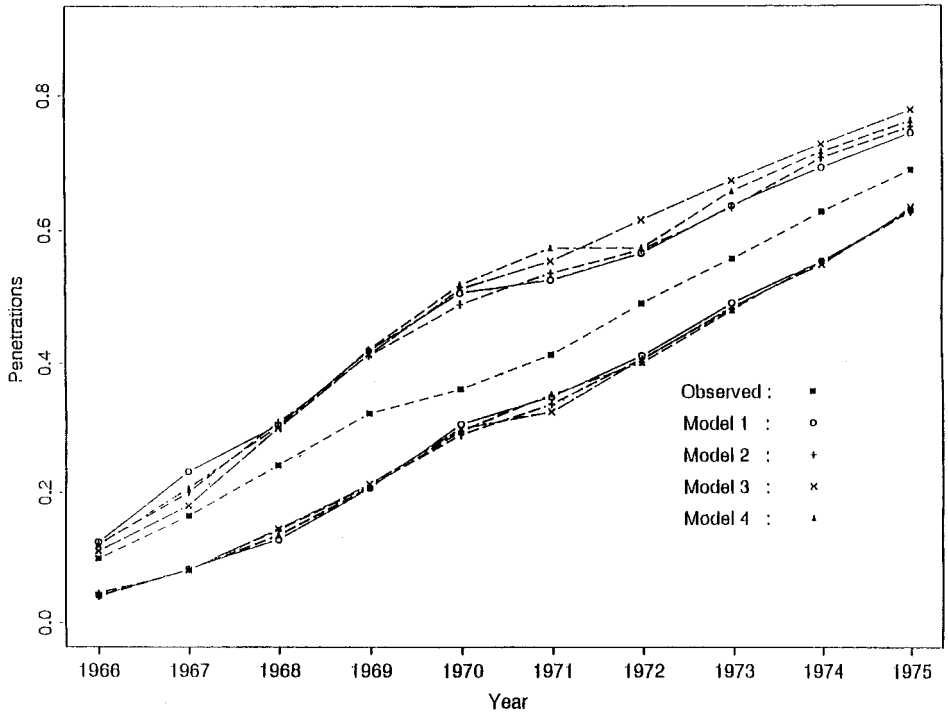


Figure 2. Plot of 95% Bayesian one-step-ahead predictive intervals for each of the four links versus years

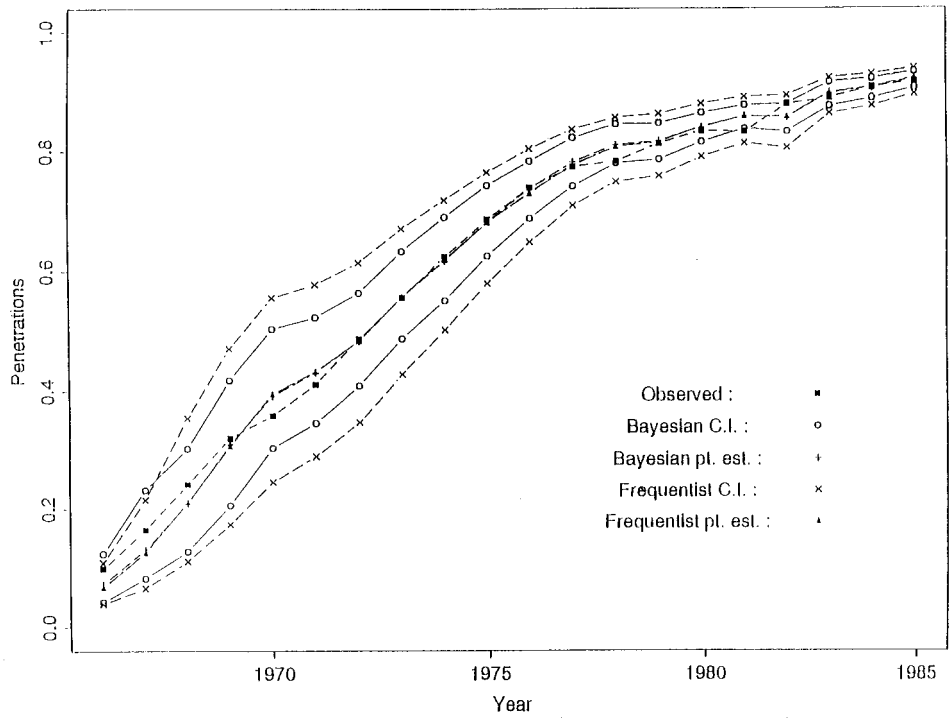


Figure 3. Plot of 95% Bayesian and frequentist's one-step-ahead predictive intervals and for the logistic link versus years

predicting intervals are not only shorter in length than the frequentist's ones but also more accurate in prediction.

Table III addresses the model selection question. The three criteria MSE, MARD, and PPBF are computed for the last 20 periods, i.e. 1966 to 1985. The three measurements are denoted by MSE(B), MARD(B), and PPBF. Furthermore, MSE and MARD are also computed using the frequentist's one-step prediction rule. They are denoted by MSE(F) and MARD(F). The table shows that the Bayesian predictor produces smaller MSE and MARD than the frequentist's predictor. Table III also shows that the logistic transformation ($M1$) yields the smallest MSE and MARD, Bayesian or frequentist's, among the four transformations. This is confirmed by Figure 2. On the other hand, the reciprocal log transformation ($M4$) yields the largest PPBF. This is confirmed by Figure 4, where the prequential CPOs for the years 1966 to 1985 are plotted for each of the link functions. It is not entirely surprising that we select different models due to different criteria. The skewedness of the predictive density revealed in Figure 2 manifests this

Table III. Model selection for the colour TV data

| | MSE(B) | MARD(B) | PPBF(log) | MSE(F) | MARD(F) |
|---|---|---|---|---|---|
| $M1$ | 0.00034 | 0.047 | 37.39 | 0.00038 | 0.052 |
| $M2$ | 0.00040 | 0.051 | 46.74 | 0.00052 | 0.058 |
| $M3$ | 0.00070 | 0.063 | 40.99 | 0.00054 | 0.064 |
| $M4$ | 0.00040 | 0.051 | 46.84 | 0.00045 | 0.059 |

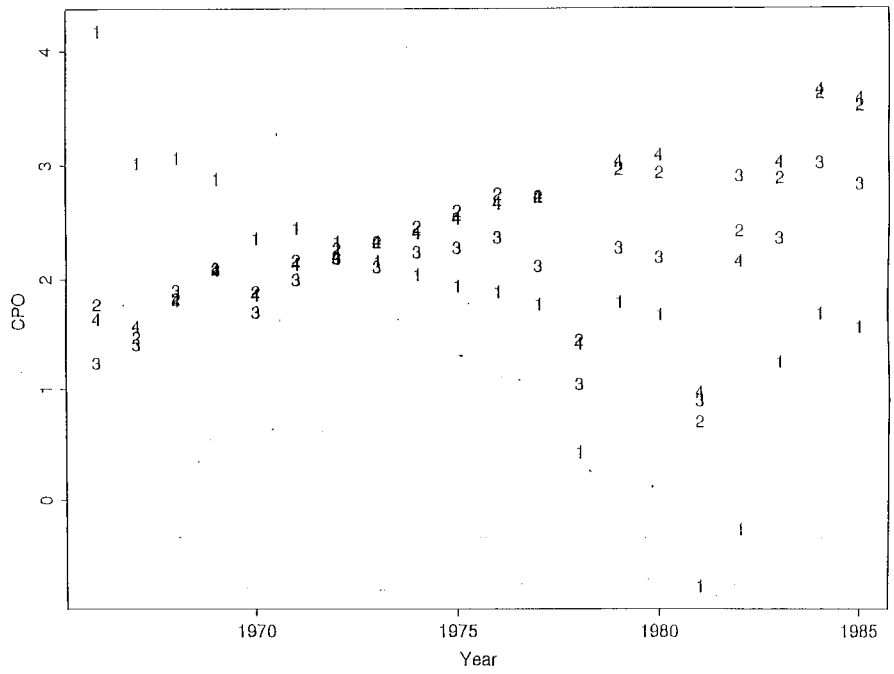


Figure 4. Plot of the prequential conditional predictive ordinates (PCPO) for each of the four links versus years

phenomenon. Figure 4 further shows that models 2 and 4 are quite comparable and the logistic ($M1$) model fits well for the earlier periods up to 1972.

Table V lists the Bayes estimates for the various parameters and its credible intervals based on the telephone switching data set. Note that the estimates for $\beta$ are positive for both data sets. This

Table IV. Penetrations of electronic telephone switching systems

| Year | Penetrations | Year | Penetrations | Year | Penetrations |
|---|---|---|---|---|---|
| 1967 | 0.00071 | 1973 | 0.11259 | 1979 | 0.30486 |
| 1968 | 0.00386 | 1974 | 0.14553 | 1980 | 0.32404 |
| 1969 | 0.00447 | 1975 | 0.16736 | 1981 | 0.35516 |
| 1970 | 0.02003 | 1976 | 0.19083 | 1982 | 0.37708 |
| 1971 | 0.02961 | 1977 | 0.20130 | 1983 | 0.41134 |
| 1972 | 0.05974 | 1978 | 0.26299 | 1984 | 0.43709 |

Table V. Gibbs approximation to the estimates for the telephone data set

| | Mean | S.D. | 2.5% | 5% | 25% | 50% | 75% | 95% | 97.5% |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\alpha}$ | −2.23 | 0.30 | −2.83 | −2.64 | −2.36 | −2.21 | −2.08 | −1.85 | −1.79 |
| $\hat{\beta}$ | 0.11 | 0.01 | 0.08 | 0.09 | 0.10 | 0.11 | 0.12 | 0.14 | 0.14 |
| $\hat{\rho}$ | 0.70 | 0.20 | 0.30 | 0.34 | 0.55 | 0.73 | 0.87 | 0.97 | 0.98 |
| $\hat{\lambda}$ | 0.46 | 0.05 | 0.35 | 0.38 | 0.43 | 0.46 | 0.50 | 0.55 | 0.57 |
| $\hat{\sigma}$ | 0.06 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.09 | 0.10 |

Table VI. Model selection for the telephone data set

| | MSE(B) | MARD(B) | PPBF(log) | MSE(F) | MARD(F) |
|---|---|---|---|---|---|
| $M1$ | 0.00028 | 0.047 | 25.72 | 0.00058 | 0.061 |
| $M2$ | 0.00021 | 0.038 | 20.92 | 0.00141 | 0.106 |
| $M3$ | 0.00049 | 0.059 | 17.88 | 0.00317 | 0.173 |
| $M4$ | 0.00024 | 0.040 | 22.37 | 0.00080 | 0.072 |

is reasonable because we would expect substitutions to increase with time. The estimate for $\lambda$ for the TV data is much smaller than that of the telephone data indicating that the dependence for the latter is much stronger. Table VI lists the model selection measures for the last eight periods, i.e. 1977 to 1984. It reveals that the logistic transformation ($M1$) is preferred by the prequential pseudo-Bayes factor. The probit transformation ($M2$) is preferred by the Bayesian MSE and MARD criteria. The Bayesian prediction rules are better than the frequentist's ones in prediction. Moreover, our figures show that the Bayesian prediction rules have shorter intervals than the frequentist's ones and more accurate; all four link transformations, with the Box–Cox transformation and AR(1) models, are adequate. These figures are omitted.

## APPENDIX 1: FREQUENTIST'S PREDICTION RULE

The frequentist's prediction rule is given here. Let $\mathbf{y}^* = (y_{n+1}, \ldots, y_{n+l})'$ be an $l$-dimensional vector of future observations to be predicted on the basis of the past observations $\mathbf{y} = (y_1, \ldots, y_n)'$. Let $\mathbf{x}^*$ denote the design matrix corresponding to the future observations. Furthermore, let $\mathrm{Cov}(\mathbf{y}^{(\lambda)'}, \mathbf{y}^{*(\lambda)'})' = (\sigma^2/(1-\rho^2))(c_{ab})$, where $c_{ab} = \rho^{|a-b|}, a, b = 1, \ldots, n+l$. When the unknown parameters are replaced by their maximum likelihood estimates from equation (A3), the forecasts of $\mathbf{y}^*$ can be written as

$$\hat{\mathbf{y}}^*_{(l\times 1)} = \begin{cases} \{1 + \hat{\lambda}[\mathbf{x}^*\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\rho}}(y_n^{(\hat{\lambda})} - \mathbf{x}_n\hat{\boldsymbol{\beta}})]\}^{1/\hat{\lambda}} - v & \text{when } \hat{\lambda} \neq 0 \\ \exp[\mathbf{x}^*\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\rho}}(y_n^{(\hat{\lambda})} - \mathbf{x}_n\hat{\boldsymbol{\beta}})] - v & \text{when } \hat{\lambda} = 0 \end{cases} \tag{A1}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{x}'\hat{V}^{-1}\mathbf{x})^{-1}\mathbf{x}'\hat{V}^{-1}\mathbf{y}^{(\hat{\lambda})}$, $\hat{\boldsymbol{\rho}} = (\hat{\rho}, \hat{\rho}^2, \ldots, \hat{\rho}^l)'$, $\mathbf{1} = (1, \ldots, 1)'$, $\hat{V} = (\hat{c}_{ab})$, $\hat{c}_{ab} = \hat{\rho}^{|a-b|}$, $\mathbf{x}_n$ and $y_n^{(\hat{\lambda})}$ are the $n$th row of $\mathbf{x}$ and $\mathbf{y}^{(\hat{\lambda})}$ respectively, and $\hat{\lambda}$ and $\hat{\rho}$ are the MLEs of $\lambda$ and $\rho$, respectively. For the one-step-ahead forecast, we set $l = 1$ in the above discussion and let $\hat{y}^*$ denote the first row of $\hat{\mathbf{y}}^*_{(l\times 1)}$. We assume $v$ is 0 in the two examples.

## APPENDIX 2: GIBBS SAMPLING

We describe briefly the Gibbs sampling procedure. Suppose we desire to estimate $f(U_1, U_2, \ldots, U_p \mid D_n)$, the posterior joint density of $(U_1, U_2, \ldots, U_p)$, given the data $D_n$. The algorithm assumes that we can generate variates from the conditional densities $f(U_i \mid \{U_j\}_{j\neq i}, D_n)$. The algorithm proceeds as follows. Let us start with initial values, $U_1^{(0)}, U_2^{(0)}, \ldots, U_p^{(0)}$. Generate a value $U_1^{(1)}$ from the conditional density $f(U_1 \mid U_2^{(0)}, \ldots, U_p^{(0)}, D_n)$. Similarly, generate a value $U_2^{(1)}$ from the conditional density $f(U_2 \mid U_1^{(1)}, U_3^{(0)}, \ldots, U_p^{(0)}, D_n)$, and continue until the value $U_p^{(1)}$ is generated from the conditional density $f(U_p \mid U_1^{(1)}, \ldots, U_{p-1}^{(1)}, D_n)$. With this new realization of the values $\mathbf{U}^{(1)} = (U_1^{(1)}, \ldots, U_p^{(1)})$ replacing the old values, we can continue to iterate until the $k$th realization. Under very mild regularity conditions, this Markov chain

converges to a stationary distribution for large $k$. The vector $(U_1^{(k)}, \ldots, U_p^{(k)})$ has a distribution that is approximately equal to $f(U, \ldots, U_p \mid D_n)$. By starting independent initial choices, we can also replicate the above iterations $r$ times. Let $\mathbf{U}^{(i,s)} = (U_1^{(i,s)}, \ldots, U_p^{(i,s)})$ denote the realization of $\mathbf{U}$ for the $i$th iteration and $s$th replication. The posterior moments, functionals, and credible sets can be computed from the empirical measure assigning weight $1/r$ to each $(U_1^{(k,s)}, \ldots, U_p^{(k,s)})$, $s = 1, \ldots, r$, to compute posterior features.

For the issue of convergence and the issue of pragmatic choices of $k$ and $r$, consult Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), Gelfand *et al.* (1990), Gelman and Rubin (1992), and Tierney (1994).

When the conditional densities are not easily identified, such as in cases without conjugate priors, the Metropolis (1953) algorithm or importance sampling methods can be employed. We describe the Metropolis algorithm briefly. Suppose we desire to sample a variate from the following generic density:

$$f(U_1 \mid U_2, \ldots, U_p, D_n) = \frac{f(D_n \mid U_1, \ldots, U_p)\pi(U_1, U_2, \ldots, U_p)}{\int_{-\infty}^{\infty} f(D_n \mid U_1, \ldots, U_p)\pi(U_1, U_2, \ldots, U_p)\, \mathrm{d}U_1} \tag{A2}$$

Let $f(U_1)$ denote the conditional density in equation (A2), suppressing the conditioning variables for brevity. Let us define a transition kernel $q(U_1, X)$, which maps $U_1$ to $X$. If $U_1$ is a real variable with range in $(-\infty, \infty)$, we can construct $q$ such that $X = U_1 + \sigma' Z$ with $Z$ being the standard normal random variate and $\sigma'^2$ reflecting the conditional variance of $U_1$ in equation (A2). Then the Metropolis algorithm proceeds as follows:

Step 1: start with any point $U_1^{(0)}$, and stage indicator $j = 0$.
Step 2: generate a point $X$ according to the transition kernel $q(U_1^{(j)}, X)$.
Step 3: update $U_1^{(j)}$ to $U_1^{(j+1)} = X$ with probability $p = \min\{1, f(X)/f(U_1^{(j)})\}$, stay at $U_1^{(j)}$ with probability $1 - p$.
Step 4: repeat steps 2 and 3 by increasing the stage indicator until the process reaches a stationary distribution.

Chapter 9 of Hammersley and Handscomb (1964) provides a discussion of why this algorithm works. Note that this algorithm is defined by using the ratio of two values of equation (A2). Therefore, all we need is to know the functional form of the likelihood and prior. This spares us the task of evaluating the normalizing constant.

If $U_1$ is a bounded variable with range in $(a, b)$, we can use a distribution with support on $(a, b)$ (for example, uniform over $(a, b)$ distribution) to generate $X$. In this case, the transition kernel $q(U_1, X)$ does not depend on $U_1$. Alternatively, we can apply the transformation, such as $U' = \log[(U_1 - a)/(b - U_1)]$, to map $(a, b)$ into $(-\infty, \infty)$, then use the normal transition kernel described earlier to generate $X$ using the density of $U'$. We implemented the latter procedure in our numerical examples because the normal transitional kernel seems to yield a more efficient algorithm than the uniform transitional kernel. On using the normal kernel, we first transform $\rho$ to $\rho' \in (-\infty, \infty)$ by $\rho' = \log[(1 + \rho)/(1 - \rho)]$. Then we apply the Metropolis algorithm to the function

$$f(\rho') \propto \frac{\mathrm{e}^{3\rho'/2}}{(1 + \mathrm{e}^{\rho'})^3} \mathrm{e}^{-(1/2\sigma^2)S'(\mathbf{y}, \boldsymbol{\beta}, \rho', \lambda)}$$

where $S'(\mathbf{y}, \boldsymbol{\beta}, \rho', \lambda)$ is obtained from $S(\mathbf{y}, \boldsymbol{\beta}, \rho, \lambda)$ with $\rho$ replaced by $(\mathrm{e}^{\rho'} - 1)/(\mathrm{e}^{\rho'} + 1)$. We also need to specify $\sigma_{\rho'}$ (denoted by $\sigma'$ in the above Metropolis algorithm). The quantity $\sigma_{\rho'}$ is usually

chosen to reflect the conditional standard deviation of $\rho'$ given $\boldsymbol{\beta}, \lambda, \sigma, \mathbf{y}$. We propose three methods to estimate the variance $\sigma^2_{\rho'}$:

(1) Start with an arbitrary guess of $\sigma_{\rho'}$; use the Gibbs sampler of the $\rho'$s to estimate empirically its variance.
(2) Obtain estimates for the mean and variance of $\rho$, denoted by $\mu_\rho$ and $\sigma^2_\rho$, from the SAS AUTOREG procedure given a fixed $\lambda$. Apply the delta method to obtain an estimate of the variance of $\rho'$, i.e. $\sigma^2_{\rho'} \approx \sigma^2_\rho(2/(1-\mu^2_\rho))^2$.
(3) Derive the profile likelihood of $\rho$ and $\lambda$ as in Lee and Lu (1987),

$$l(\rho, \lambda; \mathbf{y}) = -\frac{n}{2}\log\,\hat{\sigma}^2(\rho, \lambda) + (\lambda - 1)\sum_{t=1}^{n}\log\,y_t + \frac{1}{2}\log(1-\rho^2) \tag{A3}$$

where

$$\hat{\sigma}^2(\rho, \lambda) = \frac{1}{n}(\mathbf{y}^{(\lambda)} - \mathbf{x}\hat{\boldsymbol{\beta}})'\Sigma^{-1}(\mathbf{y}^{(\lambda)} - \mathbf{x}\hat{\boldsymbol{\beta}})$$

with

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}'\Sigma^{-1}\mathbf{x})^{-1}\mathbf{x}'\Sigma^{-1}\mathbf{y}^{(\lambda)}$$

Then, invert the sample information matrix at the maximum likelihood estimates of equation (A3) to obtain the preliminary variance estimates of $\rho$ and $\lambda$ respectively. This method is adopted in our analysis, because it has the advantage of obtaining both variances. Then we apply the delta method to obtain the variance estimates for $\rho$ and $\lambda$. These variance estimates are further refined by empirical evaluations of the variance estimates from the Gibbs sampler because the conditional variances are required in the Metropolis algorithm. Having obtained $\rho'$ from the Metropolis algorithm, we transform the $\rho'$ back to $\rho$ by $\rho = (e^{\rho'} - 1)/(e^{\rho'} + 1)$.

An operation like that applied to $\rho$ can also be applied to the variate $\lambda$, except that the above method (3) of using the SAS AUTOREG is not applicable for $\lambda$.

Conditioning on $\lambda$, our model is a special case of the models that are discussed by Albert and Chib (1993), Chib (1993) and Chib and Greenberg (1994). Therefore, the generation of $\rho$ can be done more efficiently using the idea in Chib and Greenberg. That is the full conditional distribution of $\rho$ for the uniform prior can be written as

$$\rho \,|\, \text{data}, \boldsymbol{\beta}, \sigma^2 \propto \Psi(\rho) \times N(\hat{\rho}, \omega^{-1})I_{[-1,1]}$$

where $\Psi(\rho)$ is the density of the initial observation and

$$\omega = \sum_{i=1}^{n-1}(y_i^{(\lambda)} - \alpha - \beta i)^2$$

and

$$\hat{\rho} = \omega^{-1}\sum_{i=1}^{n-1}(y_i^{(\lambda)} - \alpha - \beta i)(y_{i+1}^{(\lambda)} - \alpha - \beta(i+1)).$$

Then the Metropolis step can be done by drawing a $\rho$ (say, $\rho^*$) from $N(\hat{\rho}, \omega^{-1})$. The chain stays at the current value if $\rho^*$ does not satisfy stationarity; otherwise it moves to $\rho^*$ with probability given by

$$\min\left\{\frac{\Psi(\rho^*)}{\Psi(\rho)}, 1\right\}$$

## ACKNOWLEDGEMENTS

Kuo's research was partially supported by the National Science Foundation Grant DMS-9008021. Lee's research was partially supported by the National Research Council Grants 82-208-M009-054 and 85-2121-M-009-004. Pai's research was partially supported by the Research Foundation, University of Connecticut.

## REFERENCES

Albert, J. and Chib, S., 'Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts', *Journal of Business and Economic Statistics*, **11** (1993), 1–16.

Berger, J. O., *Statistical Decision Theory and Bayesian Analysis*, 2nd edn, New York: Springer-Verlag, 1985.

Box, G. E. P. and Cox, D. R., 'An analysis of transformations', *Journal of the Royal Statistical Society, A*, **26** (1964), 211–252.

Carlin, B., Polson, N. and Stoffer, D., 'A Monte Carlo approach to nonnormal and non-linear state-space modeling', *Journal of the American Statistical Association*, **87** (1992), 493–500.

Chib, S., 'Bayes regression with autoregressive errors, a Gibbs sampling approach', *Journal of Econometrics*, **58** (1993), 275–294.

Chib, S. and Greenberg, E., 'Bayes inference in regression models with ARMA($p$, $q$) errors', *Journal of Econometrics*, **64** (1994), 183–206.

Fisher, J. C. and Pry, R. H., 'A simple substitution model of technological change', *Technological Forecasting and Social Change*, **3** (1971), 75–88.

Geisser, S. and Eddy, W., 'A predictive approach to model selection', *Journal of the American Statistical Association*, **74** (1979), 153–160.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M., 'Illustration of Bayesian inference in normal data models using Gibbs sampling', *Journal of the American Statistical Association*, **85** (1990), 972–985.

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

Gelman, A. E. and Rubin, D., 'Inference from iterative simulation using multiple sequences', *Statistical Science*, **7** (1992), 457–472.

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

Gompertz, B., 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies', *Philosophical Transactions of the Royal Society of London*, **115** (1825), 513–585.

Granger, C. W. J. and Newbold, P., 'Forecasting transformed series', *Journal of the Royal Statistical Society, B*, **38** (1976), 189–203.

Hammersley, J. M. and Handscomb, D. C., *Monte Carlo Methods*, London: Chapman and Hall, 1964.

Keramidas, E. M. and Lee, J. C., 'Forecasting technological substitutions with concurrent short time series', *Journal of the American Statistical Association*, **85** (1990), 625–632.

Kong, A., Liu, J. S. and Wong, W. H., 'Sequential imputations and Bayesian missing data problems', *Journal of the American Statistical Association*, **89** (1994), 278–288.

Lee, J. C. and Lu, K. W., 'On a family of data-based transformed models useful in forecasting technological substitutions', *Technological Forecasting and Social Change*, **31** (1987), 61–78.
Lee, J. C. and Lu, K. W., 'Algorithm and practice of forecasting technological substitutions with data-based transformed models', *Technological Forecasting and Social Change*, **36** (1989), 401–414.
Lee, J. C. and Liu, K. C., 'Bayesian analysis of a general growth curve model with power transformations and AR(1) dependence', paper presented at the 1996 joint statistical meeting in Chicago.
Liu, J. S. and Chen, R., 'Blind deconvolution via sequential imputations', *Journal of the American Statistical Association*, **90** (1995), 567–576.
Ljung, G. M. and Box, G. E. P., 'Analysis of variance with autocorrelated observations', *Scandinavian Journal of Statistics*, **7** (1980), 172–180.
Marriott, J., Ravishanker, N., Gelfand, A. and Pai, J., 'Bayesian analysis for ARMA processes: complete sampling based inference under exact likelihoods', in Berry, D., Chaloner, K. and Geweke, J. (eds), *Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner*, (1994), 241–254.
McCullagh, P. and Nelder, J. A., *Generalized Linear Models*, 2nd edn. London: Chapman and Hall, 1989.
McCulloch, R. and Tsay, R. S., 'Bayesian analysis of autoregressive time series via the Gibbs sampler', *Journal of Time Series Analysis*, **15** (1994), 235–250.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E., 'Equations of state calculations by fast computing machines', *Journal of Chemical Physics*, **21** (1953), 1087–1091.
Nielson, A. C. Inc., private communications (1985).
Sharif, M. N. and Islam, M. N., 'The Weibull distribution as a general model for forecasting technological change', *Technological Forecasting and Social Change*, **18** (1980), 247–256.
Stapleton, E., 'The normal distribution as a model of technological substitution', *Technological Forecasting and Social Change*, **8** (1976), 325–334.
Tanner, M. and Wong, W., 'The calculation of posterior distributions by data augmentation' (with discussion), *Journal of the American Statistical Association*, **81** (1987), 82–86.
Tierney, L., 'Markov chains for exploring posterior distributions' (with discussion), *The Annals of Statistics*, **22** (1994), 1701–1758.
Zellner, A. and Tiao, G. (1964), 'Bayesian analysis of the regression model with autocorrelated errors', *Journal of the American Statistical Association*, **59** (1964), 763–778.

*Author's biographies*:

**Lynn Kuo** is a professor in the Department of Statistics at the University of Connecticut. She received her PhD from the University of California at Los Angeles. Her current research interests include Bayesian computation, software reliability, and event history analysis.

**Jack Lee** is professor and director at the Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan. He received his PhD from the State University of New York at Buffalo. His current research interests include time series, growth curve, and multivariate analysis.

**Peter Cheng** is an associate professor in the Department of Statistics at Ming Chuan University, Taipei, Taiwan. He received his PhD from Columbia University. His research interests include bioassays with historical controls and isotonic regression methods for tumour analysis.

**Jeffrey Pai** is an assistant professor at the Warren Center for Actuarial Studies and Research, Faculty of Management, University of Manitoba, Winnipeg, Canada. He received his PhD from the University of Connecticut. His research interests include time series, volatility and compound loss models.

*Authors' addresses*:

**Lynn Kuo**, Department of Statistics, University of Connecticut, Storrs, CT 06269-3120, USA.
**Jack Lee**, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.
**Peter Cheng**, Department of Statistics, Ming Chuan University, Taipei, Taiwan.
**Jefferey Pai**, Warren Center for Actuarial Studies & Research, Faculty of Management, University of Manitoba, Winnipeg, Manitoba, R3T 5V4 Canada.