On Selecting a Power Transformation in Time-series Analysis

CATHY W. S. CHEN[1]* AND JACK C. LEE[2]
[1] *Feng-Chia University, Taiwan*
[2] *National Chiao Tung University, Taiwan*

ABSTRACT
The primary aim of this paper is to select an appropriate power transformation when we use ARMA models for a given time series. We propose a Bayesian procedure for estimating the power transformation as well as other parameters in time series models. The posterior distributions of interest are obtained utilizing the Gibbs sampler, a Markov Chain Monte Carlo (MCMC) method. The proposed methodology is illustrated with two real data sets. The performance of the proposed procedure is compared with other competing procedures. © 1997 John Wiley & Sons, Ltd.

J. forecast, Vol. **16**, 343–354 (1997)

No. of Figures: 5. No. of Tables: 5. No. of References: 30.

KEY WORDS ARMA models; Forecasting; Gibbs sampler; MCMC method; power transformation

INTRODUCTION

Many biological and economic data sets encountered in practice appear to be non-normal or heteroscedastic in variance of error terms. Using power transformation to achieve normality and stable variance has no doubt occurred to data analysts from time to time. The applicability of statistical models can be enhanced through the use of power transformations, and time-series models are no exception. Box and Jenkins (1976) suggested using a power transformation to obtain an adequate autoregressive moving average (ARMA) model for the time series at hand. In particular, they advocate the use of power transformations introduced by Box and Cox (1964). The power transformations are given by

$$T(y_t) = \begin{cases} \dfrac{(y_t + v)^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y_t + v) & \text{if } \lambda = 0 \end{cases} \tag{1}$$

* Correspondence to: Cathy W. S. Chen, Department of Statistics, Feng-Chia University, Taichung 407, Taiwan

Contract grant sponsor: National Science Council of Taiwan; Contract grant number: NSC84-2121-M-035-007 and NSC86-2115-M-009-027

CCC 0277–6693/97/050343–12$17.50

where $\{y_t\}$ is an observed time series with a known constant v such that $y_t + v > 0$ for all t, λ is the index of the transformation and log stands for natural logarithm. Rather than the Box–Cox transformation, the following family of transformations is employed in practice:

$$y_t^{(\lambda)} = \begin{cases} (y_t + v)^{\lambda} & \text{if } \lambda \neq 0 \\ \log(y_t + v) & \text{if } \lambda = 0 \end{cases} \tag{2}$$

The transformation in equation (2) and Box–Cox transformation differ only in the scale and origin of the transformed data. The general analysis of the data is unaffected (Draper and Smith, 1981). However, it is important to select an adequate transformation as was suggested by several discussants of the paper by Chatfield and Prothero (1973), and Hopwood *et al*. (1984). Nelson and Granger (1979) reported that experience with using the Box–Cox transformation when forecasting economic time series shows that it does not consistently produce superior forecasts. Once a model has been constructed in a transformed scale, the forecasts obtained in that metric may need to be retransformed. The problem is that this retransformation procedure introduces bias in the forecasts. Guerrero (1993) proposed a procedure for selecting a variance-stabilizing transformation by which N observations of a time series are grouped into H subseries, so that a local estimate of mean and variance within each subseries can be obtained. The power transformation λ is then estimated by least squares. However the procedure, which is based on fitting a linear regression in logarithms, is sensitive to the size of subseries.

The Box–Cox transformation has been considered for transforming a times series before fitting an ARMA model (see Box and Jenkins, 1976). However, λ is the index of the power transformation which can be taken as a parameter to be estimated from the observed series. The primary goal of this paper is to select a power transformation of the form given in equation (2) when incorporating an instantaneous power transformation of the data into the time-series analysis. We propose a Bayesian procedure for estimating the index of the power transformation as well as other parameters in time-series models. The posterior distributions of interest are obtained utilizing the Gibbs sampler, a Markov Chain Monte Carlo (MCMC) method. Previous work on Bayesian selection of transformations in linear models include Pericchi (1981) and Sweeting (1984).

This paper is organized as follows. The next section sets forth the posterior density for the ARMA model. The third section briefly reviews the Gibbs sampler. The fourth section illustrates the methodology using some real data sets. We give conclusions in the fifth section.

MODEL AND DISTRIBUTION

A time series $\{y_t^{(\lambda)}\}$ in equation (2) is generated by an ARMA $(p,\ q)$ process if

$$y_t^{(\lambda)} = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i}^{(\lambda)} + e_t - \sum_{j=1}^{q} \theta_j e_{t-j} \tag{3}$$

where p and q are integers and $\{e_t\}$ are independently and identically distributed (i.i.d.) $N(0, \sigma^2)$. We assume that the first p observations of $\mathbf{Y} = (y_1, y_2, \ldots, y_n)$ are fixed. By conditioning on the first p observations, one can write the likelihood function similar to Broemeling and Shaarawy (1988):

$$L(\mathbf{\Theta}, \sigma, \lambda \mid \mathbf{Y}, \hat{\mathbf{e}}) \propto \sigma^{-(n-p)} \exp\left\{-\frac{1}{2\sigma^2}(Y^{(\lambda)} - \mathbf{Z}^{(\lambda)}\mathbf{\Theta})'(\mathbf{Y}^{(\lambda)} - \mathbf{Z}^{(\lambda)}\mathbf{\Theta})\right\} \times \{\lambda^{n-p} I(\lambda \neq 0) + I(\lambda = 0)\} \prod_{i=p+1}^{n} y_i^{\lambda-1} \tag{4}$$

where $\mathbf{\Theta} = (\phi_0, \phi_1, \ldots, \phi_{\rm p},\ \theta_1, \ldots, \theta_q)'$, $\hat{\mathbf{e}} = (\hat{e}_1,\ \hat{e}_2, \ldots, \hat{e}_n)$, $\mathbf{Y}^{(\lambda)} = (y_{p+1}^{(\lambda)},\ y_{p+2}^{(\lambda)}, \ldots, y_n^{(\lambda)})$ and $\mathbf{Z}^{\lambda} = (Z_{p+1}^{(\lambda)}, Z_{p+2}^{(\lambda)}, \ldots, Z_n^{(\lambda)})'$ with $Z_t^{(\lambda)} = (1,\ y_{t-1}^{(\lambda)}, \ldots, y_{t-{\rm p}}^{(\lambda)}, -\hat{e}_{t-1}, \ldots, -\hat{e}_{t-q})'$ and $\hat{e}_t = y_t^{(\lambda)} - \hat{\phi}_0 - \sum_{i=1}^{p} \hat{\phi}_i y_{t-i}^{(\lambda)} + \sum_{j=1}^{q} \hat{y}_j \hat{e}_{t-j}$.

We now turn to a Bayesian treatment. We assume that $\mathbf{\Theta}$ follows $N(\mathbf{\Theta}_0, \mathbf{V}^{-1})$ and σ^2 follows $\mathrm{IG}(v/2,\ v\eta/2)$, where IG denotes the inverse gamma distribution and the hyperparameters are assumed to be known. Moreover, we assume that a prior $\pi(\lambda)$ over Λ. $\pi(\lambda)$ could be a discrete uniform distribution.

Our interest lies in the marginal posterior distributions of $\mathbf{\Theta}$, σ^2, and λ. Denoting the conditional probability density of w given D by $p(w \mid D)$, and using some standard Bayesian techniques (e.g. DeGroot, 1970; Box and Tiao, 1973), we obtain the following results:

(1) The conditional posterior distribution of $\mathbf{\Theta}$ is

$$P(\mathbf{\Theta} \mid \mathbf{Y}, \hat{\mathbf{e}}, \sigma^2, \lambda) \sim N(\mathbf{\Theta}^*, \mathbf{V}^{*-1}) \tag{5}$$

where

$$\mathbf{\Theta}^* = \left(\frac{\mathbf{Z}^{(\lambda)'}\mathbf{Z}^{(\lambda)}}{\sigma^2} + \mathbf{V}\right)^{-1} \left(\frac{\mathbf{Z}^{(\lambda)'}\mathbf{Z}^{(\lambda)}}{\sigma^2}\hat{\mathbf{\Theta}} + \mathbf{V}\mathbf{\Theta}_0\right)$$

and

$$\mathbf{V}^* = \left(\frac{\mathbf{Z}^{(\lambda)'}\mathbf{Z}^{(\lambda)}}{\sigma^2} + \mathbf{V}\right)$$

with $\hat{\mathbf{\Theta}} = (\mathbf{Z}^{(\lambda)'}\mathbf{Z}^{(\lambda)})^{-1}\mathbf{Z}^{(\lambda)'} Y^{(\lambda)}$.

(2) The conditional posterior distribution of σ^2 is

$$P(\sigma^2 \mid \mathbf{Y}, \hat{\mathbf{e}}, \mathbf{\Theta}, \lambda) \sim \mathrm{IG}\left(\frac{v + n - p}{2}, \frac{v\eta + (n-p)s^2}{2}\right) \tag{6}$$

i.e.

$$\frac{v\eta + (n-p)s^2}{\sigma^2} \sim \chi^2_{v+n-p}$$

where $s^2 = (Y^{(\lambda)} - \mathbf{Z}^{(\lambda)}\mathbf{\Theta})'(Y^{(\lambda)} - \mathbf{Z}^{(\lambda)}\mathbf{\Theta})/(n-p)$.

(3) The conditional posterior probability function of λ is

$$P(\lambda \mid \mathbf{Y}, \hat{\mathbf{e}}, \mathbf{\Theta}, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y}^{(\lambda)} - \mathbf{Z}^{(\lambda)}\mathbf{\Theta})'(\mathbf{Y}^{(\lambda)} - \mathbf{Z}^{(\lambda)}\mathbf{\Theta})\right\} \times \{\lambda^{n-p} I(\lambda \neq 0) + I(\lambda = 0)\} \prod_{i=p+1}^{n} y_i^{\lambda-1} \pi(\lambda)$$

Suppose a prior for λ is flat on the set $\Lambda = \{a_1, a_2, \ldots, a_k\}$. Given prior probabilities $\pi_j \equiv P(\lambda = a_j)$ such that $\Sigma_{j=1}^k \pi_j = 1$, the conditional posterior probability function of λ is a multinomial distribution with probability

$$P(\lambda = a_j \mid \mathbf{Y}, \hat{\mathbf{e}}, \mathbf{\Theta}, \sigma^2) = \frac{P(\mathbf{Y}, \hat{\mathbf{e}}, \mathbf{\Theta}, \sigma^2, \lambda = a_j)}{\sum_{i=1}^{k} P(\mathbf{Y}, \hat{\mathbf{e}}, \mathbf{\Theta}, \sigma^2, \lambda = a_i)} \tag{7}$$

where

$$P(\mathbf{Y}, \hat{\mathbf{e}}, \mathbf{\Theta}, \sigma^2, \lambda = a_i) = f(\mathbf{Y}, \hat{\mathbf{e}} \mid \mathbf{\Theta}, \sigma^2, \lambda = a_i) P(\mathbf{\Theta}, \sigma^2 \mid \lambda = a_i)\pi_i$$

In many application, we are also interested in forecasting the 'future observations'. As in equation (4), the predictive distribution of $y_{n+1}^{(\lambda)}$ is

$$P(y_{n+1}^{(\lambda)} \mid \mathbf{Y}, \hat{\mathbf{e}}, \sigma^2, \lambda) \sim N(\mathbf{\Theta}' \mathrm{Z}_{n+1}^{(\lambda)}, \sigma^2) \tag{8}$$

where $\mathrm{Z}_{n+1}^{(\lambda)}$ is as defined before.

All conditional densities are available, and we will implement the Gibbs sampler which is discussed in the next section.

GIBBS SAMPLER

The Gibbs sampler is a Markov Chain Monte Carlo method for estimating desired posterior distributions from conditional distributions. A great advantage of the Gibbs sampler is its ease in implementation which makes use of the modern computational capabilities to draw inference using simulation techniques. The sampler is especially useful in extracting marginal distributions from fully conditional distributions when the joint distribution is not easily obtained. Geman and Geman (1984) showed that under mild conditions the Gibbs sampler provides a consistent estimate of the marginal distribution of interest.

In recent years, due to the work of Gelfand and Smith (1990) and Gelfand *et al.* (1990), the Gibbs sampler has been shown to be a useful tool for applied Bayesian inference in a wide variety of statistical problems. Moreover, Carlin and Chib (1995) employed MCMC method on Bayesian model choice. In time-series analysis, the Gibbs sampler has already been successfully employed in handling random level-shift models, additive outliers, missing values, and random variance-shift models in a autoregression (e.g. McCulloch and Tsay, 1993, 1994a; Tiao and Tsay, 1991). Chen (1992) and Chen and Lee (1995) applied the sampler to analyse bilinear models and threshold autoregressive models, respectively. More recently, Chen, McCulloch and Tsay (1997) proposed a unified approach to estimating and modelling univariate time series via a Gibbs sampler. The sampler has also been employed by Albert and Chib (1993) and McCulloch and Tsay (1994b) for modelling Markov switching econometric models. It is impossible to review all the recent work in this fast-growing area.

To review the method, we consider the case $(\mathbf{\Theta}, \sigma^2, \lambda)$. Denote the conditional distributions by $f_1(\mathbf{\Theta} \mid \sigma^2, \lambda, \mathbf{Y}, \hat{\mathbf{e}})$, $f_2(\sigma^2 \mid \mathbf{\Theta}, \lambda, \mathbf{Y}, \hat{e})$, and $f_3(\lambda \mid \mathbf{\Theta}, \sigma^2, \mathbf{Y}, \hat{\mathbf{e}})$. The Gibbs sampler employed in this paper proceeds as follows:

(1) Given an arbitrary value $\lambda^{(0)}$, fit a long AR model for $\{y_t^{(\lambda)}\}$ in order to obtain the estimation of $\{e_t\}$, $t = p + 1, \ldots, n$. Generate $\{e_t\}$, $t = 1, \ldots, p$, from $N(0, \sigma^{2(0)})$, where $\sigma^{2(0)}$ is the residual variance of the long AR model.
(2) Use $\sigma^{2(0)}$ as the starting value for σ^2.
(3) Draw $\mathbf{\Theta}^{(1)}$ from $f_1(\mathbf{\Theta} \mid \sigma^{2(0)}, \lambda^{(0)}, \mathbf{Y}, \hat{\mathbf{e}})$, then draw $\sigma^{2(1)}$ from $f_2(\sigma^2 \mid \mathbf{\Theta}^{(1)}, \lambda^{(0)}, \mathbf{Y}, \hat{\mathbf{e}})$, and we complete the first iteration by drawing $\lambda^{(1)}$ from $f_3(\lambda \mid \mathbf{\Theta}^{(1)}, \sigma^{2(1)}, \mathbf{Y}, \hat{\mathbf{e}})$.
(4) Use the ith realization $(\mathbf{\Theta}^{(i)}, \sigma^{2(i)}, \lambda^{(i)})$ in step 3 to refine the series $\hat{\mathbf{e}}$.

After the chain has converged, say at the sth iteration, we can obtain $(\mathbf{\Theta}^{(i)}, \sigma^{2(i)}, \lambda^{(i)})$, $i = s + 1, \ldots, s + N$ as a set of samples from the desired posterior marginals. However, the sample could be highly dependent, each realization being generated from the previous one. To monitor and reduce this dependence, we compute the sample autocorrelations of this chain and select a lag l such that the sample autocorrections at lag $i \geqslant l$ are very small. Then we collect $(\mathbf{\Theta}^{(i)}, \sigma^{2(i)}, \lambda^{(i)})$, $i = s + l, \ldots, s + Nl$, as a random sample. We now employ the parallel chain strategy recommended by Gelman and Rubin (1992), instead of a single string. We run M parallel Markov chains and collect the samples from each chain after convergence has been achieved, in which case we would get MN/l independently and identically distributed observations. The main idea of the recommended strategy is to use independent sequences based on the overdispersion criterion. We monitor convergence of the iterative simulation by estimating the factor by which the scale of the current distribution might be reduced if the simulations were continued in the limit $N \to \infty$. The potential scale reduction is estimated by $(\hat{R})^{1/2}$, where $\hat{R}$ is the ratio of the current variance estimate to the within-sequence variance, with a factor to account for the extra variance of the Student t-distribution. If the potential scale reduction is high, then we have reason to believe that proceeding with further simulations may improve our inference about the target distribution. Once $\hat{R}$ is near 1 for all scalar estimands of interest, it is typically desirable to summarize the target distribution by a set of simulations.

ILLUSTRATIVE EXAMPLES

In this section we illustrate the proposed methodology with two real data sets, focusing on inferences about $\mathbf{\Theta}, \sigma^2$, and λ. The convergence of the Gibbs samplers are monitored by examining the Gelman and Rubin (1992) statistics based on $M = 6$ independent parallel Markov chains.

Example 1

We consider the well-known Nicholson (1950) blowfly data. A fixed number of adult blowflies with balanced sex ratios were kept inside a cage and given a fixed amount of food daily. The blowfly population was then counted every other day for approximately two years, giving a total

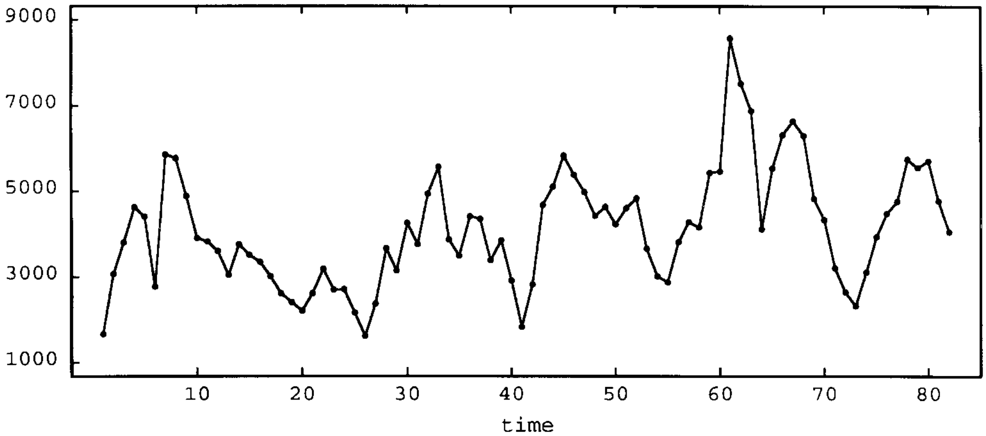

Figure 1. Time plot of blowfly data. The x-axis is day

of 364 observations. Tong (1990) fitted a threshold autoregressive model for the period from 20 to 145 after taking a $\log_{10}$ transformation. Wei (1990) used the series from 218 to 299 in his analysis. For comparison, we are interested in analyzing the series from 218 to 299 which consist of 82 observations as shown in Figure 1. Wei (1990) calculated the following preliminary residual sum of squares:

$$S(\lambda) = \sum_{t=1}^{n} (y_t^{(\lambda)} - \hat{\mu})^2$$

for $\lambda = \{-1.0, -0.5, 0.0, 0.5, 1.0\}$, where $\hat{\mu}$ is the corresponding sample mean of the transformed series. These calculations suggest that a square root or a logarithmic transformation is needed. The maximum likelihood estimates of AR(1) parameters with $\lambda = 0.5$ was given by Wei (1990). We are interested in making inferences about λ as well as other parameters in the AR(1) model.

The hyperparameters used are $\mathbf{\Theta}_0 = \mathbf{o}$, $\mathbf{V} = 0.1 I_2$, $v_i = 3$, and $\eta = \tilde{\sigma}^2/3$, where I_2 denotes the 2×2 identity matrix and $\tilde{\sigma}^2$ is the residual mean squared error of fitting an AR(2) model to the data. The prior on λ is uniform $\Lambda_1 = \{-2, -1.75, -1.5, \ldots, 2\}$. The Gibbs sampler is run for 2000 iterations. We record every fifth value in the sequence of the last 800 in order to have more clearly independent contributions. The point estimates and estimates of the standard deviations for each parameters are given in Table I(a). The result suggests that the square root

Table I. The parameter estimates of blowfly data

	Par.	ϕ_0	ϕ_1	σ^2	λ
(a)	Mean	4.737	0.923	49.763	0.499
λ on $\Lambda_1 = \{-2, -1.75, \ldots, 2\}$	s.e.	0.091	0.002	0.287	0.0004
	Median	4.714	0.925	49.687	0.500
	Mode	4.989	0.924	49.644	0.500
(b)	Mean	4.565	0.829	7.161	0.390
λ on $\Lambda_2 = \{0.30, 0.31, \ldots, 0.5\}$	s.e.	0.079	0.003	0.165	0.001
	Median	4.527	0.822	5.673	0.39
	Mode	4.508	0.865	8.353	0.38

transformation should be adopted rather than the logarithmic transformation. We also employ extended sample autocorrelation function (EACF) proposed by Tsay and Tiao (1984) to the series by taking the square root and the logarithmic transformations separately. The choice of ARMA orders for two transformed data are all AR(1). Under the AR(1) model, our proposed procedure which is uniform over Λ_1 strongly favours the square root transformation. It is worth mentioning that $\lambda = 0.5$ for almost all realizations. The reason is that the probability of $\lambda = 0.5$ in equation (7) is approximately 1.0 and the probability of other values of λ is approximately 0.0. Moreover, if we partition the interval more finely on Λ, from -2 to $+2$ with increment 0.1, then the suggested transformation is $\hat{\lambda} = 0.4$. Due to limited space, the detailed results are omitted. We re-analysed further with λ on $\Lambda_2 = \{0.30, 0.31, \ldots, 0.50\}$, a much finer partition on 0.3 to 0.5. The results are given in Table I(b). When we adopt a uniform prior over Λ_2 for λ, the estimate of λ turns out to be 0.39.

The estimates of λ are 0.5 and 0.39 with uniform priors on Λ_1 and Λ_2, respectively. Figures 2 and 3 show the respective frequency distributions of λ for the two selected priors. We would like to evaluate these estimates of λ with respect to the performance of one-step-ahead forecasts. Using the post-sample data reserved for the purpose, 10 one-step-ahead forecasts done sequentially with $\hat{\lambda} = 0.5$ and $\hat{\lambda} = 0.39$ are produced and compared. We apply the Gibbs sampler using the three conditional distributions in equations (5), (6), and (8) for 2000 iterations, where $N = 800$ and $l = 5$. The estimated posterior medians are collected. An inverse transformation is required in order to produce equivalent forecasts. Table II lists RMSE (root mean square errors), MAE (mean absolute errors), MPE (mean percentage errors), and MAPE (mean absolute percentage errors). The results in Table II show that when $\hat{\lambda} = 0.39$ its RMSE, MAE, MPE, MAPE are smaller. Better forecasts are obtained with $\hat{\lambda} = 0.39$ than with the square root transformed data.

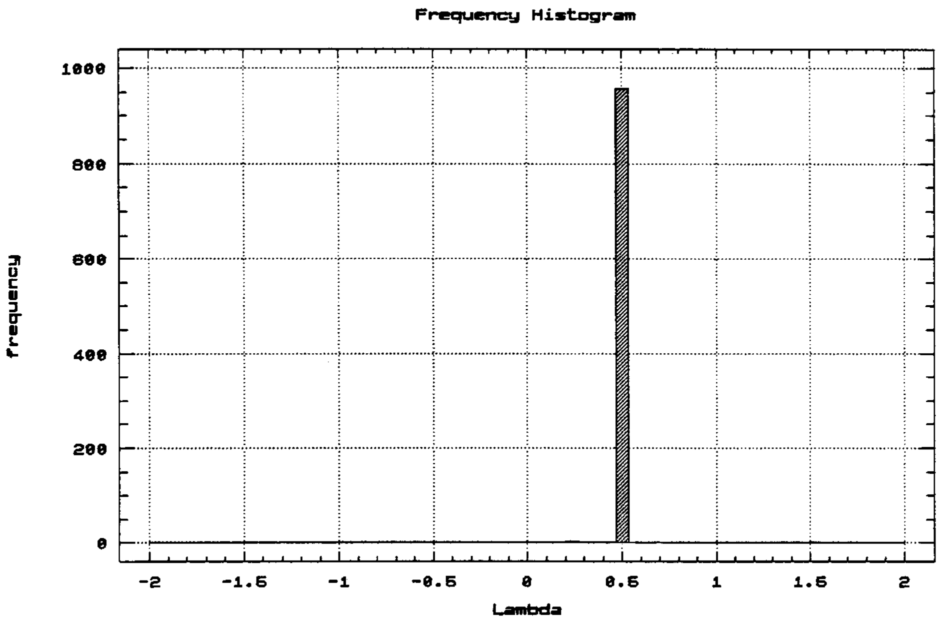

Figure 2. Frequency distributions of λ with uniform prior over $\Lambda_1 = \{-2, -1.75, -1.5, \ldots, 2\}$ in Example 1

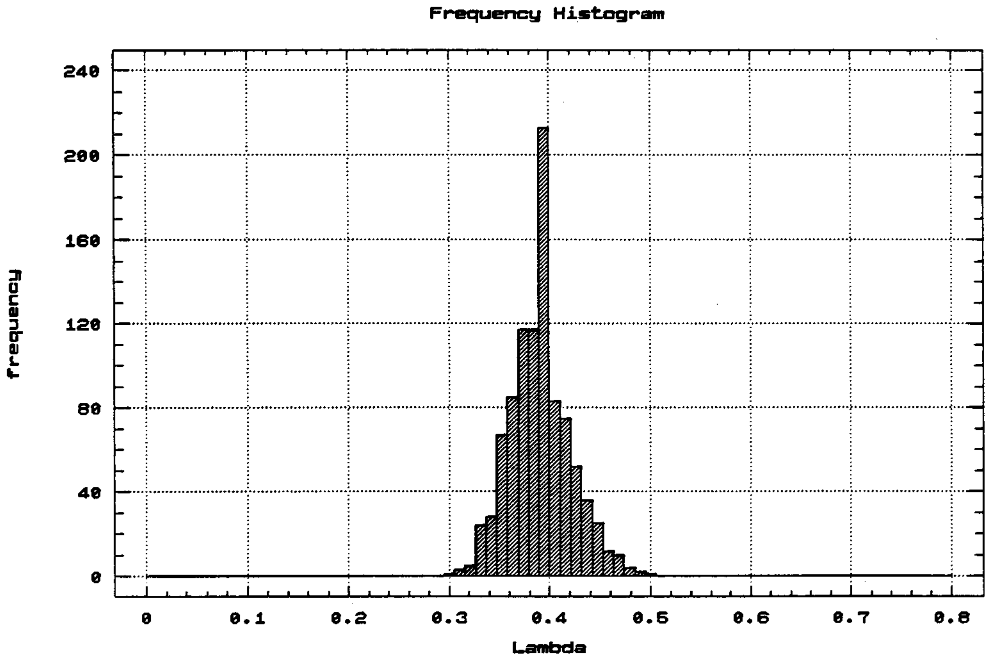

Figure 3. Frequency distributions with λ with uniform prior over $\Lambda_2 = \{0.30, 0.31, \ldots, 0.50\}$ in Example 1

Table II. Comparison of the one-step-ahead prediction errors

Estimate of λ	RMSE	MAE	MPE	MAPE
$\hat{\lambda} = 0.39$	830.138	738.338	0.993%	18.568%
$\hat{\lambda} = 0.5$	992.055	884.535	5.276%	21.377%

We compare our results to a variance-stabilizing power transformation suggested by Guerrero (1993). The method is briefly described as follows:

$$\log(S_h) = \log(a) + (1 - \lambda)\log(\bar{y}_h) + \varepsilon_h, \quad h = 1, \ldots, H$$

with

$$\bar{y}_h = \sum_{r=1}^{R} y_{(h-1)R+r}/R \quad \text{and} \quad S_h = \left[\sum_{r=1}^{R}(y_{(h-1)R+r} - \bar{y}_h)^2/(R-1)\right]^{1/2}$$

where $a > 0$ and the ε_h's are a random sample of errors uncorrelated with $\log(\bar{y}_h)$, whose mean is zero and variance σ_ε^2. Guerrero (1993) suggested taking $R = 2$ when no seasonality is present. The resulting least squares estimates of λ are presented in Table III.

We clearly see that the procedure suggested by Guerrero (1993) based on fitting a linear regression in logarithms is sensitive to the size of subseries, R. The interval of $\hat{\lambda}$ is $(-1, 1.1)$ which is heavily dependent on the choice of the subseries size. Moreover, the estimate of λ with $R = 2$ as suggested by Guerrero (1993) is different from our result as well as Wei's.

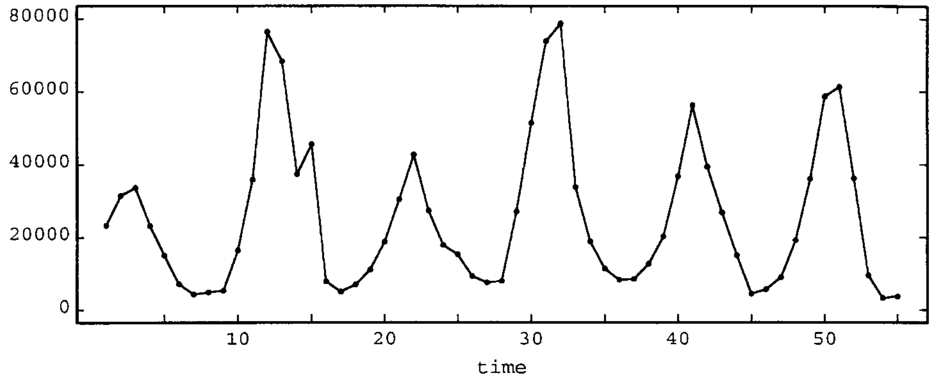

Figure 4. The yearly number of lynx pelts sold in Canada between 1857 and 1911

Example 2

In this example we would like to study the yearly number of lynx pelts sold by the Hudson's Bay Company in Canada between 1857 and 1911. Wei (1990) fitted an ARMA(2,1) as well as AR(3) for the logged Canadian lynx data. The series is plotted in Figure 4. A related series that has been extensively analysed in the literature is the number of Canadian lynx trapped between the years 1821 to 1934. Various linear and non-linear models have been proposed for this data set after taking logarithmic transformation. In general, for this series it seems that different data spans would suggest different models.

We entertain a model ARMA(2,1) in our analysis. We employ the residuals of AR(6) of $y_t^{(\lambda)}$ with $\lambda = 0$ as an initial estimation of $\mathbf{e}$ and use hyperparameters:

$$\mathbf{\Theta}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{V} = \begin{bmatrix} 0.1 & 0.04 & 0.04 & 0.04 \\ 0.04 & 0.1 & 0.04 & 0.04 \\ 0.04 & 0.04 & 0.1 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.1 \end{bmatrix}$$

$v_i = 3, \eta = \tilde{\sigma}^2/3$, where $\tilde{\sigma}$ is the residual standard error of the AR(2) model. We choose a prior on λ which is uniform over $\Lambda = \{x : x = -1 + kd,\ r, d = 0.1,\ k = 0, \ldots, 2/d\}$. Table IV lists the point estimates and estimates of the standard error by the Gibbs sampler with 2200 iterations which are recorded at every fifth value in the sequence of the last 800. Figure 5 plots the frequency distribution of λ. The result in Table IV indicates that taking a transformation with $\lambda = 0.1$ is required which is slightly different from $\lambda = 0.0$ as is usually given in the literature. However, we would probably use logarithmic transformation in practice. Note that the choice of hyperparameters is not sensitive. We also use $\mathbf{V} = 0.1\mathbf{I}_4$ as a hyperparameter which produces a similar result and hence is omitted.

Table III. Selection of λ for different subseries sizes for blowfly data suggested by Guerrero (1993)

Size of the subseries (R)	2	3	4	5	7	8	10
$\hat{\lambda}$	1.078	0.398	0.514	0.053	0.323	−0.994	−0.094

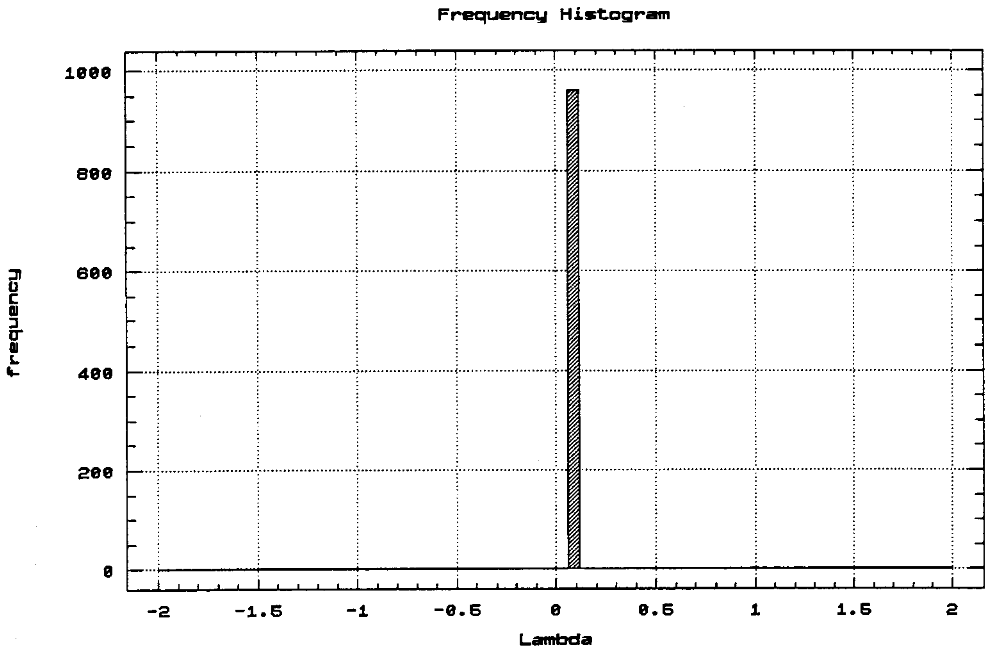

Figure 5. Frequency distributions of λ with uniform prior $\Lambda = \{x : x = -1 + kd, \quad r, d = 0.1, k = 0, \ldots, 2/d\}$ in Example 2

Table IV. The parameter estimates of Canadian lynx data from 1857 to 1911

Par.	ϕ_0	ϕ_1	ϕ_2	θ_1	σ^2	λ
Mean	0.984	1.580	−0.946	0.591	0.014	0.100
s.e.	0.011	0.008	0.007	0.013	0.0004	0.000
Median	1.009	1.540	−0.909	0.601	0.013	0.100
Mode	1.002	1.439	−0.852	0.600	0.011	0.100

Table V. Selection of λ for different subseries sizes for Canadian lynx data suggested by Guerrero (1993)

Size of the subseries (R)	2	3	4	5	6
$\hat{\lambda}$	−0.1215	−0.1488	0.1773	0.1434	−0.1694

Selection of λ for different choices of subseries size suggested by Guerrero (1993) is given in Table V. Again, the procedure suggested by Guerrero (1993) based on fitting a linear regression in logarithms is somewhat sensitive to the size of subseries, R.

CONCLUSIONS

Many biological and economic time series are non-stationary in the variance. To overcome this problem, we need a proper variance stabilizing transformation. To stabilize the variance, we can use power transformation. Frequently, the transformation not only stabilizes the variance but also improves the approximation to normality.

In this work we have developed a practical Bayesian approach on selecting a power transformation in time-series analysis. From the illustrative examples, we find that the estimates of λ almost mass on some certain points. The reason is that the conditional posterior probability function of λ is a multinomial distribution. The numerator in the computed posterior probability for a given value of λ induces a large difference for each probability. It is worth mentioning that the choice of λ depends on the chosen prior space Λ. The chosen λ might just represent the best out of several poor choices of the model. Hence, it is advisable to select an appropriate prior for λ, e.g., the range of a uniform prior from -2 to $+2$ with increment 0.1. The estimates of λ may be slightly different if the prior space of the power transformation were discretized more finely. However, the minor difference is often overlooked in practice.

The results obtained in the illustrative examples show that the Gibbs sampler indeed offers an attractive alternative to the other methods. Based on this work, it should be possible to analyse other models, for example the transfer function model and bilinear time-series models.

ACKNOWLEDGEMENTS

The authors thank an anonymous referee for constructive comments which have significantly improved this article. They are also grateful for research support from the National Science Council of Taiwan (grant NSC84-2121-M-035-007 and NSC86-2115-M-009-027).

REFERENCES

Albert, J. and Chib, S., 'Bayesian inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts', *J. Bus. Econ. Statist.*, **11** (1993), 1–15.

Broemeling, L. D. and Shaarawy, S., 'Time series: A Bayesian analysis in the time domain', in Spall, J. (ed.), *Bayesian Analysis of Time Series and Dynamic Models*, New York: Marcel Dekker, 1988, 1–12.

Box, G. E. P. and Cox, D. R., 'An analysis of transformations', *Journal of the Royal Statistical Society,* **B26** (1964), 211–43.

Box, G. E. P. and Jenkins, G. M., *Time Series Analysis Forecasting and Control*, 2nd edition, San Francisco: Holden-Day, 1976.

Box, G. E. P. and Tiao, G. C., *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley, 1973.

Carlin, B. P. and Chib, S., 'Bayesian model choice via Markov Chain Monte Carlo methods', *Journal of the Royal Statistical Society,* **B57** (1995), 473–84.

Chatfield, C. and Prothero, D. L., 'Box–Jenkins seasonal forecasting: problems in a case-study', *Journal of the Royal Statistical Society,* **A136** (1973), 295–315.

Chen, C. W. S., 'Bayesian analysis of bilinear time series models: a Gibbs sampling approach', *Commun. Statist.—Theory Meth.,* **21** (1992), 3407–25.

Chen, C. W. S. and Lee, J. C., 'Bayesian inference of threshold autoregressive models', *Journal of Time Series Analysis,* **16** (1995), 483–92.

Chen, C. W. S., McCulloch, R. E. and Tsay, R. S., 'A unified approach to estimating and modeling linear and nonlinear time series', *Statistica Sinica*, **7** (1997), 451–72.

DeGroot, M., *Optimal Statistical Decisions*, New York: McGraw-Hill, 1970.

Draper, N. and Smith, H., *Applied Regression Analysis*, 2nd edition, New York: John Wiley, 1981.

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

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), 972–85.
Gelman, A. and Rubin, D. B., 'Inference from iterative simulation using multiple sequences with discussion', *Statistical Science,* **7(4)** (1992), 457–511.
Geman, S. and Geman, D., 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Trans. on Pat. Anal. and Mach. Intel.,* **6** (1984), 721–41.
Guerrero, V. M., 'Time-series analysis supported by power transformations', *Journal of Forecasting,* **12** (1993), 37–48.
Hopwood, W. S., Mckeown, J. C. and Newbold, P., 'Time series forecasting models involving power transformation', *Journal of Forecasting,* **3** (1984), 57–61.
McCulloch, R. E. and Tsay, R. S., 'Bayesian inference and prediction for mean and variance shifts in autoregressive time series', *Journal of the American Statistical Association,* **88** (1993), 968–78.
McCulloch, R. E. and Tsay, R. S., 'Bayesian analysis of autoregressive time series via the Gibbs sampler', *Journal of Time Series Analysis,* **15** (1994a), 235–50.
McCulloch, R. E. and Tsay, R. S., 'Statistical inference of economic time series via Markov switching models', *Journal Time Series Analysis,* **15** (1994b), 523–39.
Nelson, H. L. and Granger, C. W. J., 'Experience with using the Box–Cox transformation when forecasting economic time series', *Journal of Econometrics,* **10** (1979), 57–69.
Nicholson, A. J., 'Population oscillations caused by competition for food', *Nature,* **165** (1950), 476–77.
Pericchi, L. R., 'A Bayesian approach to transformations to normality', *Biometrika*, **68** (1981), 35–43.
Sweeting, T. J., 'On the choice of prior distribution for the Box-Cox transformed linear model', *Biometrika*, **71** (1984), 127–34.
Tsay, R. S. and Tiao, G. C., 'Consistent estimates of autoregressive parameters and extended sample autocorrection function for stationary and nonstationary ARIMA models', *Journal of the American Statistical Association,* **88** (1984), 968–78.
Tiao, G. C. and Tsay, R. S., 'Some advances in nonlinear and adaptive modeling in time series analysis', *Journal of Forecasting,* **13** (1994), 109–31.
Tong, H., *Nonlinear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press, 1990.
Tong, H. and Lim, K. S., 'Threshold autoregression, limit cycles and cyclical data', (with discussion), *Journal of the Royal Statistical Society,* **B42** (1980), 245–92.
Wei, W. W. S., *Time Series Analysis: Univariate and Multivariate Methods*, MA: Addison-Wesley, 1990.

Authors' biographies:

Cathy W. S. Chen is Associate Professor of Statistics, Feng-Chia University, Taiwan. She received her Ph.D. in Statistics from National Central University in 1993. Her research interests include time series modelling, nonlinear time series, MCMC method, and capture–recapture methods. She has published papers in *Computational Statistics and Data Analysis, Communications in Statistics—Theory and Methods, Journal of Time Series Analysis*, and *Statistica Sinica*.

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

Authors' addresses:

Cathy W. S. Chen, Department of Statistics, Feng-Chia University, Taichung 407, Taiwan. E-mail: chenws@fcu.edu.tw.

Jack C. Lee, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

