BAYESIAN INFERENCE OF THRESHOLD AUTOREGRESSIVE MODELS

By Cathy W. S. Chen and Jack C. Lee

Feng-Chia University and National Chiao Tung University

First version received November 1993

Abstract. The study of non-linear time series has attracted much attention in recent years. Among the models proposed, the threshold autoregressive (TAR) model and bilinear model are perhaps the most popular ones in the literature. However, the TAR model has not been widely used in practice due to the difficulty in identifying the threshold variable and in estimating the associated threshold value. The main focal point of this paper is a Bayesian analysis of the TAR model with two regimes. The desired marginal posterior densities of the threshold value and other parameters are obtained via the Gibbs sampler. This approach avoids sophisticated analytical and numerical multiple integration. It also provides an estimate of the threshold value directly without resorting to a subjective choice from various scatterplots. We illustrate the proposed methodology by using simulation experiments and analysis of a real data set.

Keywords. Arranged autoregression; changepoints; Gibbs sampler; Metropolis algorithm; threshold autogressive model.

1. INTRODUCTION

Non-linear time series has drawn much attention in recent years, where many classes of models have been proposed. The threshold autoregressive (TAR) model and bilinear model are perhaps the most popular non-linear time series models in the literature. The TAR model was proposed by Tong (1978, 1983) and Tong and Lim (1980) as an alternative model for describing periodic time series. The major features of this class of models include limit cycles, amplitude dependent frequencies and jump phenomena which linear models fail to capture. Nevertheless, the TAR model has not been widely used in practice. Tsay (1989) pointed out that the reason for this is that it is hard in practice to identify the threshold variable and to estimate the associated threshold values. The procedure proposed by Tong and Lim (1980) is complicated. Tsay (1989) adopted an arranged autoregression for TAR models to produce some predictive residuals which can be used to locate the threshold values by using various scatterplots. McCulloch and Tsay (1993a) proposed a Bayesian procedure for detecting threshold values in a TAR model via some posterior probability plots. However, it is a matter of subjective choice to determine threshold values from the plots. Hence for a given data set, it is conceivable that different people would choose different threshold values. The main objective of this work, therefore, is to propose a procedure for estimating the threshold values and other parameters objectively.

The idea of our study is to transform a TAR model into a changepoint problem in linear regression via arranged autoregression. An arranged autoregression is an autoregression with cases rearranged based on the values of a particular regressor. Our focus is on a Bayesian analysis of threshold autoregression models. In particular, desired marginal posterior densities of interest are obtained utilizing the Gibbs sampler, a Monte Carlo method. Therefore, the method of Carlin *et al.* (1992) for changepoint problems in linear regression analysis is applicable. Broemeling and Cook (1992) and Geweke and Terui (1993) also proposed Bayesian analysis of two-regime threshold autoregressions. However, our approach avoids sophisticated analytical and numerical multiple integration.

Section 2 presents the TAR models. Section 3 briefly reviews the Gibbs sampler as well as the Metropolis algorithm. In Section 4, the necessary conditional distributions for implementing the Gibbs sampler are given. Section 5 illustrates the methodology using simulated and real data sets. We give concluding remarks in Section 6.

2. THE THRESHOLD AUTOREGRESSIVE MODEL

A time series y_t is said to follow a TAR model with threshold variable y_{t-d} if it satisfies

$$y_t = \phi_0^{(k)} + \sum_{i=1}^{p_k} \phi_i^{(k)} y_{t-i} + a_t^{(k)} \qquad \text{for } r_{k-1} \leqslant y_{t-d} < r_k \tag{1}$$

where $k = 1, \ldots, g$ and d is a positive integer. The innovation $\{a_t^{(k)}\}$ is a sequence of independent and identically distributed (i.i.d.) normal random variates with mean zero and variance σ_k^2, where $\{a_t^{(i)}\}$ and $\{a_t^{(j)}\}$ are independent if $i \neq j$. The real numbers r_j satisfy $-\infty = r_0 < r_1 < \cdots < r_g = \infty$ and form a partition of the space of y_{t-d}. The positive integer d is commonly referred to as the delay (or threshold lag) of the model. The partition $r_{k-1} \leqslant y_{t-d} < r_k$ is referred to as the kth regime of the TAR model in (1). We denote the TAR model in (1) by TAR$(g; p_1, \ldots, p_g)$. The TAR model is a piecewise linear model in the space of y_{t-d}, but not a piecewise linear model in time. Tong (1990) provides an excellent review of properties of the TAR model.

3. GIBBS SAMPLER

The Gibbs sampler is a 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.

Recently, 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. For example, Carlin *et al.* (1992) applied the sampler to handle changepoint problems, George and McCulloch (1993) proposed a variable selection method in multiple regression via Gibbs sampling and Zeger and Karim (1991) used it to study generalized linear models with random effects.

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 autoregression, e.g. McCulloch and Tsay (1993b, 1994a) and Tiao and Tsay (1994). The sampler has also been employed by Albert and Chib (1991) and McCulloch and Tsay (1994b) for modeling Markov switching econometric models, and by Chen (1992) for estimating bilinear models. It is impossible to review all the recent work in this fast-growing area.

To review the method, we consider the case of k parameters $(\theta_1, \theta_2, \ldots, \theta_k)$ and suppose that the conditional distribution of each parameter given the others has a simple form from which a random sample can be drawn. Denote the conditional distributions by $f_1(\theta_1|\theta_2, \theta_3, \ldots, \theta_k, y)$, $f_2(\theta_2|\theta_1, \theta_3, \ldots, \theta_k, y)$, $\ldots$, $f_k(\theta_k|\theta_1, \theta_2, \ldots, \theta_{k-1}, y)$, where $y = (y_1, y_2, \ldots, y_n)'$ denotes the data. The Gibbs sampler employed in this paper then works as follows. Given arbitary values $(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)})$ draw $\theta_1^{(1)}$ from $f_1(\theta_1| \theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_k^{(0)}, y)$, then draw $\theta_2^{(1)}$ from $f_2(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_k^{(0)}, y)$ and continue until we complete the first iteration by drawing $\theta_k^{(1)}$ from $f_k(\theta_k|\theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_{k-1}^{(1)}, y)$. After a large number, say M, of iterations, we obtain $(\theta_1^{(M)}, \theta_2^{(M)}, \ldots, \theta_k^{(M)})$.

The desired posterior marginals can be approximated by the empirical marginal distributions of the N values $(\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_k^{(i)})$, $i = M + 1, \ldots, M + N$, where M is large enough to give sufficient precision to the empirical distribution of interest.

When the conditional densities are not easily identified, such as in cases without conjugate priors, the Metropolis algorithm (Metropolis *et al.* 1953) or importance sampling methods can be employed. Chapter 9 of Hammersley and Handscomb (1964) and Kuo *et al.* (1993) provide some discussions of why the Metropolis algorithm works. We describe the Metropolis algorithm briefly at the end of Section 4.

4. THE POSTERIOR DISTRIBUTIONS

Consider a TAR$(2; p_1, p_2)$ process,

$$y_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p_1} \phi_i^{(1)} y_{t-i} + a_t^{(1)} & y_{t-d} \leq r \\ \phi_0^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} y_{t-i} + a_t^{(2)} & y_{t-d} > r \end{cases} \tag{2}$$

The process described by (2) is a special case of the TAR$(g; p_1, \ldots, p_g)$ of (1).

Let $p = \max\{p_1, p_2\}$. We assume that the first p observations $(y_1, y_2, \ldots, y_p)$ are fixed. Let π_i be the time index of the ith smallest observation of $\{y_{p+1-d}, y_{p+2-d}, \ldots, y_{n-d}\}$. By conditioning on the first p observations, one can write the likelihood function as

$$L(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \sigma_1^2, \sigma_2^2, r, d|\boldsymbol{Y}) \propto \sigma_1^{-s}\sigma_2^{-(n-p-s)} \times \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{s}\left(y_{\pi_i+d} - \phi_0^{(1)} - \sum_{k=1}^{p_1}\phi_k^{(1)}y_{\pi_i+d-k}\right)^2 - \frac{1}{2\sigma_2^2}\sum_{i=s+1}^{n-p}\left(y_{\pi_i+d} - \phi_0^{(2)} - \sum_{k=1}^{p_2}\phi_k^{(2)}y_{\pi_i+d-k}\right)^2\right\} \tag{3}$$

where s satisfies $y_{\pi_s} \leq r < y_{\pi_{s+1}}$, $\boldsymbol{Y} = (y_{\pi_1+d}, y_{\pi_2+d}, \ldots, y_{\pi_{n-p}+d})'$, $\boldsymbol{\Theta}_1 = (\phi_0^{(1)}, \phi_1^{(1)}, \ldots, \phi_{p_1}^{(1)})'$ and $\boldsymbol{\Theta}_2 = (\phi_0^{(2)}, \phi_1^{(2)}, \ldots, \phi_{p_2}^{(2)})'$. The parameters of the TAR$(2; p_1, p_2)$ model to be estimated are $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$, σ_1^2, σ_2^2, r and d. Let $Y_1^* = (y_{\pi_1+d}, y_{\pi_2+d}, \ldots, y_{\pi_s+d})'$ be the observations generated by regime I in order of occurrence and, similarly, $Y_2^* = (y_{\pi_{s+1}+d}, \ldots, y_{\pi_{n-p}+d})'$ for regime II; $X_1^* = (x_{1,\pi_1+d}, x_{1,\pi_2+d}, \ldots, x_{1,\pi_s+d})'$, $X_2^* = (x_{2,\pi_{s+1}+d}, \ldots, x_{2,\pi_{n-p}+d})'$, where $x_{1,t} = (1, y_{\pi_t+d-1}, \ldots, y_{\pi_t+d-p_1})'$ and $x_{2,t} = (1, y_{\pi_t+d-1}, \ldots, y_{\pi_t+d-p_2})'$. This is an arranged autoregression with the first s cases of $\boldsymbol{Y}$ in the first regime and $n - p - s$ in the second regime.

To implement the proposed analysis, it remains to derive the conditional posterior distribution of an unknown parameter given all the others. To this end, we choose the priors as follows.

We take $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$ to be independent $\mathrm{N}(\boldsymbol{\Theta}_{0i}, \boldsymbol{V}_i^{-1})$ and σ_1^2, σ_2^2 independent $\mathrm{IG}(v_i/2, v_i\lambda_i/2)$, $i = 1, 2$, where IG denotes the inverse gamma distribution and the hyper-parameters are assumed to be known. Similar to Geweke and Terui (1993), we assume that r follows a uniform distribution on (a, b) and d follows a discrete uniform distribution on the intergers $1, 2, \ldots, d_0$.

Our interest lies in the marginal posterior distributions of $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$, σ_1^2, σ_2^2, r and d. Denoting the conditional probability density of w given D by $p(w|D)$, and using some standard Bayesian techniques, e.g. DeGroot (1970) and Box and Tiao (1973), we obtain the following results.

(i) The conditional posterior distribution of $\boldsymbol{\Theta}_i$ is independent of $\boldsymbol{\Theta}_j$ for $i \neq j$.

$$p(\boldsymbol{\Theta}_i|\boldsymbol{Y}, \sigma_1^2, \sigma_2^2, r, d) \sim \mathrm{N}(\boldsymbol{\Theta}_i^*, \boldsymbol{V}_i^{*-1})$$

where

$$\boldsymbol{\Theta}_i^* = \left(\frac{X_i^{*\prime}X_i^*}{\sigma_i^2} + V_i\right)^{-1}\left(\frac{X_i^{*\prime}X_i^*}{\sigma_i^2}\hat{\boldsymbol{\Theta}}_i + V\boldsymbol{\Theta}_{0i}\right)$$

and

$$V_i^* = \left(\frac{X_i^{*\prime}X_i^*}{\sigma_i^2} + V_i\right)$$

with $\hat{\boldsymbol{\Theta}}_i = (X_i^{*\prime}X_i^*)^{-1}X_i^{*\prime}Y_i^*$.

(ii) The conditional posterior distribution of σ_i^2 is independent of σ_j^2, for $i \neq j$.

$$p(\sigma_i^2|Y, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, r, d) \sim \mathrm{IG}\left(\frac{\upsilon_i + n_i}{2}, \frac{\upsilon_i\lambda_i + n_i s_i^2}{2}\right)$$

i.e.

$$\frac{\upsilon_i\lambda_i + n_i s_i^2}{\sigma_i^2} \sim \chi^2_{\upsilon_i + n_i} \qquad i = 1, 2$$

where

$$n_1 = \sum_{t=1}^{n-p} I_{\{y_{\pi_t} \leq r\}} \qquad n_2 = \sum_{t=1}^{n-p} I_{\{y_{\pi_t} > r\}} \qquad s_i^2 = n_i^{-1}(Y_i^* - \hat{Y}_i)'(Y_i^* - \hat{Y}_i)$$

with $\hat{Y}_i = X_i^{*\prime}\boldsymbol{\Theta}_i$.

(iii) The conditional posterior probability function of r is

$$p(r|Y, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \sigma_1^2, \sigma_2^2, d) \propto \frac{\exp\{-\sum_{i=1}^2(1/2\sigma_i^2)(Y_i^* - X_i^{*\prime}\boldsymbol{\Theta}_i)'(Y_i^* - X_i^{*\prime}\boldsymbol{\Theta}_i)\}}{\sigma_1^{n_1}\sigma_2^{n_2}}. \tag{4}$$

Notice that n_1 and n_2 are functions of r.

(iv) The conditional posterior probability function of d is a multinomial distribution with probability

$$p(d|Y, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \sigma_1^2, \sigma_2^2, r) = \frac{L(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \sigma_1^2, \sigma_2^2, r, d|Y)}{\sum_{d=1}^{d_0} L(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \sigma_1^2, \sigma_2^2, r, d|Y)}$$

where $d = 1, 2, \ldots, d_0$ and

$$L(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \sigma_1^2, \sigma_2^2, r, d|Y) = \frac{\exp\{-\sum_{i=1}^2(1/2\sigma_i^2)(Y_i^* - X_i^{*\prime}\boldsymbol{\Theta}_i)'(Y_i^* - X_i^{*\prime}\boldsymbol{\Theta}_i)\}}{\sigma_1^{n_1}\sigma_2^{n_2}}.$$

All conditional densities are identified with the exception of r. With regard to r, we will employ the Metropolis algorithm (Metropolis *et al.* 1953). Let $f(r)$ denote the conditional density in (4), suppressing the conditioning variables for simplicity. Since we assume the prior distribution of r is uniform over (a, b), we can use a transition kernel $h(r, r^*)$, where $r^* = \log\{(r - a)/$

$(b - r)\}$, which maps (a, b) into $(-\infty, \infty)$. Then the Metropolis algorithm works as follows.

STEP 1. Start with any point $r^{(0)}$ from the prior $U(a, b)$, and stage indicator $j = 0$.

STEP 2. Generate a point r^* according to the transition kernel $h(r^{(j)}, r^*)$.

STEP 3. Update $r^{(j)}$ to $r^{(j+1)} = r^*$ with probability $p = \min\{1, f(r^*)/f(r^{(j)})\}$, remain at $r^{(j)}$ with probability $1 - p$.

STEP 4. Repeat Steps 2 and 3 by increasing the stage indicator until the process reaches a stationary distribution.

5. ILLUSTRATIVE EXAMPLES

In this section, we illustrate the proposed methodology with a simulation study and a real data set, focusing on inferences about $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$, σ_1^2, σ_2^2, r and d. The results generally are not dependent on the priors selected, and therefore a sensitivity analysis with respect to the prior inputs is not provided here.

5.1. *Simulation*

We first consider 100 data sets generated by a TAR (2; 1, 1).

$$y_t = \begin{cases} \phi_1^{(1)} y_{t-1} + a_t^{(1)} & y_{t-d} \leq r \\ \phi_1^{(2)} y_{t-1} + a_t^{(2)} & y_{t-d} > r \end{cases} \tag{5}$$

where $a_t^{(k)}$ are i.i.d. $N(0, \sigma_k^2)$, for $k = 1, 2$, and $\{a_t^{(1)}\}$, $\{a_t^{(2)}\}$ are independent. The parameter values used to generate the data sets of $n = 200$ observations are given by

$$\begin{array}{ll} \phi_1^{(1)} = -0.5 & \sigma_1^2 = 2.0 \\ \phi_1^{(2)} = 0.5 & \sigma_2^2 = 1.0 \\ r = 0.40 & d = 1 \end{array}$$

The hyper-parameters used are $\boldsymbol{\Theta}_{0i} = 0$, $V_i = 0.1$, $\upsilon_i = 3$ and $\lambda_i = \tilde{\sigma}^2/3$ for $i = 1$ and 2, where $\tilde{\sigma}^2$ is the residual mean squared error of fitting an AR(2) model to the data. Moreover, we choose $d_0 = 4$, $a = p_{10}$ and $b = p_{90}$, where p_k stands for the kth percentile of the data. The Gibbs sampler is run for 2200 iterations. We record every tenth value in the sequence of the last 1000 in order to have more nearly independent contributions. The simulation results are shown in Table I. In columns 3 through 5, the entries for d are the mean, median and standard deviation of 100 posterior modes while those of the other parameters are the corresponding values of 100 posterior means.

TABLE I
RESULTS OF SIMULATION WITH 100 DATA SETS GENERATED FROM MODEL (5)

Parameter	True value	Mean	Median	Standard deviation
$\phi_1^{(1)}$	−0.50	−0.384	−0.376	0.14
$\phi_1^{(2)}$	0.50	0.484	0.478	0.06
σ_1^2	2.0	1.878	1.865	0.30
σ_2^2	1.0	1.151	1.152	0.15
r	0.40	0.401	0.403	0.12
d	1	1	1	0

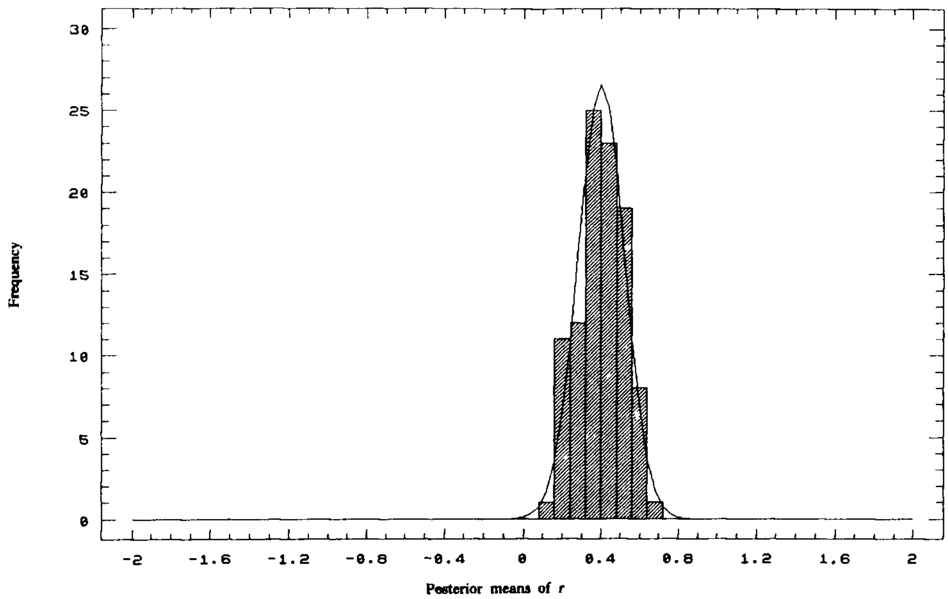

FIGURE 1. Frequency histogram of posterior means of r for 100 realizations.

Figure 1 displays the posterior means of r for the 100 realizations. It appears that the distribution is nearly symmetric.

The results in Table I indicate that the Gibbs sampler gives sound inference in this finite sample case. It is worthwhile mentioning that the posterior modes of d correctly indicate $d = 1$ for all 100 data sets. The estimate of $\phi_1^{(1)}$ still has slightly positive bias, while the estimates of other parameters are approximately unbiased. The bias of $\phi_1^{(1)}$ can be alleviated somewhat by using the posterior median rather than posterior mean.

5.2. *Unemployment rates*

We study the series of changes in US quarterly unemployment rates from the second quarter of 1948 to the first quarter of 1991 which were analyzed recently by McCulloch and Tsay (1993a) (referred to as MT hereinafter). Figures 2(a) and 2(b) give time plots of the unemployment rates and the change series. MT uses the analysis of TAR models in Tsay (1989) to obtain a TAR(2; 4, 2) with $d = 1$. In this work, we do not need to apply any methods to get an *a priori* estimate of d. We are interested in making inferences about d as well as other parameters. The hyper-parameters used are $\boldsymbol{\Theta}_{0i} = 0$, $V_1 = 0.1I_4$, $V_2 = 0.1I_3$, $v_i = 3$, $\lambda_i = \tilde{\sigma}^2/3$ for $i = 1, 2$, and $\tilde{\sigma}^2$ is the residual

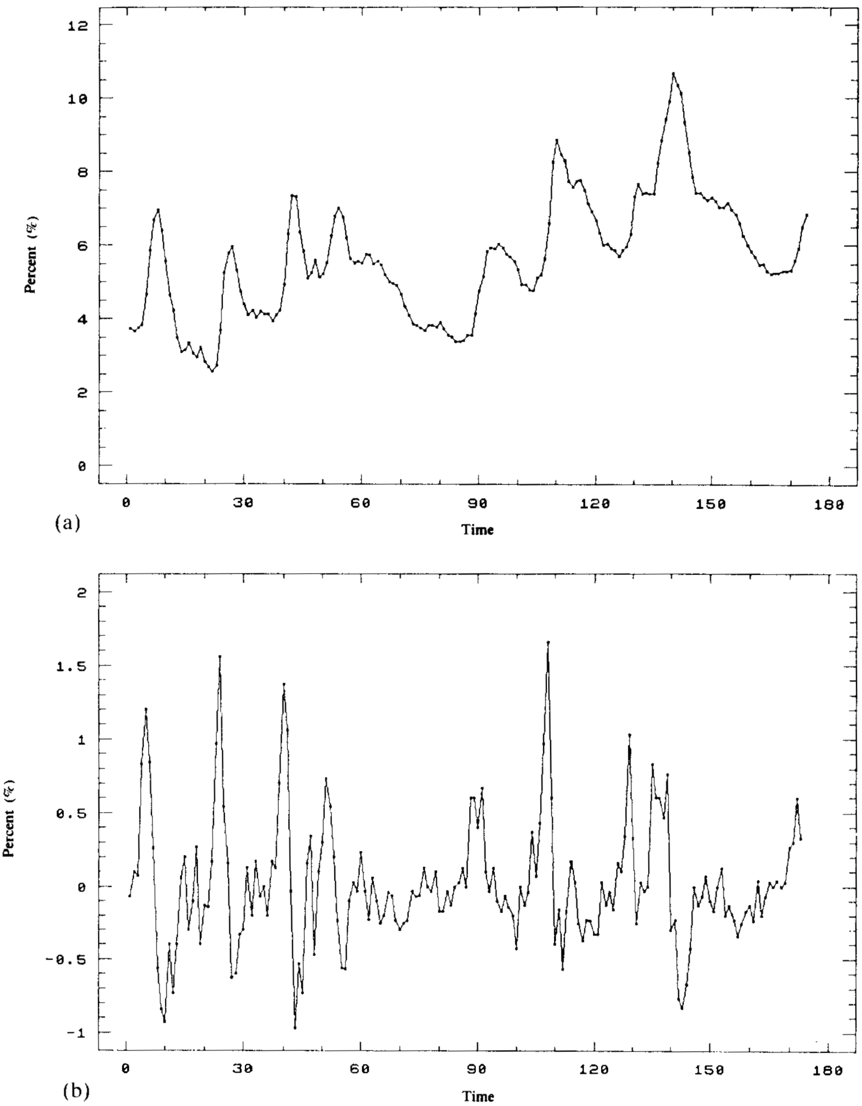

FIGURE 2. (a) Time plot of unemployment rates in the USA; (b) time plot of the change of unemployment rates.

mean squared error of fitting an AR(7) to the data. We also choose $d_0 = 4$, $a = p_{10}$ and $b = p_{90}$, where p stands as before. We ran the Gibbs sampler for 2800 iterations but used every tenth of the last 1000 iterations for making inference. The parameter estimates together with standard errors are given in Table II.

By using our approach, the posterior mode of d is 1 which is identical to that of MT. Figure 6 in MT plots the posterior probability of being a changepoint against the threshold variable y_{t-1} and suggests a single changepoint around $y_{t-1} = 0.30$ which is close to the changepoint value 0.299 in our changepoint estimate. Moreover, the number of observations for each regime is the same for both results although the threshold value is slightly different. A major advantage of using our procedure here is that it avoids determining threshold values subjectively via the plots which might cause imprecision. We should point out, of course, that our procedure is more restrictive because it focuses only on the two-regime TAR model.

6. CONCLUDING REMARKS

We have proposed a Bayesian analysis of a TAR(2; p_1, p_2) process. The idea of the proposed procedure is to transform a TAR model into a changepoint problem in linear regression via arranged autoregression. The main advantage of using the proposed procedure is that it is simple and requires no subjective specification of threshold value via scatterplots. The results obtained in illustrative examples show that the Gibbs sampler indeed offers an attractive alternative to the other methods. Nevertheless, much remains to be done for the general TAR model with possibly multiple regimes.

ACKNOWLEDGEMENTS

The authors thank the anonymous referee and Professor Ruey S. Tsay for constructive comments which have significantly improved this paper. C. W. S. Chen and J. C. Lee acknowledge research support from National Science

TABLE II

THE PARAMETER ESTIMATES OF CHANGES IN THE US UNEMPLOYMENT RATES

		$\phi_1^{(1)}$	$\phi_2^{(1)}$	$\phi_3^{(1)}$	$\phi_4^{(1)}$	n_1	σ_1^2	r	d
Mean		0.564	0.086	−0.086	−0.178	137	0.067	0.299	
Standard error		0.010	0.012	0.010	0.008			0.020	
Mode		0.576	0.078	−0.085	−0.182	137	0.065	0.273	1
	$\phi_0^{(2)}$	$\phi_1^{(2)}$	$\phi_2^{(2)}$			n_2	σ_2^2		
Mean	0.182	0.776	−0.520			32	0.274		
Standard error	0.028	0.030	0.024						
Mode	0.107	0.770	−0.487			32	0.240		

Council of R.O.C. grants NSC82-0115-M-035-012-T and NSC83-0208-M-009-017, respectively.

REFERENCES

ALBERT, J. and CHIB, S. (1991) Bayesian regression analysis of binary data. Paper presented at Ohio State University, February, 1991.

BOX, G. E. P. and TIAO, G. C. (1973) *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

BROEMELING, L. D. and COOK, P. (1992) Bayesian analysis of threshold autoregressions. *Commun. Statist. - Theory Meth.* 21, 2459–82.

CARLIN, B. P., GELFAND, A. E., and SMITH, A. F. M. (1992) Hierarchical Bayesian analysis of changepoint problems. *Appl. Statist.* 41, 389–405.

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

DEGROOT, M. (1970) *Optimal Statistical Decisions*. New York: McGraw-Hill.

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

—, HILLS, S. E., RACINE-POON, A. and SMITH, A. F. M. (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Am. Statist. Assoc.* 85, 972–85.

GEMAN, S. and GEMAN, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pat. Anal. Mach. Intel.* 6, 721–41.

GEORGE, E. I. and MCCULLOCH, R. E. (1993) Variable selection via Gibbs sampling. *J. Am. Statist. Assoc.* 88, 881–89.

GEWEKE, J. and TERUI, N. (1993) Bayesian threshold autoregressive models for nonlinear time series. *J. Time Ser. Anal.* 14, 441–54.

HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1964) *Monte Carlo Methods*. London: Chapman and Hall.

KUO, L., LEE, J., PAI, J. and CHENG, P. (1993) Bayes inference for technological substitution data with data-based transformation. Unpublished manuscript.

MCCULLOCH, R. E. and TSAY, R. S. (1993a) Bayesian analysis of threshold autoregressive processes with a random number regimes. Technical report, Statistics Research Center, Graduate School of Business, University of Chicago.

— and — (1993b) Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *J. Am. Statist. Assoc.* 88, 968–78.

— and — (1994a) Bayesian analysis of autoregressive time series via the Gibbs sampler. *J. Time Ser. Anal.* 15, 235–50.

— and — (1994b) Statistical inference of economic time series via markov switching models. *J. Time Ser. Anal.* 15, 523–39.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N. and TELLER, A. H. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087–91.

TIAO, G. C. and TSAY, R. S. (1994) Some advances in nonlinear and adaptive modeling in time series analysis. *J. Forecasting*, 13, 109–31.

TONG, H. (1978) *On a Threshold Model in Pattern Recognition and Signal Processing* (ed. C. H. Chen). Amsterdam: Sijhoff & Noordhoff.

— (1983) *Threshold Models in Non-linear Time Series Analysis*, Vol. 21 of Lecture Notes in Statistics (ed. K. Krickegerg). New York: Springer-Verlag.

— (1990) *Nonlinear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

— and LIM, K. S. (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *J. R. Statist. Soc. Ser. B.* 42, 245–92.

TSAY, R. S. (1989) Testing and modeling threshold autoregressive process. *J. Am. Statist. Assoc.* 84, 231–40.

ZEGER, S. L. and KARIM, M. R. (1991) Generalized linear models with random effects: a Gibbs sampling approach. *J. Am. Statist. Assoc.* 86, 79–86.