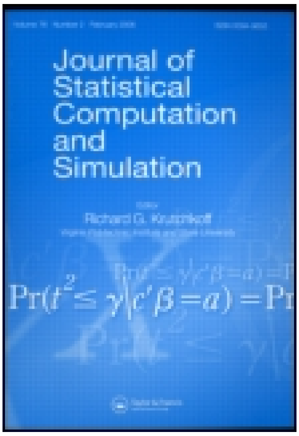

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

Bayesian estimation for time-series regressions improved with exact likelihoods

Cathy W. S. Chen [a] , Jack C. Lee [b c] , Hsiang-Yu Lee [b] & W. F. Niu [b]

[a] Graduate Institute of Statistics and Actuarial Science , Feng Chia University , Taichung, 407, Taiwan, R.O.C.

[b] Institute of Statistics , National Chiao Tung University , Hsinchu, Taiwan, R.O.C.

[c] Graduate Institute of Finance , National Chiao Tung University , Hsinchu, Taiwan, R.O.C.

Published online: 01 Feb 2007.

To cite this article: Cathy W. S. Chen , Jack C. Lee , Hsiang-Yu Lee & W. F. Niu (2004) Bayesian estimation for time-series regressions improved with exact likelihoods, Journal of Statistical Computation and Simulation, 74:10, 727-740, DOI: 10.1080/00949650310001643270

To link to this article: http://dx.doi.org/10.1080/00949650310001643270

Journal of Statistical Computation and Simulation
Vol. 74, No. 10, October 2004, pp. 727–740

BAYESIAN ESTIMATION FOR TIME-SERIES REGRESSIONS IMPROVED WITH EXACT LIKELIHOODS

CATHY W. S. CHEN[a,*], JACK C. LEE[b,c], HSIANG-YU LEE[b] and W. F. NIU[b]

[a]*Graduate Institute of Statistics and Actuarial Science, Feng Chia University, Taichung 407, Taiwan, R.O.C.;*
[b]*Institute of Statistics; and* [c]*Graduate Institute of Finance, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.*

(Received 30 March 2002; In final form 30 October 2003)

We propose an estimation procedure for time-series regression models under the Bayesian inference framework. With the exact method of Wise [Wise, J. (1955). The autocorrelation function and spectral density function. *Biometrika*, **42**, 151–159], an exact likelihood function can be obtained instead of the likelihood conditional on initial observations. The constraints on the parameter space arising from the stationarity conditions are handled by a reparametrization, which was not taken into consideration by Chib [Chib, S. (1993). Bayes regression with autoregressive errors: A Gibbs sampling approach. *J. Econometrics*, **58**, 275–294] or Chib and Greenberg [Chib, S. and Greenberg, E. (1994). Bayes inference in regression model with ARMA(p, q) errors. *J. Econometrics*, **64**, 183–206]. Simulation studies show that our method leads to better inferential results than their results.

Keywords: Autoregressive process; Exact likelihood; Markov chain Monte Carlo; Partial autocorrelations

1 INTRODUCTION

Regression methods have been an integral part of time-series analysis for a long time. Consider a regression model possessing error terms with zero mean and unknown variance. The assumption that the errors are uncorrelated is generally unrealistic. Violations of the independence assumption can be checked using residual plots, the runs test, the Durbin–Watson test, and so on. In most time-series analysis, it is often shown that the covariance matrix of the regression model disturbance terms has a Markov pattern. Thus the regression model can be written in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2\mathbf{V}), \tag{1}$$

where

$$\mathbf{Y} = (y_1, \ldots, y_n)^{\mathrm{T}}, \quad \mathbf{X} = [\mathbf{1}, \mathbf{X}_1, \ldots, \mathbf{X}_k]_{n\times(k+1)} \text{ with } \mathbf{X}_j = (X_{j1}, \ldots, X_{jn})^{\mathrm{T}},$$
$$\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^{\mathrm{T}}, \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}},$$

* Corresponding author. E-mail: chenws@fcu.edu.tw

ISSN 0094-9655 print; ISSN 1563-5163 online
DOI: 10.1080/00949650310001643270

$$\mathbf{V} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-2} & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-3} & \rho_{n-2} \\ \rho_2 & \rho_1 & \cdots & \rho_{n-4} & \rho_{n-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & \rho_1 & 1 \end{bmatrix},$$

and $\sigma_\varepsilon^2\mathbf{V}$ is an autocovariance matrix with theoretical autocorrelations $\rho_0, \rho_1, \ldots, \rho_{n-1}$. In this article, we develop an exact method to analyze time-series regression models in a Bayesian framework. Parameter estimation is done using a Markov chain Monte Carlo (MCMC) method which is a hybrid of the Gibbs sampler and the Metropolis–Hastings (MH) algorithm. Compared with a simple regression model, the major obstacle in estimating a time-series regression model involves treating its initial observations. Bayesian inference for time-series regression regarding autoregressive processes conditional on initial observations has been considered by Chib (1993), McCulloch and Tsay (1994), and Albert and Chib (1993). However, when conditioning on initial observations, the problem setting loses its time-series feature completely and becomes a pure regression problem. In this paper, we consider the exact likelihood function instead of the likelihood conditional on the initial observations. We employ the results of Wise (1955) to obtain an inverse autocovariance matrix in the exact likelihood function. Alternatively, the unobserved history prior to the first observation can be treated as latent variables in the model (*e.g.* Marriott *et al.*, 1996; Chen, 1999). Chib and Greenberg (1994) (denoted C&G henceforth) work with the state space form for ARMA processes. Using the same time-series regression models, both approaches of Chib (1993) and C&G are used to compare with the proposed approach. The constraints on the parameter space arising from the stationarity conditions are handled by a useful reparametrization which was not taken into consideration by Chib (1993) or C&G. Simulation studies show that our method leads to better inferential results when compared with the results presented in Chib (1993) and C&G.

This paper is structured as follows. Section 2 describes the Bayesian setup for time-series regression models. A useful transformation enabling us to handle the stationarity constraint is also given in Section 2. Section 3 illustrates the MCMC methods and shows how the simulation is conducted. Simulation results comparing the performance of our Bayesian method with those of Chib (1993) and C&G are given in Section 4. Applications of our proposed method to three data sets are also given. Finally, Section 5 provides concluding remarks.

2 ESTIMATION

Suppose that $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^{\mathrm{T}}$. The relationship between $\boldsymbol{\rho}$ and $\boldsymbol{\phi}$ follows the Yule–Walker equations. Therefore, the time-series regression model in Eq. (1) is equivalent to

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ \varepsilon_t &= \phi_1\varepsilon_{t-1} + \cdots + \phi_p\varepsilon_{t-p} + a_t, \end{aligned} \tag{2}$$

where a_t is distributed as $N(0, \sigma_a^2)$ and is independent of other errors over time and ε_t is distributed as $N(0, \sigma_\varepsilon^2)$ but is not independent of other errors over time.

We will focus on a general approach to estimate parameters, $(\boldsymbol{\beta}, \sigma_a^2, \boldsymbol{\phi})$, from model (2). The stationarity restrictions on $\boldsymbol{\phi}$ can be applied as in Chib (1993) and C&G. However, their constrained approach becomes more difficult in the absence of explicit constraints on each autoregressive component when the order of the AR(p) model increases (*i.e.* for $p > 2$). Alternatively, we apply the following one-to-one transformation which reparametrizes $\boldsymbol{\phi}$ in terms of the partial autocorrelations η_j (Barndorff-Nielsen and Schou, 1973):

$$\begin{aligned} \phi_k^{(k)} &= \eta_k, \\ \phi_i^{(k)} &= \phi_i^{(k-1)} - \eta_k \phi_{k-i}^{(k-1)}, \quad i = 1, \ldots, k-1, \end{aligned} \tag{3}$$

where $\phi_j^{(p)}$ is the jth coefficient from an AR(p) process and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)^{\mathrm{T}}$. The stationarity condition of $\boldsymbol{\phi}$ becomes $|\eta_i| < 1$, $i = 1, \ldots, p$. Furthermore, we then apply a 'Fisher-type' transformation from $\boldsymbol{\eta}$ to $\boldsymbol{\gamma}$ (Marriott and Smith, 1992):

$$\gamma_j = \log \frac{(1+\eta_j)}{(1-\eta_j)}, \quad j = 1, \ldots, p.$$

The parameter space of $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^{\mathrm{T}}$ is the entire real line. Hence, we assume a multivariate normal distribution for $\boldsymbol{\gamma}$. In the case of AR(1), $\eta_1 = \phi_1$ and $\gamma_1 = \log\{(1+\eta_1)/(1-\eta_1)\}$, which is assumed to be a normally distributed. For the AR(2) process, $\phi_1 = \eta_1(1-\eta_2)$ and $\phi_2 = \eta_2$, where $-1 < \eta_i < 1$, $i = 1, 2$. For the AR(3) process, $\phi_1 = \eta_1 - \eta_1\eta_2 - \eta_2\eta_3$, $\phi_2 = \eta_2 - \eta_1\eta_3 - \eta_1\eta_2\eta_3$, and $\phi_3 = \eta_3$, where $-1 < \eta_i < 1$, $i = 1, 2, 3$. Again, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)^{\mathrm{T}}$ is assumed to follow the normality assumption. All random generation is done in the $\boldsymbol{\gamma}$-space, inverting back to $\boldsymbol{\phi}$ at the end.

The exact likelihood function for the time-series regression in Eq. (1) is

$$L(\boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\rho}|\mathbf{Y}) = (2\pi\sigma_\varepsilon)^{-(n/2)}|\mathbf{V}|^{-(1/2)} \exp\left\{-\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma_\varepsilon^2}\right\}. \tag{4}$$

Let $\sigma_\varepsilon^{-2}\mathbf{V}^{-1} = \sigma_a^{-2}\boldsymbol{\Sigma}^{-1}$, which is the exact inversion of the autocovariance matrix of a pth order autoregressive process obtained by Wise (1955). The definition of $\boldsymbol{\Sigma}^{-1}$ is given later.

Therefore, Eq. (4) can be reparametrized as follows:

$$L(\boldsymbol{\beta}, \sigma_a^2, \boldsymbol{\gamma}|\mathbf{Y}) = (2\pi\sigma_a)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left\{\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma_a^2}\right\}. \tag{5}$$

Suppose $\boldsymbol{\beta}$, σ_a^2, and $\boldsymbol{\gamma}$ are *a priori* independent. That is,

$$\pi(\boldsymbol{\beta}, \sigma_a^2, \boldsymbol{\gamma}) = \pi(\boldsymbol{\beta})\pi(\sigma_a^2)\pi(\boldsymbol{\gamma}).$$

The prior distributions used are as follows:

$$\boldsymbol{\beta} \sim N_k(\boldsymbol{\beta}_a, \mathbf{A_0}), \quad \sigma_a^2 \sim \mathrm{IG}\left(\frac{\upsilon_0}{2}, \frac{\delta_0}{2}\right), \quad \boldsymbol{\gamma} \sim N_p(\boldsymbol{\phi_0}, \boldsymbol{\Phi_0}),$$

where the symbol IG($\upsilon_0/2$, $\delta_0/2$) stands for the inverse gamma distribution, which is equivalent to $\delta_0/\sigma_a^2 \sim \chi_{\upsilon_0}^2$. Also, it is assumed that $\boldsymbol{\gamma}$ satisfies stationarity conditions. It follows that the complete conditional posterior distributions are given by

$$\boldsymbol{\beta} \mid \mathbf{Y}, \sigma_a^2, \boldsymbol{\gamma} \sim N_k(\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{A}}), \tag{6}$$

where

$$\tilde{\boldsymbol{\beta}} = \left(\mathbf{A_0}^{-1} + \frac{\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}}{\sigma_a^2}\right)^{-1}\left(\mathbf{A_0}^{-1}\boldsymbol{\beta}_a + \frac{\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Y}}{\sigma_a^2}\right),$$

$$\tilde{\mathbf{A}} = \left(\mathbf{A_0}^{-1} + \frac{\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}}{\sigma_a^2}\right)^{-1}.$$

$$\sigma_a^2 \mid \mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\gamma} \sim \mathrm{IG}\left(\frac{n+\upsilon_0}{2}, \frac{\delta_0 + d_B}{2}\right), \tag{7}$$

where $d_B = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$,

$$\mathbf{p}(\boldsymbol{\gamma} \mid \mathbf{Y}, \boldsymbol{\beta}, \sigma_a^2) \propto |\boldsymbol{\Sigma}|^{-1/2} \times \exp\left\{-\frac{1}{2}\left[\frac{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})}{\sigma_a^2} + (\boldsymbol{\gamma} - \boldsymbol{\phi_0})^{\mathrm{T}}\Phi_0^{-1}(\boldsymbol{\gamma} - \boldsymbol{\phi_0})\right]\right\}. \tag{8}$$

The advantage of the above reparametrization in implementing the Gibbs sampler is clear. Rather than sampling the awkwardly constrained conditional distributions for $\boldsymbol{\phi}$, we can sample the unconstrained distribution for $\boldsymbol{\gamma}$ based on the posterior in Eq. (8) and invert back to $\boldsymbol{\phi}$. Note that all conditional posterior distributions have closed forms except that of $\boldsymbol{\gamma}$. As the posterior of $\boldsymbol{\gamma}$ is nonstandard, we perform the sampling using the MH algorithm (Metropolis *et al.*, 1953; Hastings, 1970; Chib and Greenberg, 1995).

The exact inversion of the autocovariance matrix of an autoregressive process obtained by Wise (1955) can be found in Chen and Wen (2001). Due to limited space, the inverse autocovariance matrix for $p = 3$ is given below:

$$\frac{\mathbf{V}^{-1}}{\sigma_\varepsilon^2} = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 & -\phi_3 & \cdots & 0 & 0 & 0 \\ -\phi_1 & 1+\phi_1^2 & -\phi_1+\phi_1\phi_2 & -\phi_2+\phi_1\phi_3 & \cdots & 0 & 0 & 0 \\ -\phi_2 & -\phi_1+\phi_1\phi_2 & 1+\phi_1^2+\phi_2^2 & -\phi_1+\phi_1\phi_2+\phi_2\phi_3 & \cdots & 0 & 0 & 0 \\ -\phi_3 & -\phi_2+\phi_1\phi_3 & -\phi_1+\phi_1\phi_2+\phi_2\phi_3 & 1+\phi_1^2+\phi_2^2+\phi_3^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+\phi_1^2+\phi_2^2 & -\phi_1+\phi_1\phi_2 & -\phi_2 \\ 0 & 0 & 0 & 0 & \cdots & -\phi_1+\phi_1\phi_2 & 1+\phi_1^2 & -\phi_1 \\ 0 & 0 & 0 & 0 & \cdots & -\phi_2 & -\phi_1 & 1 \end{bmatrix}$$

$$= \frac{\boldsymbol{\Sigma}^{-1}}{\sigma_a^2}$$

Note that $\mathbf{\Sigma}^{-1}$ is an $n \times n$ banded matrix with bandwidth $2p + 1$, where p is the order of the autoregressive process.

3 BAYESIAN INFERENCE

In order to make inferences about model parameters, we need to integrate over high-dimensional probability distributions. The MCMC methods are very helpful for solving our problems. The MCMC is essentially Monte Carlo integration using Markov chains. It draws samples from the required distribution by running a cleverly constructed Markov chain for a long time and then forms sample averages to approximate expectations. The Gibbs sampler is used in conjunction with the MH algorithm to make inferences and to make predictions. Let f denote the target density for notational convenience. In the simulation of $\boldsymbol{\gamma}$, $f(\boldsymbol{\gamma})$ is the posterior in Eq. (8). Details of the MH steps for $\boldsymbol{\gamma}$ are as follows:

Step 1. At iteration j, generate a point $\boldsymbol{\gamma}^*$ from the random walk kernel,

$$\boldsymbol{\gamma}^* = \boldsymbol{\gamma}^{(j-1)} + \varepsilon, \quad \varepsilon \sim N(0, a\,\Omega),$$

where $\boldsymbol{\gamma}^{(j-1)}$ is the $(j-1)$th iteration of $\boldsymbol{\gamma}$.

Step 2. Accept $\boldsymbol{\gamma}^*$ as $\boldsymbol{\gamma}^{(j)}$ with probability $p = \min\{1, f(\boldsymbol{\gamma}^*)/f(\boldsymbol{\gamma}^{(j-1)})\}$. Otherwise, set $\boldsymbol{\gamma}^{(j)} = \boldsymbol{\gamma}^{(j-1)}$.

To yield good convergence properties, the choices of Ω and a are important and can be found in So and Chen (2003). In summary, we use the following iterative sampling scheme to construct the desired posterior sample:

1. Draw $\boldsymbol{\beta}$ from the multivariate normal distribution in Eq. (6).
2. Draw σ_a^2 from the inverse Gamma distribution in Eq. (7).
3. Draw $\boldsymbol{\gamma}$ from the posterior in Eq. (8) using the MH algorithm.

We make an inverse transformation of $\boldsymbol{\phi}$ after we draw $\boldsymbol{\gamma}$. This completes one iteration.

4 ILLUSTRATIVE EXAMPLES

In this section, we illustrate the proposed methodology with a simulation study and three real data sets. We analyze the simulated data to calibrate the results against the known situation. The convergence of the Gibbs samplers are monitored by examining a procedure developed in Raftery and Lewis (1992).

4.1 Simulation Study

We now apply our proposed methodology to three examples in this simulation study. For comparisons, the first two examples follow the specification given in Chib (1993), and the third one adopts the specification used by C&G. Following Chib (1993), the regression for the first two examples is defined through

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t = 1, \ldots, n,$$

where $\beta_0 = 0$ and $\beta_1 = 1$. As mentioned by Chib (1993), the covariate X_t is generated according to the autoregression

$$X_t \sim N(0, 1/0.36),$$
$$X_t = 0.8X_{t-1} + V_t,$$
$$V_t \sim N(0, 1), \quad t \geq 2.$$

In this experiment, two sample sizes, $n = 100$ and $n = 500$, with 500 replications are used for demonstration. We generated the AR(1) and AR(2) processes for ε_t.

Example 1 Regression model with AR(1) errors, ε_t follows the process

$$\varepsilon_t - \phi_1 \varepsilon_{t-1} = a_t, \quad a_t \sim \text{iid } N(0, \sigma_a^2).$$

The values of (ϕ_1, σ_a^2) are (0.2, 0.96) and (0.6, 0.64) generated for each case.

Example 2 Regression model with AR(2) errors, ε_t is given by a second-order stationary autoregressive process

$$\varepsilon_t - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} = a_t, \quad a_t \sim \text{iid } N(0, \sigma_a^2),$$

where the parameters $(\phi_1, \phi_2, \sigma_a^2)$ are set at (0.2, 0.5, 1.59) and (0.7, 0.2, 4.44) to ensure stationarity and a variance of unity in each of the models.

TABLE I Summary Statistics for Regression Model with AR(1) Errors Obtained from 500 Replications.

	Real values	*Mean*	*(Std. Dev.)*	*Median*	*(Std. Dev.)*	*Chib*	*(Std. Dev.)*
$n = 100$							
ϕ_1	0.6	0.5953	(0.0776)	0.5949	(0.0763)	0.5431	(0.105)
β_0	0.0	0.0589	(0.1826)	0.0573	(0.1873)	−0.1035	(0.2013)
β_1	1.0	1.0032	(0.0853)	1.0044	(0.0873)	0.9781	(0.0844)
σ_a^2	0.64	0.6368	(0.0848)	0.6338	(0.0854)	0.6782	(0.1103)
$n = 500$							
ϕ_1	0.6	0.6005	(0.0534)	0.5998	(0.0537)	0.5512	(0.083)
β_0	0.0	0.0516	(0.0892)	0.0526	(0.0873)	−0.0935	(0.1711)
β_1	1.0	0.9753	(0.0347)	0.9753	(0.0345)	0.9832	(0.0612)
σ_a^2	0.64	0.6793	(0.0443)	0.6777	(0.00461)	0.6693	(0.0911)
$n = 100$							
ϕ_1	0.2	0.2027	(0.0911)	0.2047	(0.0934)	0.1421	(0.1091)
β_0	0.0	−0.0084	(0.1243)	−0.0118	(0.1222)	−0.0951	(0.1134)
β_1	1.0	1.0708	(0.074)	1.0695	(0.0812)	0.9841	(0.0911)
σ_a^2	0.96	0.9377	(0.1338)	0.9246	(0.1342)	1.0311	(0.1633)
$n = 500$							
ϕ_1	0.2	0.2003	(0.0636)	0.2004	(0.0671)	0.1423	(0.879)
β_0	0.0	−0.0676	(0.0565)	−0.0693	(0.0553)	−0.0891	(0.0831)
β_1	1.0	1.0472	(0.0312)	1.0452	(0.0331)	0.9877	(0.061)
σ_a^2	0.96	0.9652	(0.0636)	0.9628	(0.00612)	0.9835	(0.1351)

We choose $\boldsymbol{\beta_a} = 0.0$, $\mathbf{A_0} = 10^{-1}\mathbf{I}$, $\nu_0 = 3$, $\delta_0 = \tilde{\sigma}_a^2$, $\boldsymbol{\phi}_0 = 0.0$, and $\boldsymbol{\Phi}_0 = 10^{-1}\mathbf{I}$, where $\tilde{\sigma}_a^2$ is the residual mean squared error of fitting a pure regression model to the data. We carry out 4000 MCMC iterations and discard the first 2000 burn-in iterates for each series. We record every second value in the sequence of the last 2000 iterations in order to have more clearly independent contributions. The simulations have been redone based on the approaches of Chib (1993) and C&G with the same prior input for the parameters and with the same stationarity assumption. The results are shown in Tables I and II. For each data set, we obtain posterior means and posterior medians. Columns 3 and 4 contain the means and the standard errors of 500 posterior means while columns 5 and 6 contain the corresponding values of 500 posterior medians. The posterior means and standard errors of Chib (1993) are reproduced in columns 7 and 8. Histograms of posterior medians are given in Figures 1–4. The means of the estimators are close to the respective true values, indicating that the posterior mean obtained by our sampling scheme is a reliable estimator. We observe that the bias of posterior means slightly decreases when the sample size increases from 100 to 500. Moreover, the standard errors diminish substantially with the larger sample size.

Example 3 Following C&G, we simulate data from a regression model with AR(3) errors. The construction is described as follows:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t = 1, \ldots, n,$$
$$X_t \sim N(0, 1),$$

TABLE II Summary Statistics for Regression Model with AR(2) Errors Obtained from 500 Replications.

	Real values	*Mean*	*(Std. Dev.)*	*Median*	*(Std. Dev.)*	*Chib*	*(Std. Dev.)*
$n = 100$							
ϕ_1	0.2	0.1897	(0.0917)	0.1901	(0.912)	0.1341	(0.0921)
ϕ_2	0.5	0.4946	(0.0873)	0.4946	(0.0843)	0.5512	(0.093)
β_0	0.0	−0.143	(0.2534)	−0.1494	(0.2571)	−0.3122	(0.4231)
β_1	1.0	1.0882	(0.1122)	1.0864	(0.1219)	0.941	(0.1141)
σ_a^2	1.59	1.6081	(0.2405)	1.5749	(0.2425)	1.3411	(0.2031)
$n = 500$							
ϕ_1	0.2	0.2001	(0.0714)	0.2011	(0.0712)	0.1413	(0.0734)
ϕ_2	0.5	0.5006	(0.0658)	0.4993	(0.0643)	0.5443	(0.081)
β_0	0.0	−0.1092	(0.1783)	−0.1179	(0.1771)	−0.2956	(0.3621)
β_1	1.0	0.9892	(0.0481)	0.9894	(0.0479)	0.938	(0.0943)
σ_a^2	1.59	1.6189	(0.1111)	1.6167	(0.1135)	1.3741	(0.1816)
$n = 100$							
ϕ_1	0.7	0.6826	(0.0687)	0.6816	(0.0682)	0.6421	(0.097)
ϕ_2	0.2	0.1933	(0.0662)	0.1925	(0.0623)	0.1639	(0.098)
β_0	0.0	−0.0039	(0.3142)	−0.0057	(0.311)	−0.3433	(0.4111)
β_1	1.0	1.0827	(0.1676)	1.0898	(0.1561)	1.3313	(0.2123)
σ_a^2	4.44	4.2853	(0.6310)	4.252	(0.6292)	4.831	(0.6411)
$n = 500$							
ϕ_1	0.7	0.6867	(0.0698)	0.6902	(0.0683)	0.6333	(0.074)
ϕ_2	0.2	0.1916	(0.0613)	0.1902	(0.0623)	0.1691	(0.075)
β_0	0.0	0.3011	(0.3724)	0.3115	(0.371)	−0.3114	(0.3312)
β_1	1.0	1.1214	(0.0878)	1.1235	(0.0831)	1.2133	(0.1795)
σ_a^2	4.44	4.3838	(0.3604)	4.3387	(0.361)	4.613	(0.5314)

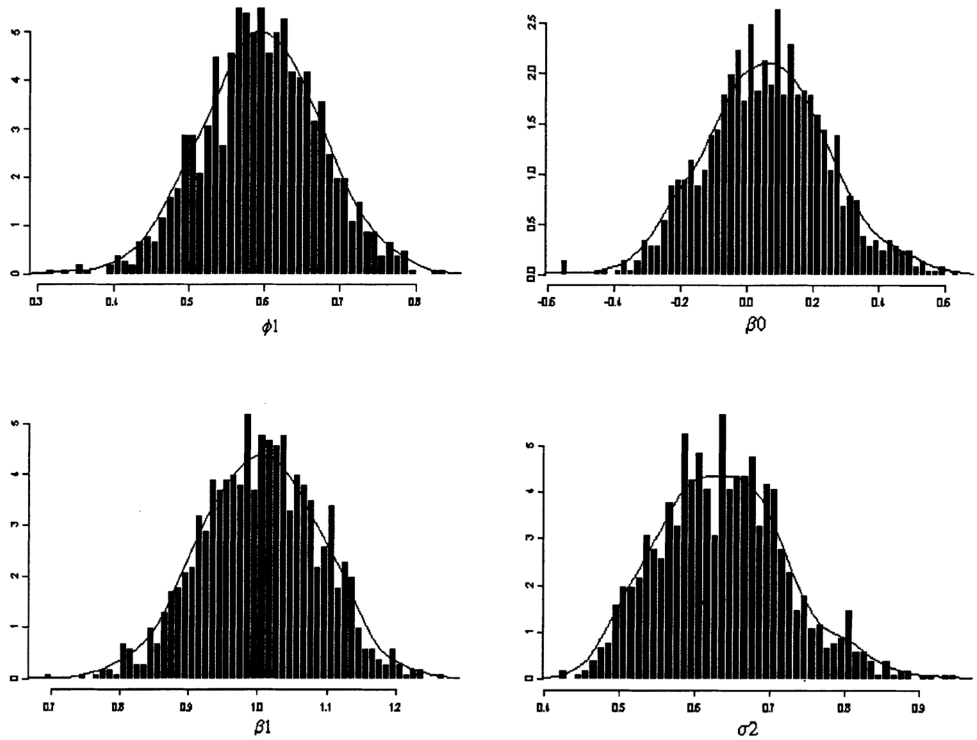

FIGURE 1 Simulation results for the regression model with AR(1) errors for $n = 100$ when true values are $(\phi_1, \beta_0, \beta_1, \sigma^2) = (0.6, 0.0, 1.0, 0.64)$.

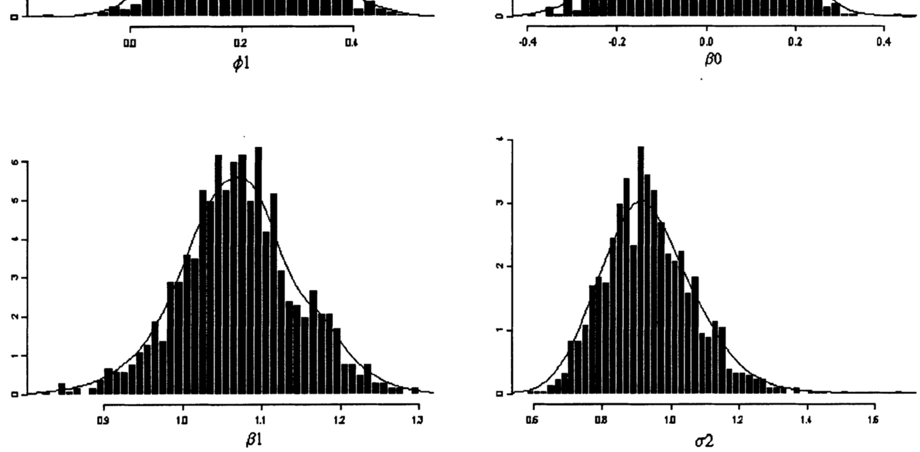

FIGURE 2 Simulation results for the regression model with AR(1) errors for $n = 100$ when true values are $(\phi_1, \beta_0, \beta_1, \sigma^2) = (0.2, 0.0, 1.0, 0.96)$.

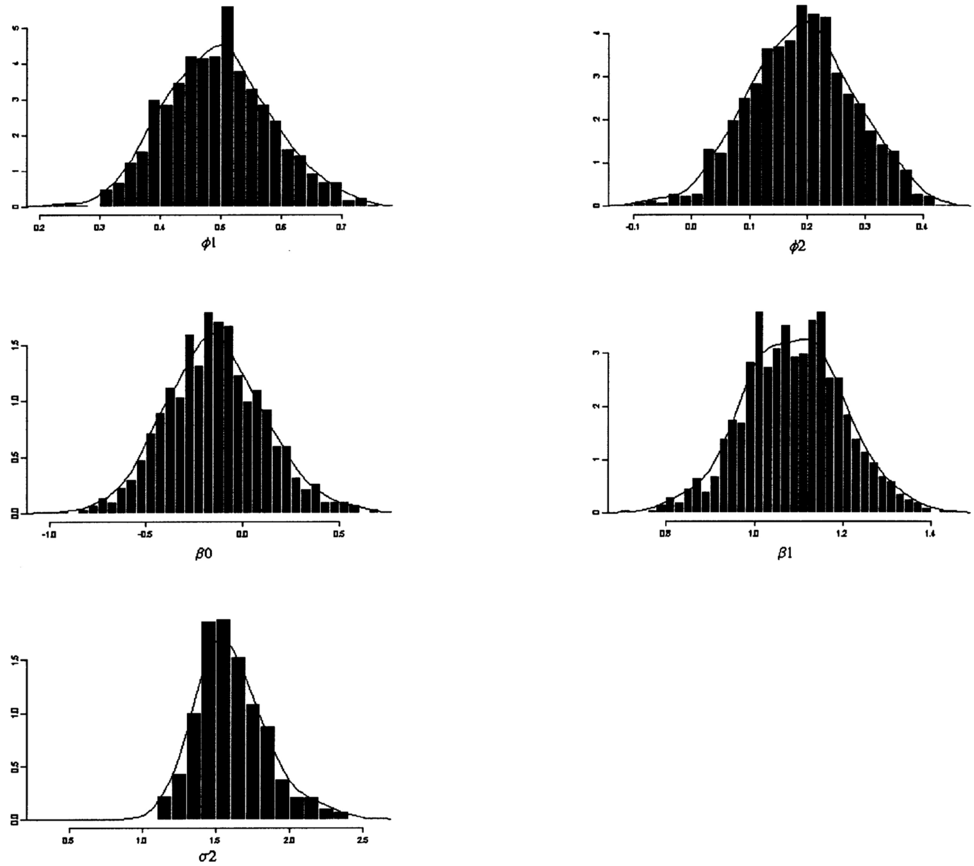

FIGURE 3 Simulation results for the regression model with AR(2) errors for $n = 100$ when true values are $(\phi_1, \phi_2, \beta_0, \beta_1, \sigma^2) = (0.2, 0.5, 0.0, 1.0, 1.59)$.

$$X_t = 0.8X_{t-1} + V_t,$$
$$V_t \sim N(0, 8), \quad t \geq 2,$$
$$\varepsilon_t - \phi_1\varepsilon_{t-1} - \phi_2\varepsilon_{t-2} - \phi_3\varepsilon_{t-3} = a_t, \quad a_t \sim N(0, \sigma_a^2).$$

The true values are

$$\boldsymbol{\beta} = (1, 1)^{\mathrm{T}}, \quad \boldsymbol{\phi} = (1.2, -0.2, -0.2)^{\mathrm{T}}, \quad \sigma_a^2 = 1.$$

Results of C&G are used to make a comparison with the proposed approach. Table III presents the summary statistics of comparative results from our approach and the approach of C&G. Histograms for the posterior medians of each parameter are given in Figure 5. We can see the estimates obtained by our sampling scheme are more accurate than those of C&G generally. Note that the inversion of the autocovariance matrix for an AR(p) model with sample size n in our procedure is an $n \times n$ banded matrix with bandwidth $2p + 1$. Thus, the computation burden is intolerable.

4.2 Applications

To illustrate the proposed procedure, three real data sets are considered in this subsection. In what follows, the estimation is done using S-PLUS 2000.

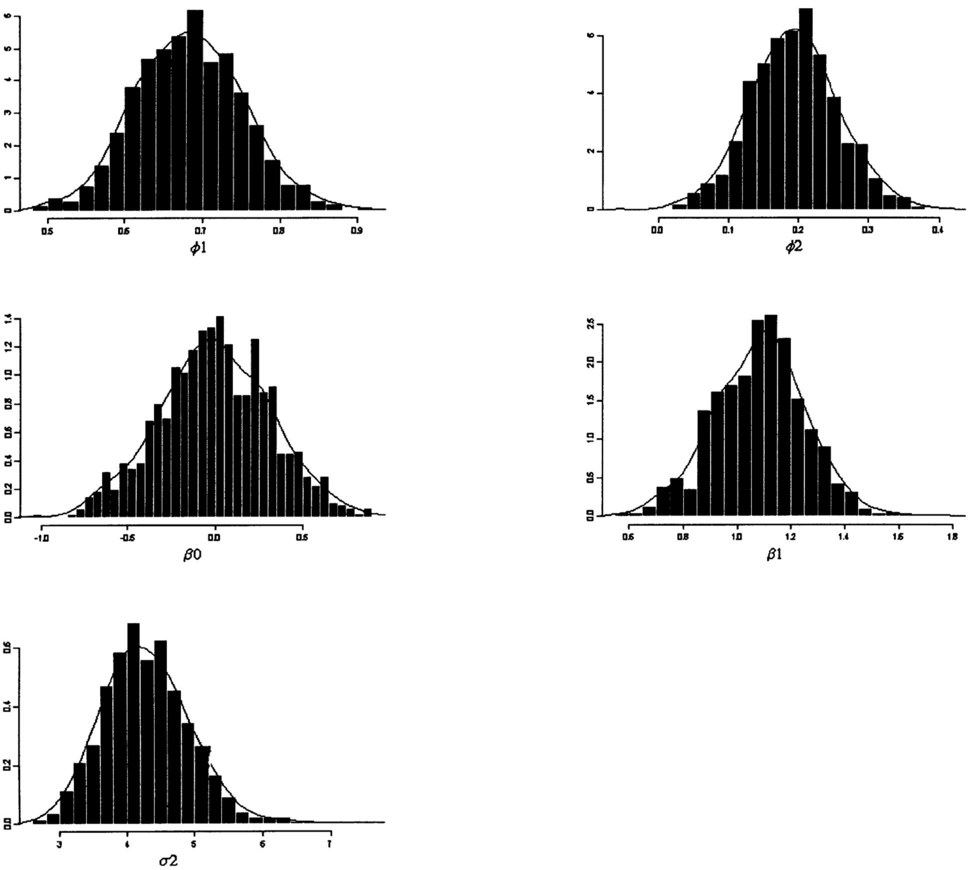

FIGURE 4 Simulation results for the regression model with AR(2) errors for $n = 100$ when true values are $(\phi_1, \phi_2, \beta_0, \beta_1, \sigma^2) = (0.7, 0.2, 0.0, 1.0, 4.44)$

TABLE III Summary Statistics for Regression Model with AR(2) Errors Obtained from 500 Replications.

	True value	*Mean*	*(Std. Dev.)*	*Median*	*Lower 95% limit*	*Upper 95% limit*	*C&G*	*(Std. Dev.)*
$n = 100$								
ϕ_1	1.2	1.1992	(0.0693)	1.2015	1.0579	1.3294	1.3521	(0.1131)
ϕ_2	−0.2	−0.1894	(0.1230)	−0.1872	−0.4286	0.052	−0.5211	(0.1726)
ϕ_3	−0.2	−0.1994	(0.1036)	−0.2028	−0.4084	−0.0051	−0.0721	(0.1181)
β_0	1.0	0.8228	(0.2994)	0.7949	0.3162	1.4349	1.5218	(0.3107)
β_1	1.0	0.9923	(0.0311)	0.9921	0.9319	1.0565	1.195	(0.0911)
σ_a^2	1.0	1.002	(0.2105)	0.9669	0.7085	1.5419	0.984	(0.144)
$n = 500$								
ϕ_1	1.2	1.2018	(0.0702)	1.1978	1.0575	1.3409	1.343	(0.093)
ϕ_2	−0.2	−0.2026	(0.1249)	−0.2011	−0.4427	0.0273	−0.477	(0.136)
ϕ_3	−0.2	−0.2019	(0.1011)	−0.2027	−0.4016	−0.004	−0.091	(0.098)
β_0	1.0	1.1916	(0.2290)	1.2151	0.6634	1.5781	1.498	(0.217)
β_1	1.0	1.0185	(0.0166)	1.0187	0.9844	1.0497	1.044	(0.0542)
σ_a^2	1.0	1.0774	(0.1908)	1.0264	0.8594	1.589	0.977	(0.1031)

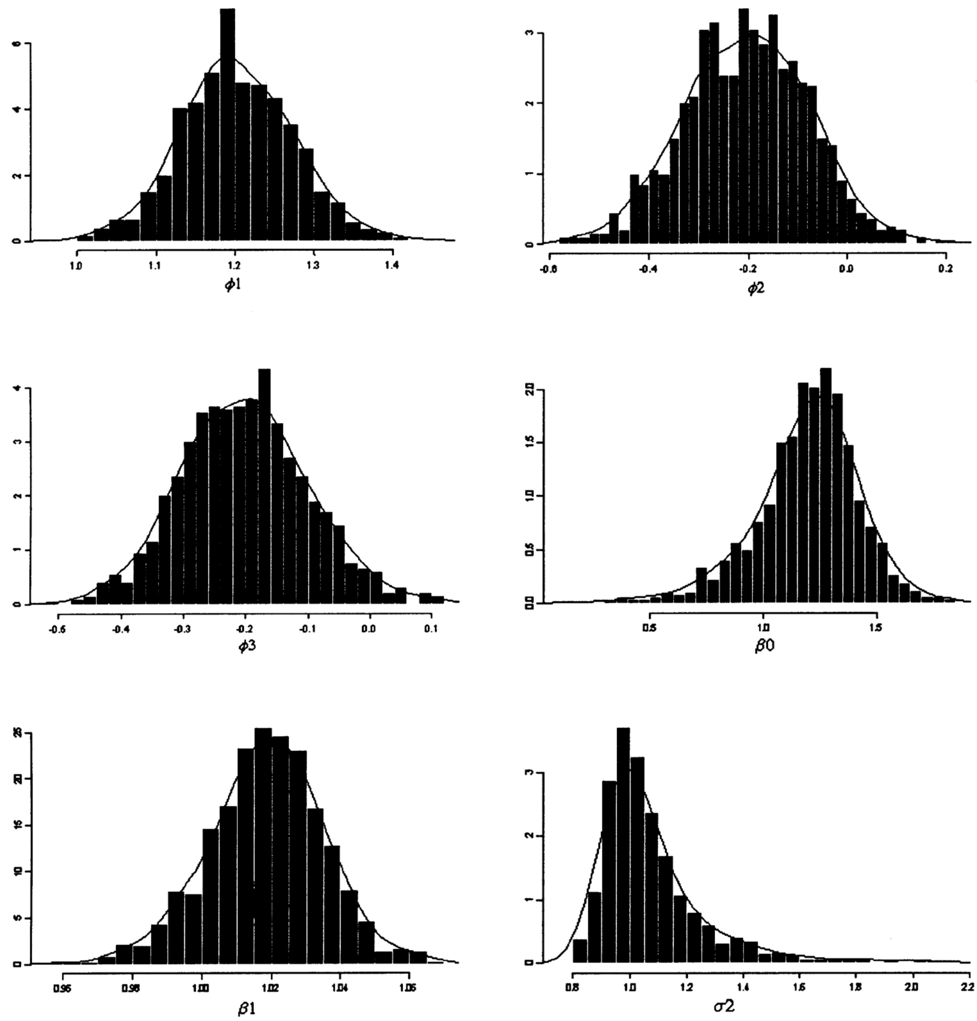

FIGURE 5 Simulation results for the regression model with AR(3) errors for $n = 100$ when true values are $(\phi_1, \phi_2, \phi_3, \beta_0, \beta_1, \sigma^2) = (0.7, 0.2, 0.0, 1.0, 4.44)$.

Example 4 Consider ice cream consumption measured over 30 four-week periods from March 18 1951, to July 11 1953. The data can be obtained from the Data and Story Library (DASL) at www.stat.cmu .edu/www/cmu-stats/DASL (datafile name 'ice cream'). The variables that underlie the time-series regression model are as follows: IC and P denote ice cream

TABLE IV Estimation for Ice Cream Consumption Data.

Parameter	*Mean*	*(Std. Dev.)*	*Median*	*Lower 95% limit*	*Upper 95% limit*
ϕ_1	0.6799	(0.0973)	0.6824	0.4828	0.8593
β_1	0.0754	(0.2621)	0.0648	−0.4313	0.625
β_2	0.0022	(0.0009)	0.0022	0.0005	0.0039
β_3	0.0034	(0.0006)	0.0034	0.0021	0.0043
σ_a^2	1.11E−3	(0.0003)	1.11E−3	0.0006	0.0018

TABLE V Parameter Estimates for Retail Sales Data.

Parameter	*Mean*	*(Std. Dev.)*	*Median*	*Lower 95% limit*	*Upper 95% limit*
ϕ_1	0.9829	(0.0031)	0.9828	0.977	0.9891
ϕ_2	−0.0134	(0.0097)	−0.0135	−0.0327	0.0038
β_1	0.0196	(0.0014)	0.0195	0.0172	0.0223
β_2	−0.0088	(0.0022)	−0.0089	−0.0135	−0.0048
β_3	0.0026	(0.0009)	0.0027	0.0008	0.0044
σ_a^2	0.1191	(0.0357)	0.1137	0.0699	0.1882

consumption in pints per capita and the price of ice cream per pint in dollars, respectively, and I is weekly family income in dollars, and Temp is mean temperature in degrees F. The model is specified as

$$\text{IC}_t = \beta_1 P_t + \beta_2 I_t + \beta_3 \text{Temp}_t + \varepsilon_t,$$
$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + a_t,$$

where $|\phi_1| < 1$ and $a_t \sim N(0, \sigma_a^2)$. The results are summarized in Table IV. We find a significant autoregressive dynamic in error terms. $\hat{\phi}_1 = 0.6799$ which indicates positive autocorrection in the error terms. To check model adequacy, we examine the residual ACF and PACF, which are small and exhibit no patterns.

Example 5 In this example, the data are also obtained from the DASL (datafile name 'predicting retail sales'). The datafile contains quarterly sales for four kinds of retail establishments together with nonagricultural employment, wage and salary disbursements. The goal is to develop a model that predicts retail sales. Variable names for data from the first quarter of 1979 to the fourth quarter of 1989 are

WS = national income wage and salary disbursements($ billions),
EMP = employees on payrolls of nonagriculture establishments (thousands),
BLD = building material dealer sales($ millions),
AUTO = automotive dealer sales($ millions),
FURN = furniture and home furnishings dealer sales($ millions),
GMER = general merchandise dealer sales($ millions).

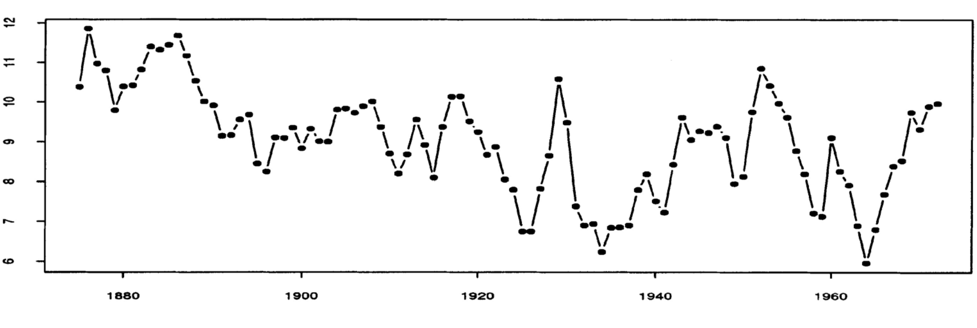

FIGURE 6 Level of Lake Huron in feet, reduced by 570 (1875–1972).

TABLE VI Regression Model with AR(2) Errors for Lake Huron Data.

Parameter	*Mean*	*(Std. Dev.)*	*Median*	*Lower 95% limit*	*Upper 95% limit*
ϕ_1	0.9850	(0.0140)	0.9848	0.9576	1.0132
ϕ_2	−0.2582	(0.0241)	−0.2575	−0.3054	−0.2109
β_0	10.0253	(0.2644)	10.0271	9.5122	10.5519
β_1	−0.0205	(0.0058)	−0.0204	−0.0319	−0.0091
σ_a^2	0.4654	(0.0668)	0.4596	0.3530	0.6179

After first fitting a full model, results showed that two explanatory variables, FURN and GMER, are insignificant. Therefore, the model is specified as

$$\begin{aligned} \mathrm{WS}_t &= \beta_1 \mathrm{EMP}_t + \beta_2 \mathrm{BLD}_t + \beta_3 \mathrm{AUTO}_t + \varepsilon_t, \\ \varepsilon_t &= \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + a_t. \end{aligned}$$

Parameter estimations are given in Table V. Note that the parameter ϕ_2 is marginally insignificant, its 95% confidence interval is (−0.0327, 0.0038). Moreover, the estimate of ϕ_1 is 0.98, which is close to 1. The goodness-of-fit for time-series regression models can be found in Chen and Wen (2001). The results of goodness-of-fit tests given in Chen and Wen (2001) for the two data sets used in Examples 4 and 5 indicate model adequacy.

Example 6 We consider series A in Brockwell and Davis (1991), which consists of lake levels in feet (reduced by 570) of Lake Huron for July of each year from 1875 through 1972. A time plot of the level of Lake Huron is shown in Figure 6. The level appears to fluctuate around an average level that decreases over time in a linear fashion. A time-series regression model with an explanatory variable t (time) is given as follows:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + \varepsilon_t, \\ \varepsilon_t &= \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + a_t. \end{aligned}$$

The results, with $t = 1$ for 1875, are given in Table VI. All parameters are significantly different from zero. The sign of $\hat{\beta}_1$ is negative which indicates that the water level of Lake Huron is descending as time progresses.

5 CONCLUDING REMARKS

We propose a Bayesian estimation procedure for a simple but most frequently used model in practice – namely, time-series regression models. Following Wise (1955), we use the exact likelihood instead of the likelihood conditional on initial observations. With the application of a useful reparametrization, the stationarity condition of $\boldsymbol{\phi}$ becomes $|\eta_i| < 1, i = 1, \ldots, p$. Consequently, the proposed procedure is valid for any order of the AR(p) process. Therefore, there is no difficulty in dealing with the AR(p) model for orders $p > 3$. We obtained better inferential results on simulations when compared with those of Chib (1993) and C&G. Results from real data sets show that the fitted models are adequate for the data sets. The proposed methodology can be extended in the following directions.

1. The Bayesian estimation procedure can be extended to regression with ARMA errors. The exact likelihood function of a general ARMA model can be found in Newbold (1974) and Hillmer and Tiao (1979).
2. To allow heteroscedasticity in the error variance, we can assume GARCH-type conditional variance, a model that has become very common in econometric and financial research.
3. When we deal with financial data, typical empirical evidence in the literature indicates that the distribution of errors is usually fat-tailed. In future work, we could consider leptokurtic distributions for the error terms, such as a student t-distribution or a generalized error distribution.

Acknowledgements

The authors thank Nan-Yu Wang for initial simulation study. C. W. S. Chen and Jack C. Lee acknowledge research support from National Science Council of Taiwan grants NSC91-2118-M-035-002 and NSC91-2118-M-009-002, respectively.

References

Albert, J. and Chib, S. (1993). Bayesian inference for autoregressive time series with mean and variance subject to Markov jumps. *J. Business Econ. Stat.*, **11**, 1–15.

Barndorff-Nielsen, O. E. and Schou, G. (1973). On the parametrization of autoregressive models by partial autocorrelations. *J. Multivar. Anal.*, **3**, 408–419.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, New York.

Chen, C. W. S. (1999). Subset selection of autoregressive time series models. *J. Forecast.*, **18**, 505–516.

Chen, C. W. S. and Wen, Y.-W. (2001). On goodness of fit for time series regression models. *J. Stat. Comput. Sim.*, **6**, 239–256.

Chib, S. (1993). Bayes regression with autoregressive errors: A Gibbs sampling approach. *J. Econometrics*, **58**, 275–294.

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

Chib, S. and Greenberg, E. (1995). Understanding the metropolis-Hastings algorithm. *Am. Stat.*, **49**, 327–335.

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

Hillmer, S. C. and Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *J. Am. Stat. Assoc.*, **74**, 652–660.

Marriott, J. M. and Smith, A. F. M. (1992). Reparametrization aspects of numerical Bayesian methods for autoregressive moving average models. *J. Time Ser. Anal.*, **13**, 327–343.

Marriott, J., Ravishanker, N., Gelfand, A. and Pai, J. (1996). Bayesian analysis for ARMA processes: Complete sampling-based inference under exact likelihoods. In: Berry, D. A., Chaloner, K. M. and Geweke, J. K. (Eds.), *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*. John Wiley & Sons, pp. 243–256.

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**, 1087–1091.

McCulloch, R. E. and Tsay, R. S. (1994). Bayesian analysis of autoregressive time series via the Gibbs sampler. *J. Time Ser. Anal.*, **15**, 235–250.

Newbold, P. (1974). The exact likelihood function for a mixed autoregressive moving average process. *Biometrika*, **61**, 423–426.

Raftery, A. E. and Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In: Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (Eds.), *Bayesian Statistics 4* Oxford University Press, Oxford, pp. 765–776.

So, M. K. P. and Chen, C. W. S. (2003). Subset threshold autoregression. *J. Forecast.*, **22**, 49–66.

Wise, J. (1955). The autocorrelation function and spectral density function. *Biometrika*, **42**, 151–159.