Journal of Econometrics 58 (1993) 275–294. North-Holland

Bayes regression with autoregressive errors*

A Gibbs sampling approach

Siddhartha Chib
Washington University, St. Louis, MO 63130, USA

Received June 1991, final version received December 1991

This paper develops a practical framework for the Bayesian analysis of Gaussian and Student-t regression models with autocorrelated errors. As is customary in classical estimation procedures, the posteriors are conditioned on the initial observations. Recourse is taken to the method of Gibbs sampling, an iterative Markovian sampling method, and it is shown that the proposed approach can readily deal with high-order autoregressive processes without requiring an importance sampling function or other tuning constants. Several examples, including one with AR(4) errors, are used to illustrate the ideas.

1. Introduction

The regression model with autocorrelated errors is one of the most heavily analyzed model in econometrics. Although frequentist ideas have dominated the literature, there is a quite substantial amount of Bayesian work in this area [cf. Zellner and Tiao (1964), Zellner (1971), Richard (1975), Fomby and Guilkey (1978), Judge et al. (1985, 1988), and Kennedy and Simons (1991)]. However, for the usual computational reasons, the prior–posterior analysis is difficult to implement with high-order autoregressions. Most Bayesian empirical illustrations have, therefore, been confined to AR(1) and AR(2) processes. On the other hand, sampling theory estimates, if conditioned on the initial observations, can be routinely calculated by nonlinear least squares, iterated Cochrane–Orcutt, or maximum likelihood [cf. Harvey (1990)].

In this paper we develop a practical Bayesian approach to Gaussian and Student-t versions of this model, conditioning on the initial observations, that is

Correspondence to: Siddhartha Chib, Olin School of Business, Washington University, St. Louis, MO 63130, USA.

*Comments made by the associate editor, two anonymous referees, and discussions with Ed Greenberg are gratefully acknowledged.

quite competitive with available classical methods. The approach rests on successive synthetic sampling from certain conditional distributions all of which are in standard form (multivariate normal, inverted gamma). An important feature of the approach is that it avoids the use of any importance sampling function, or other tuning constants, that would be required with standard Monte Carlo integration methods. Actually, the approach is fully automatic since it may be implemented without necessarily specifying any prior inputs about parameters, or any starting values for the iterations. Of course, the method offers the flexibility of incorporating prior information, for instance about the stationarity of the error process. Furthermore, the framework yields a complete posterior sample of the parameters from which the posterior density and functionals of the posterior (e.g., the prediction density) can be readily calculated.

The framework presented here is motivated by the Gibbs sampling scheme. Although we review the method in section 3 below, detailed discussions can be found elsewhere [Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), and Gelfand et al. (1990)]. The method is extremely useful for exploring intractable joint distributions that have a convenient conditional structure (in a sense that is made precise later). The Gibbs sampler has been widely used in recent years for Bayesian problems. [Cf. Carlin and Polson (1991), Albert and Chib (1991), McCulloch and Tsay (1991), Zeger and Karim (1991), and the references cited therein. The first use of the Gibbs algorithm in econometrics appears in Chib (1992) and Geweke (1992).] In particular, the McCulloch and Tsay paper, which deals with pure autoregressive models with outliers and structural breaks, is closely related.

The rest of the paper is organized as follows. Section 2 contains the model that is studied along with the inference problems that are discussed in the paper. An overview of the Gibbs sampler is provided in section 3. The development of the Gibbs sampling approach to AR(p) error models is taken up in section 4. In section 5 all the ideas are illustrated with several examples, one involving the AR(4) process. Concluding remarks are made in the final section.

2. The model

Consider the following model in which an observation y_t at time t, is generated by the regression with autocorrelated errors:

$$y_t = x_t'\beta + \varepsilon_t, \qquad \phi(L)\varepsilon_t = u_t, \tag{2.1}$$

where x_t: $k \times 1$ is a vector of covariates, $\beta \in \boldsymbol{R}^k$, $\phi(L)$ is a polynomial in the lag operator, L, given by $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, and the error u_t is white noise Gaussian with mean zero and variance $\sigma^2 > 0$, i.e., $N(0,\sigma^2)$. In the sequel, the Gaussian assumption will be relaxed in the direction of the Student-t family.

Assume for convenience that x_t does not contain any lagged values of y. Note that a special case of the model above is the autoregressive model $\phi(L)y_t = u_t$, which is obtained from (2.1) by excluding the covariates.

For future use note that if we multiply both sides in (2.1) by $\phi(L)$, then the distribution of y_t given H_{t-1} and θ, where H_{t-1} is the past history, is

$$y_t \mid H_{t-1}, \theta \sim \mathrm{N}(\hat{y}_{t|t-1,\theta}, \sigma^2), \tag{2.2}$$

where $\hat{y}_{t|t-1,\theta} = (1 - \phi(L))y_t + \phi(L)x_t'\beta$. By the law of total probability, the joint density of the observations, conditioned on the first p, is

$$\begin{aligned} f(y_{p+1},\ldots,y_n \mid y_p,\ldots,y_1,\theta) &= \prod_{t=p+1}^{n} f(y_t \mid H_{t-1},\theta) \\ &\propto \sigma^{-(n-p)} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=p+1}^{n} (y_t - \hat{y}_{t|t-1,\theta})^2 \right). \end{aligned} \tag{2.3}$$

Inference problems: The main problem of interest is to combine the data $y = (y_1,\ldots,y_n)$, with the prior information, if available, to obtain the posterior of the unknown parameters $\theta = (\beta, \sigma^2, \phi)$, where ϕ is a p vector of coefficients of the polynomial. By Bayes theorem, for a given prior pdf $\pi(\theta)$, the posterior of interest is given by

$$\pi(\theta \mid y) \propto \pi(\theta) f(y_{p+1},\ldots,y_n \mid y_p,\ldots,y_1,\theta), \tag{2.4}$$

with the normalizing constant given by $K = \int \pi(\theta) f(y_{p+1},\ldots,y_n \mid y_p,\ldots,y_1,\theta) \times \mathrm{d}\theta$.[1] Other questions of interest include making inferences about functionals of the parameters, for instance the prediction density of future y's. It should be noted that due to the nonlinearity arising from the parametrization of the model, these posterior computations are rendered cumbersome if approached directly via Bayes theorem even with diffuse priors on the parameters. However, the model has a convenient conditional structure that lends itself to the method of Gibbs sampling. This method is described next.

3. The method of Gibbs sampling

The algorithm: The Gibbs sampler is an iterative Monte Carlo method designed to extract marginal distributions from intractable joint distributions.

[1] As is common in classical approaches for high-order processes, we condition on the initial observations. Under the assumption that $n - p$ is large, this simplification is not likely to have an adverse impact on the conclusions.

Consider the three variables θ_1, θ_2, θ_3, with joint distribution $[\theta_1, \theta_2, \theta_3]$, where the conditioning on y is suppressed. Suppose the complete conditional distributions, $[\theta_1 | \theta_2, \theta_3]$, $[\theta_2 | \theta_1, \theta_3]$, and $[\theta_3 | \theta_1, \theta_2]$, have a much simpler form and are easily sampled. Then the algorithm for obtaining a draw $(\theta_1, \theta_2, \theta_3)$ from $[\theta_1, \theta_2, \theta_3]$ proceeds as follows.

Step 1. Specify initial values, $\theta_1^{(0)}$, $\theta_2^{(0)}$, $\theta_3^{(0)}$, and set $i = 1$.

Step 2. Cycle through the full conditionals drawing

$$\begin{aligned} &\text{(a)} \quad \theta_1^{(i)} \quad \text{from} \quad [\theta_1 | \theta_2^{(i-1)}, \theta_3^{(i-1)}] , \\ &\text{(b)} \quad \theta_2^{(i)} \quad \text{from} \quad [\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)}] , \\ &\text{(c)} \quad \theta_3^{(i)} \quad \text{from} \quad [\theta_3 | \theta_1^{(i)}, \theta_2^{(i)}] . \end{aligned} \tag{3.1}$$

Step 3. Set $i = i + 1$, and go to step 2.

After M iterations of the above scheme, the sample $\theta_1^{(M)}, \theta_2^{(M)}, \theta_3^{(M)}$ is obtained. Under regularity conditions [cf. Tierney (1991)], as $M \to \infty$, the sampled values converge in distribution to the relevant marginal and joint distribution, i.e., $\theta_j^{(M)} \to \theta_j \sim [\theta_j]$ and $(\theta_1^{(M)}, \theta_2^{(M)}, \theta_3^{(M)}) \to (\theta_1, \theta_2, \theta_3) \sim [\theta_1, \theta_2, \theta_3]$.

For M large, so that the desired convergence in distribution has been attained, the N values $(\theta_1^{(j)}, \theta_2^{(j)}, \theta_3^{(j)})$, $j = M + 1, \ldots, M + N$, are a sample from $[\theta_1, \theta_2, \theta_3]$. Alternatively, it is possible to use multiple runs wherein the sampler is replicated with different starting values, and the Mth iterate from each stream is retained [cf. Gelfand and Smith (1990)]. In either case, posterior inference is straightforward since the entire posterior sample is available.

Posterior inference: The marginal density of θ_1 (say), is obtained as a finite mixture

$$[\hat{\theta}_1] = \frac{1}{N} \sum_{j=M+1}^{M+N} [\theta_1 | \theta_2^{(j)}, \theta_3^{(j)}] , \tag{3.2}$$

while the expectation of a function, $g(\theta)$, of the parameters is estimated via the sample average

$$\hat{\mathrm{E}}(g(\theta)) = \frac{1}{N} \sum_{j=M+1}^{M+N} g(\theta^{(j)}) . \tag{3.3}$$

Note that is possible to reduce the numerical variance of the estimator by conditioning. Suppose for any s the conditional expectation, $g_s(\theta_r, r \neq s) \equiv \mathrm{E}(g(\theta) | \theta_r, r \neq s)$, is available in closed form.

Then by a Rao–Blackwell argument, the desired estimator is

$$\hat{g}_s = N^{-1} \sum_{j=M+1}^{M+N} g_s(\theta_r^{(j)},\ r \neq s) \tag{3.4}$$

[cf. Gelfand and Smith (1990)]. Thus, to calculate $\mathrm{E}(\theta_1)$ and $\mathrm{V}(\theta_1)$, for example, one uses

$$\hat{\mathrm{E}}(\theta_1) = N^{-1} \sum_{j=M+1}^{M+N} g_1(\theta_2^{(j)}, \theta_3^{(j)}) \tag{3.5}$$

and

$$\hat{\mathrm{V}}(\theta_1) = N^{-1} \sum_{j=M+1}^{M+N} \tilde{g}_1(\theta_2^{(j)}, \theta_3^{(j)}) - \{\hat{\mathrm{E}}(\theta_1)\}^2 , \tag{3.6}$$

where $g_1 \equiv \mathrm{E}(\theta_1 \mid \theta_2, \theta_3)$ and $\tilde{g}_1 \equiv \mathrm{E}(\theta_1^2 \mid \theta_2, \theta_3)$.

Numerical standard errors: The estimates above are based on correlated draws, and therefore the numerical standard errors have to be computed by time series methods. One possibility is the spectral approach of Geweke (1992). Alternatively, one can use the well-known batch means method [cf. Ripley (1987)]. In batching the idea is to divide the data, for example, the summand in (3.5), into b batches of length B. If we let batch means be m_i and the average of the batches be $\bar{m}$, then the standard error of the estimate is given by $\{b(b-1)\}^{-1} \sum_{i=1}^{b} (m_i - \bar{m})^2$. We illustrate this method in section 5.

4. Conditional distributions

We now show that the model under study has a convenient conditional structure. Each of the complete conditional distributions, i.e., $[\beta \mid y, \sigma^2, \phi]$, $[\sigma^2 \mid y, \beta, \phi]$, and $[\phi \mid y, \beta, \sigma^2]$, have a simple form and are easily simulated, thus providing via the Gibbs sampler a complete sample from the joint posterior of the parameters. Suppose that the prior distribution of (β, σ^2, ϕ) is given by

$$\pi(\beta, \sigma^2, \phi) = \pi(\beta \mid \sigma^2)\, \pi(\sigma^2)\, \pi(\phi) , \tag{4.1}$$

which asserts that (β, σ^2) is *a priori* independent of ϕ. In addition, let

$$\beta \mid \sigma^2 \sim \mathrm{N}_k(\beta_0, \sigma^2 A_0^{-1}) , \quad \sigma^2 \sim \mathrm{IG}\left(\frac{\nu_0}{2}, \frac{\delta_0}{2}\right) , \quad \phi \propto \mathrm{N}_p(\phi_0, \Phi_0^{-1}) I_{\phi \in S_\phi} , \tag{4.2}$$

a combination of multivariate normal and inverted gamma distributions for (β, σ^2), and a multivariate normal distribution for ϕ truncated to the region, S_ϕ, that implies a stationary error process; I_{S_ϕ} denotes the indicator function of the set S_ϕ. The indicator function can be dropped if this restriction is not being imposed. Diffuse, automatic priors over (β, σ^2, ϕ) are obtained if the prior hyperparameters in (4.2) are set at $\beta_0 = 0$, $A_0 = $, $v_0 = -k$, $\delta_0 = 0$, $\phi_0 = 0$, and $\Phi_0 = 0$.[2]

$[\boldsymbol{\beta} | \boldsymbol{y}, \boldsymbol{\phi}, \boldsymbol{\sigma^2}]$: Since we condition on (σ^2, ϕ), perform a change of variable in (2.4) from y_t to $y_t^* = \phi(L) y_t$. It is easy to confirm the fact that $y_t^* | H_{t-1} \sim \text{IN}(x_t^* \beta, \sigma^2)$, where $x_t^* = \phi(L) x_t$, $t = p+1, \ldots, n$, and IN denotes independent normal. Therefore due to the independence of the $\{y_t^*\}$, the model in terms of the transformed data is given by

$$y^* = X^* \beta + u, \qquad u \sim \text{N}(0, \sigma^2 I_{n-p}), \tag{4.3}$$

where $y^* = (y_t^*)$ and $X^* = (x_t^*)$ and $(n-p) \times 1$ and $(n-p) \times k$, respectively.[3] Note that the rank of X^* will be k if for all covariates $j, j = 1, \ldots, k$, and some t, $x_{jt} - \phi_1 x_{jt-1} - \cdots - \phi_p x_{jt-p} \neq 0$ [cf. Zellner (1971) for the AR(1) case].

Combining the normal prior of β in (4.2) with the likelihood of β defined by (4.3), we obtain using standard results that

$$\beta | y, \sigma^2, \phi \sim \text{N}_k(\tilde{\beta}, \sigma^2 \tilde{A}^{-1}), \tag{4.4}$$

where $\tilde{\beta} = (A_0 + X^{*\prime} X^*)^{-1} (A_0 \beta_0 + X^* y^*)$ and $\tilde{A} = (A_0 + X^{*\prime} X^*)$. Simulation of β from (4.4) is straightforward. Note, however, that if stationarity is not imposed and the ϕ's satisfy the unit root condition $\sum \phi_i = 1$, then the constant β_0 is not identified since $X^{*\prime} X^*$ is of rank $k-1$, as noted by Zellner (1971, p. 93) in the AR(1) case. As long as the (1, 1) element of A^* is positive, one effectively samples from the prior at the point in the Gibbs sampler.[4]

[2] Other prior distributions can be easily used. For example, independence between β and σ^2 can be permitted, as pointed out by a referee. As a result, in (4.4) the covariance matrix of β would be given by $(A_0 + \sigma^{-2} X^{*\prime} X^*)$, the posterior mean by the product of this covariance and $(A_0 \beta_0 + \sigma^{-2} X^{*\prime} y^*)$, and (4.5) would not be required.

[3] For example, if we have the AR(2) process, then the tth element of y^* is $y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2}$ and the tth row of X^* is $x_t' - \phi_1 x_{t-1}' - \phi_2 x_{t-2}'$. Note that the dependence of y^* and X^* on the ϕ's is not emphasized in the notation.

[4] If the Gibbs run reveals considerable probability mass for, say, the region $0.999 < \sum \phi_i < 1.001$, the model can be estimated without a constant. Of course, these issues do nor arise if ϕ is restricted to S_ϕ.

$[\sigma^2 \mid y, \beta, \phi]$: Given β and ϕ, the likelihood of σ^2 is immediately obtained from (4.3). Also from (4.1), a straightforward calculation shows that the prior $\sigma^2 \mid \beta$ is given by

$$\sigma^2 \mid \beta \sim \text{IG}\left(\frac{v_0 + k}{2}, \frac{\delta_0 + Q_\beta}{2}\right), \qquad Q_\beta = (\beta - \beta_0)' A_0 (\beta - \beta_0) . \tag{4.5}$$

Applying the usual Bayesian updating formulas to (4.3) and (4.5) gives the second conditional as

$$\sigma^2 \mid y, \beta, \phi \sim \text{IG}\left(\frac{n - p + v_0 + k}{2}, \frac{\delta_0 + Q_\beta + d_\beta}{2}\right), \tag{4.6}$$

where $d_\beta = (y^* - X^* \beta)'(y^* - X^* \beta)$.

$[\phi \mid y, \beta, \sigma^2]$: The derivation of the conditional posterior of ϕ is surprisingly not difficult even when ϕ is high-dimensional. The main point is that given y, β, σ^2, the error $\{\varepsilon_t\}$ in (2.1) becomes degenerate. Thus, letting $\varepsilon_t = y_t - x_t'\beta$ and suppressing the dependence on β, the model for the errors is given by

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} + u_t , \qquad t = p + 1, \ldots, n, \tag{4.7}$$

or in matrix notation as

$$\varepsilon = E\phi + u , \tag{4.8}$$

where E is a $(n - p) \times p$ matrix whose tth row is $(\varepsilon_{t-1}, \ldots, \varepsilon_{t-p})$. From (4.8), it is readily observed that the conditional likelihood of ϕ is multivariate normal. Therefore, combining the likelihood with the normal prior on ϕ leads to an updated truncated multivariate normal posterior given by

$$\phi \mid y, \beta, \sigma^2 \propto \text{N}_p(\tilde{\phi}, \tilde{\Phi}^{-1}) I_{S_\phi} , \tag{4.9}$$

where $\tilde{\phi} = \tilde{\Phi}^{-1}(\Phi_0 \phi_0 + \sigma^{-2} E'\varepsilon)$ and $\tilde{\Phi} = (\Phi_0 + \sigma^{-2} E'E)$. As compared to the unrestricted case, sampling from the posterior is only a bit more complicated. We can draw from the untruncated normal posterior, and retain the draw if the roots of $\phi(L)$ lie outside the unit circle. The proportion of draws thus accepted provide the conditional probability that the process is stationary. The main virtue of this simulation scheme is that it can be used without regard to the

dimension of ϕ, and it will be quite efficient if the conditional posterior is concentrated over S_ϕ.[5]

With the complete conditionals available for sampling, the algorithm described in (3.1) may now be run by cycling through (4.9), (4.4), and (4.6), generating ϕ, β, and σ^2 in that order. The algorithm may be initialized by using the first-round $[\phi(L) \equiv 1]$ least squares estimates of β and σ^2; note that this choice is purely automatic and requires no user input. Based on the resulting posterior sample, $(\beta^{(j)}, \sigma^{2(j)}, \phi^{(j)})$, $j = M+1, \ldots, M+N$, for some suitable values of M and N, the posterior mean and standard deviation of β and σ^2 may be obtained via (3.5) and (3.6) since the conditional mean and variance of each is available in closed form. The posterior moments of ϕ can be obtained directly from the posterior sample on ϕ as in (3.3).

It is possible to provide an interesting connection with the iterative Cochrane-Orcutt procedure if we consider the conditional posteriors that emerge under a diffuse prior. Then the Cochrane–Orcutt method may be viewed as iterating on the computation of the mode of (4.9), and then of the mode of (4.4), both conditioned on estimates from the previous round.

Prediction density: Based on the posterior sample it is straightforward to derive the prediction density, say, for the first out of sample observation. Since $[y_{n+1} \mid y] \equiv \int [y_{n+1} \mid y, \theta]\, d[\theta \mid y]$, it follows that the prediction density can be obtained as

$$N^{-1} \sum_{i=1}^{N} f_n(y_{n+1} \mid \hat{y}_{n+1 \mid n, \theta^{(i)}}, \sigma^{2(i)}),$$

where f_n is the density of the normal distribution given in (2.2). Prediction densities for other lead times are obtained in the same way by averaging the relevant conditional densities.

Student-errors: In passing we mention that the assumption of Gaussian errors can be relaxed in the direction of the Student-t family. Let $\{\lambda_t\} \sim$ Gamma $(\nu/2, \nu/2)$ and assume

$$\varepsilon_t \mid H_{t-1}, \sigma^2, \lambda_t \sim \mathrm{N}((1 - \phi(L))\varepsilon_t, \lambda_t^{-1}\sigma^2).$$

[5] In the one-dimensional case, $S_\phi = (-1, 1)$, and a one-for-one draw from (4.9) may be generated via the result in Devroye (1986).

On integrating w.r.t. λ_t, it can be confirmed that $[\varepsilon_t | H_{t-1}, \sigma^2] = T_\nu((1-\phi(L))\varepsilon_t, \sigma^2)$, a t distribution with ν degrees of freedom. Let Λ be a $n-p$ diagonal matrix with entries given by $(\lambda_{p+1}, \ldots, \lambda_n)$. Then, under the prior in (4.1) we have that the complete conditionals are now given by

$$\beta | y, \sigma^2, \phi, \{\lambda_t\} \sim \mathrm{N}_k(\tilde{\beta}_\lambda, \sigma^2 \tilde{A}_\lambda^{-1}),$$

$$\sigma^2 | y, \beta, \phi, \{\lambda_t\} \sim \mathrm{IG}\left(\frac{n-p+\nu_0+k}{2}, \frac{\delta_0 + Q_\beta + d_{\beta\lambda}}{2}\right),$$

$$\phi | y, \beta, \sigma^2, \{\lambda_t\} \propto \mathrm{N}_p(\tilde{\phi}_\lambda, \tilde{\Phi}^{-1}) I_{S_\phi},$$

$$\lambda_t | y, \beta, \sigma^2, \phi \sim \mathrm{Gamma}\left(\frac{\nu+1}{2}, \frac{\nu+\delta_t}{2}\right), \qquad t = p+1, \ldots, n,$$

where $\tilde{\beta}_\lambda = \tilde{A}_\lambda^{-1}(A_0\beta_0 + X^{*\prime}\Lambda y^*)$, $\tilde{A}_\lambda = (X^{*\prime}\Lambda X^* + A_0)$, $\tilde{\Phi}_\lambda = (\Phi_0 + \sigma^{-2}E'\Lambda E)$, $d_{\beta\lambda} = (y^* - X^*\beta)'\Lambda(y^* - X^*\beta)$, $\tilde{\phi}_\lambda = \tilde{\Phi}_\lambda^{-1}(\Phi_0\phi_0 + \sigma^{-2}E'\Lambda\varepsilon)$, and $\delta_t = (y_t^* - x_t^{*\prime}\beta)^2/\sigma^2$. Calculations are straightforward despite the $n-p$ new random variables (illustrative examples are available from the author). See Carlin and Polson (1991) and Albert and Chib (1991) for use of the Student-t distribution in other models.

5. Examples

We now apply the proposed method to three examples involving different orders of autocorrelation and sample sizes. Our objective is to illustrate how the methods actually perform in realistic settings that were previously considered intractable from a Bayesian perspective. The analysis is conducted with vague priors with and without the stationarity restriction. For comparison purposes, estimates by NLS which are computed via Micro-TSP are provided along with the Bayes estimates. The programs for the Gibbs sampler are coded in GAUSS and implemented on a 80386-33Mhz computer. Before presenting the results, we summarize the specifications used in the examples.

The first two examples that we consider deal with simulated data in which the true data generating process (DGP) is known. The regression DGP is defined through

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \qquad t = 1, \ldots, n,$$

where $\beta_1 = 0$ and $\beta_2 = 1$. As in Keener, Kmenta, and Weber (1991), the covariate x_t is generated according to the autoregression

$$x_1 \sim \mathrm{N}(0, 1/0.36)\,, \quad x_t = 0.8x_{t-1} + v_t\,, \quad v_t \sim \mathrm{N}(0, 1), \quad t \geq 2.$$

For sample sizes of $n = 50$, 100, and 150, we consider the following AR(1) and AR(2) DGP for ε_t.

Example 1: AR(1) models. In this case ε_t follows the process

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + u_t\,, \qquad u_t \sim \text{iid } \mathrm{N}(0, \sigma^2)\,.$$

The values of (ϕ_1, σ^2) are (0.2, 0.96) and (0.6, 0.64) which implies that $\mathrm{var}(\varepsilon_t) = 1$ for each case.

Example 2: AR(2) models. Now ε_t is given by a second-order stationary autoregression,

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + u_t, \qquad u_t \sim \text{iid } \mathrm{N}(0, \sigma^2)\,,$$

where the parameters $(\phi_1, \phi_2, \sigma^2)$ 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.

Example 3: Real data with AR(4) errors [cf. Ramanathan (1990)]. The problem is to explain the consumption of electricity by residential customers that are served by the San Diego Gas and Electric Company. The model has six covariates (including the constant). The dependent variable is measured as the logarithm of electricity consumption in kilowatt-hours (*KWH*) per residential consumer. The covariates in order are a constant (*CNST*), logarithm of per capita real income (*PCI*), logarithm of the real price of electricity (*PE*), logarithm of the real price of gas (*PG*), cooling degree days (*CDD*), and heating degrees days (*HDD*). The data set, which runs from the first quarter of 1970 to the first quarter of 1983, is reproduced in the appendix.

In the implementation of the Gibbs sampler, the first 50 draws are discarded and the algorithm is run to obtain 1200 draws from the posterior. Numerical standard errors are obtained by the method of batching. Following the recommendation of Ripley (1987), the batch size B is selected such that the first-order correlation between the batch means is at most 0.05. Because of the nature of the Markovian sampling, this usually results in small to negligible higher-order

correlations between the batch means. To provide an idea of the computing effort entailed, we mention that it takes 2 to 3 minutes (depending on the sample size) to estimate the models. Thus, computing time is not of much concern and is likely to be even less important on more powerful computers than the one used in the present case.

The results for examples 1 and 2 are presented in tables 1 and 2, where the posterior mean and standard deviation of the parameters is calculated based on (3.3)–(3.6), as described earlier. The hyperparameters in (4.2) are set at $\beta_0 = 0$, $A_0 = 10^{-6} I_k$, $\nu_0 = 0.02$, $\delta_0 = 2$, $\phi_0 = 0$, and $\Phi_0 = 10^{-6} I_p$. From the tables we can ascertain that the Bayes estimates are very close to the true values that are used to generate the data. Except in a few cases, the lag 1 correlation of the Gibbs run displays only moderate dependence. The batch standard errors indicate that the estimates are very reliable. In many cases, there is an improvement over the NLS estimates although, as is to be expected, the difference is not significant for the larger samples. Generally, the posterior standard deviation exceeds the standard error of the NLS estimate.

Next, consider tables 3 and 4 and figs. 1 and 2 which deal with two models based on example 3. Two priors, A and B, are used for the analysis. In both cases, Jeffreys' prior is specified for σ^2, while the other hyperparameters are as before. In addition, prior A does not impose the stationarity restriction on the roots of the error polynomial. The first point to be noted is that the results are resistant to the prior used, except for the constant which is insignificant under prior A. This is not surprising since the probability of the unit root is 0.11 with model 1 and 0.20 with model 2 [these are computed by the proportion of times $\sum \phi_i$ lies in the interval (0.999, 1.001)]. Next note that for some of the parameters there are significant differences between the NLS estimates and the Bayes posterior moments. For example, the posterior mean of *PCI* is about four times as large as the NLS; the variable *CDD* appears not be statistically significant; and the posterior mean of ϕ_2 is half that of NLS and has the opposite sign, and that of ϕ_3 is negative. On the basis of the posterior moments for model 1, we reestimate the model after dropping the variables *PG* and *CDD*. We observe from table 4 that the posterior moments of the parameters of model 2 are resistant to changes in the specification of the covariates. However, the NLS estimates change considerably and become similar to those from the Gibbs sampler. Finally, for model 2 with prior B the marginal posterior pdfs of all the parameters (excluding σ^2) are obtained (cf. figs. 1 and 2). The skewness in the posteriors of β_3, ϕ_1, and ϕ_3 is relatively pronounced. The pdfs are computed on a 65-point grid, applying (3.2) to (4.4) and (4.9); the domain of ϕ is restricted to the minimum and maximum values of ϕ observed in the 1200 draws. All the results in model 2, including the eight pdfs, are obtained in just four minutes.

Table 1
AR(1) models (cf. example 1).[a]

DGP		β_1 0	β_2 1	ϕ_1 0.2	σ^2 0.96
$n = 50$	NLS	−0.331 (0.191)	0.900 (0.112)	0.134 (0.150)	1.113 (---)
	Bayes	−0.313 (0.223) [0.001] {0.29}	0.926 (0.128) [0.001] {0.29}	0.190 (0.163) [0.005] {0.26}	1.157 (0.247) [0.002] {0.17}
$n = 100$	NLS	−0.104 (0.121)	0.992 (0.060)	0.111 (0.102)	1.050 (---)
	Bayes	−0.101 (0.129) [0.0003] {0.17}	0.996 (0.063) [0.0004] {0.17}	0.135 (0.106) [0.004] {0.16}	1.070 (0.156) [0.0008] {0.06}
$n = 150$	NLS	−0.026 (0.101)	0.971 (0.053)	0.172 (0.082)	0.960 (---)
	Bayes	−0.025 (0.105) [0.00007]{0.11}	0.973 (0.055) [0.0002] {0.11}	0.186 (0.085) [0.003] {0.12}	0.973 (0.115) [0.0005]{0.02}
DGP		β_1 0	β_2 1	ϕ_1 0.6	σ^2 0.64
$n = 50$	NLS	−0.038 (0.219)	1.313 (0.109)	0.387 (0.135)	0.778 (---)
	Bayes	−0.042 (0.295) [0.0004] {0.40}	1.314 (0.114) [0.00001] {0.17}	0.435 (0.143) [0.005] {0.23}	0.822 (0.176) [0.002] {0.22}
$n = 100$	NLS	−0.097 (0.165)	1.105 (0.066)	0.499 (0.088)	0.642 (---)
	Bayes	−0.095 (0.187) [0.0001] {0.19}	1.105 (0.068) [0.00001] {0.16}	0.522 (0.091) [0.003] {0.17}	0.662 (0.097) [0.0005] {0.10}
$n = 150$	NLS	−0.055 (0.160)	1.098 (0.058)	0.586 (0.068)	0.628 (---)
	Bayes	−0.051 (0.177) [0.0003] {0.14}	1.098 (0.059) [0.00001] {0.15}	0.600 (0.070) [0.002] {0.15}	0.641 (0.075) [0.0003]{0.06}

[a] NLS denotes the nonlinear least squares estimate. The Bayes estimates are based on $M = 50$ and $N = 1200$. Standard deviation (standard error for NLS) is in brackets; the numerical standard error of the posterior mean is in square brackets; the lag 1 correlation of the Gibbs run is in curly brackets.

Table 2

AR(2) models (cf. example 2).[a]

		β_1	β_2	ϕ_1	ϕ_2	σ^2
DGP		0	1	0.2	0.5	1.59
$n = 50$	NLS	0.226 (0.246)	0.872 (0.112)	− 0.028 (0.143)	0.381 (0.148)	1.080 (− − −)
	Bayes	0.103 (0.932) [0.019] [0.58}	0.883 (0.119) [0.0008] {0.39}	0.036 (0.154) [0.006] {0.25}	0.448 (0.153) [0.006] {0.22}	1.137 (0.249) [0.003] {0.33}
$n = 100$	NLS	− 0.374 (0.324)	0.955 (0.091)	0.106 (0.088)	0.530 (0.088)	1.295 (− − −)
	Bayes	− 0.433 (0.616) [0.008] {0.43}	0.955 (0.092) [0.0001] {0.03}	0.129 (0.091) [0.003] {0.13}	0.557 (0.090) [0.003] {0.10}	1.321 (0.195) [0.002] {0.24}
$n = 150$	NLS	− 0.470 (0.277)	0.937 (0.080)	0.162 (0.075)	0.473 (0.076)	1.460 (− − −)
	Bayes	− 0.510 (0.504) [0.008] {0.69}	0.936 (0.081) [0.0002] {0.04}	0.177 (0.077) [0.002] {0.11}	0.450 (0.077) [0.002] {0.07}	1.464 (0.174) [0.001] {0.24}
		β_1	β_2	ϕ_1	ϕ_2	σ^2
DGP		0	1	0.7	0.2	4.44
$n = 50$	NLS	2.651 (4.760)	1.786 (0.299)	0.641 (0.146)	2.286 (0.149)	4.850 (− − −)
	Bayes	1.885 (6.532) [0.059] {0.35}	1.791 (0.298) [0.002] {0.09}	0.645 (0.147) [0.004] {0.03}	0.282 (0.147) [0.004] {0.02}	4.835 (1.05) [0.008] {0.02}
$n = 100$	NLS	− 0.240 (2.62)	1.371 (0.208)	0.636 (0.100)	0.276 (0.100)	4.711 (− − −)
	Bayes	− 1.618 (6.532) [0.119] {0.51}	1.373 (0.206) [0.0003] {0.29}	0.646 (0.096) [0.003] {0.03}	0.284 (0.096) [0.003] {0.04}	4.723 (0.694) [0.004] {0.09}
$n = 150$	NLS	− 0.063 (1.41)	1.082 (0.162)	0.649 (0.083)	0.227 (0.082)	4.445 (− − −)
	Bayes	− 0.407 (3.24) [0.032]{0.49}	1.085 (0.161) [0.0003] {0.13}	0.660 (0.084) [0.002] {0.04}	0.239 (0.083) [0.002] {0.04}	4.470 (0.531) [0.003] {0.20}

[a] NLS denotes the nonlinear least squares estimate. The Bayes estimates are based on $M = 50$ and $N = 1200$. Standard deviation (standard error for NLS) is in brackets; the numerical standard error of the posterior mean is in square brackets; the lag 1 correlation of the Gibbs run is in curly brackets.

AR(4) models (cf. example 3), model 1.[a]

Variable	NLS	Bayes	
		Prior A	Prior B
CNST	− 7.836 (0.457)	− 8.749 (8.114) [0.629] {0.74}	− 7.927 (2.425) [0.237] {0.86}
PCI	0.193 (0.150)	0.669 (0.144) [0.004] {0.59}	0.653 (0.146) [0.004] {0.52}
PE	− 0.175 (0.051)	− 0.188 (0.065) [0.0001] {0.38}	− 0.187 (0.065) [0.001] {0.37}
PG	− 0.076 (0.071)	− 0.106 (0.069) [0.001] {0.38}	− 0.102 (0.068) [0.001] {0.37}
CDD	$2.46 \times 10^{-4}(4.16 \times 10^{-5})$	$2.44 \times 10^{-5}(2.26 \times 10^{-5})$ [0.000001] {0.23}	$2.50 \times 10^{-5}(2.31 \times 10^{-5})$ [0.000001] {0.16}
HDD	$2.33 \times 10^{-4}(4.15 \times 10^{-5})$	$3.36 \times 10^{-4}(2.67 \times 10^{-5})$ [0.000001] {0.29}	$3.36 \times 10^{-4}(2.74 \times 10^{-5})$ [0.000001] {0.29}
ϕ_1	0.535 (0.161)	0.554 (0.146) [0.006] {0.29}	0.552 (0.140) [0.006] {0.25}
ϕ_2	− 0.625 (0.151)	0.360 (0.129) [0.004] {0.12}	0.335 (0.130) [0.004] {0.05}
ϕ_3	0.521 (0.161)	− 0.524 (0.146) [0.006] {0.27}	− 0.493 (0.141) [0.006] {0.22}
ϕ_4	− 0.338 (0.148)	0.578 (0.124) [0.004] {0.14}	0.560 (0.124) [0.004] {0.03}
σ^2	7.73×10^{-4}(—)	$7.92 \times 10^{-4}(1.86 \times 10^{-4})$ [0.0002] {0.20}	$7.84 \times 10^{-4}(1.85 \times 10^{-4})$ [0.000002] {0.23}

[a] NLS denotes the nonlinear least squares estimate. Prior B imposes stationarity. The Bayes estimates are based on $M = 50$ and $N = 1200$. Standard deviation (standard error for NLS) is in brackets; the numerical standard error of the posterior mean is in square brackets; the lag 1 correlation of the Gibbs run is in curly brackets.

Table 4

AR(4) models (cf. example 3 continued), model 2.[a]

Variable	NLS	Bayes	
		Prior A	Prior B
CNST	− 9.117 (0.437)	− 8.014 (10.16) [0.776] {0.68}	− 8.329 (1.950) [0.164] {0.84}
PCI	0.630 (0.140)	0.653 (0.139) [0.003] {0.49}	0.634 (0.141) [0.004] {0.51}
PE	− 0.213 (0.060)	− 0.216 (0.063) [0.001] {0.30}	− 0.213 (0.063) [0.001] {0.45}
PG	—	—	—
CDD	—	—	—
HDD	$3.46 \times 10^{-4}(1.51 \times 10^{-5})$	$3.45 \times 10^{-4}(1.60 \times 10^{-5})$ [0.000001] {0.28}	$3.44 \times 10^{-4}(1.75 \times 10^{-5})$ [0.000001] {0.50}
ϕ_1	0.543 (0.140)	0.573 (0.142) [0.006] {0.24}	0.563 (0.147) [0.006] {0.28}
ϕ_2	0.389 (0.134)	0.392 (0.130) [0.004] {0.06}	0.363 (0.125) [0.005] {0.07}
ϕ_3	0.548 (0.138)	− 0.546 (0.146) [0.005] {0.19}	− 0.520 (0.144) [0.005] {0.23}
ϕ_4	0.525 (0.132)	0.550 (0.122) [0.004] {0.04}	0.531 (0.120) [0.004] {0.06}
σ^2	7.53×10^{-4} (—)	$8.06 \times 10^{-4}(1.87 \times 10^{-4})$ [0.000002] {0.19}	$7.85 \times 10^{-4}(1.82 \times 10^{-4})$ [0.000002] {0.23}

[a] NLS denotes the nonlinear least squares estimate. Prior A does not impose stationarity; prior B imposes stationarity. The Bayes estimates are based on $M = 50$ and $N = 1200$. Standard deviation (standard error for NLS) is in brackets; the numerical standard error of the posterior mean is in square brackets; the lag 1 correlation of the Gibbs run is in curly brackets.

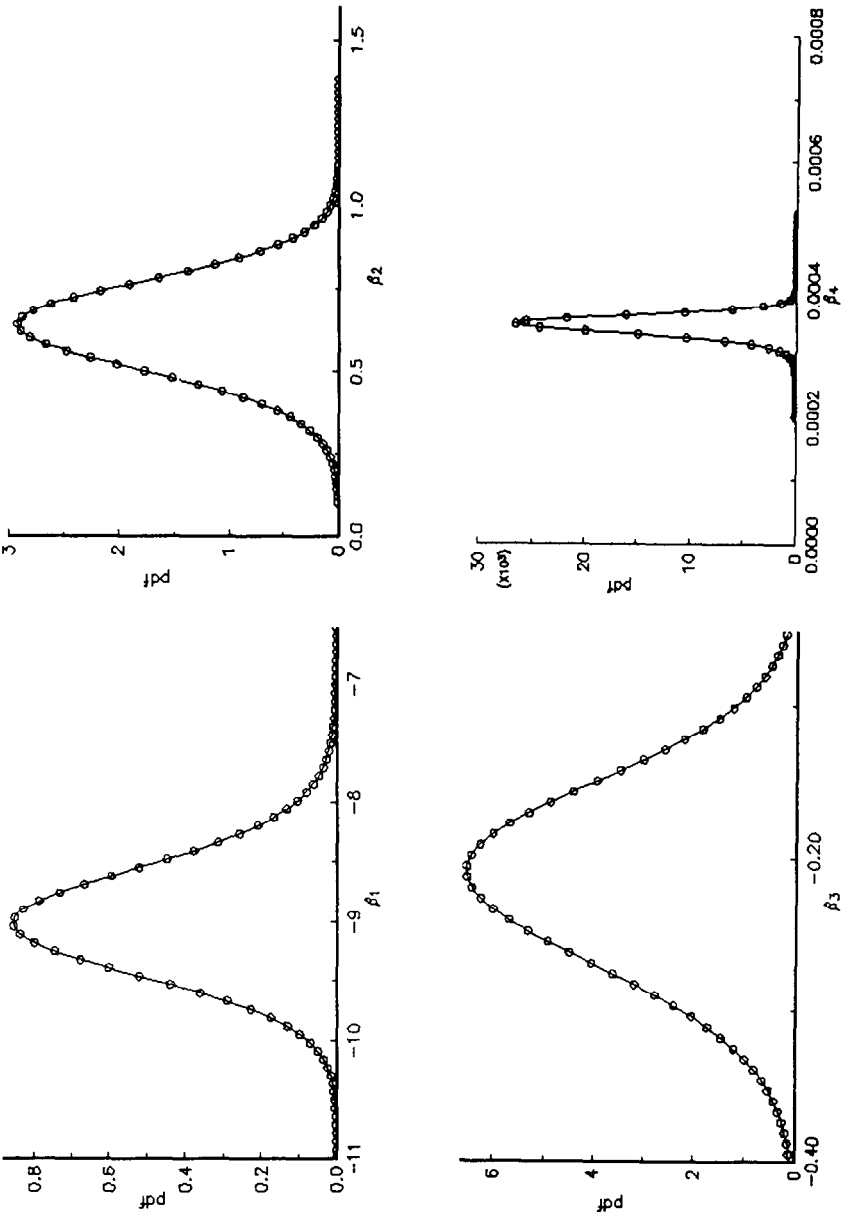

Fig. 1. Posterior pdf's of β in example 3, model 2.

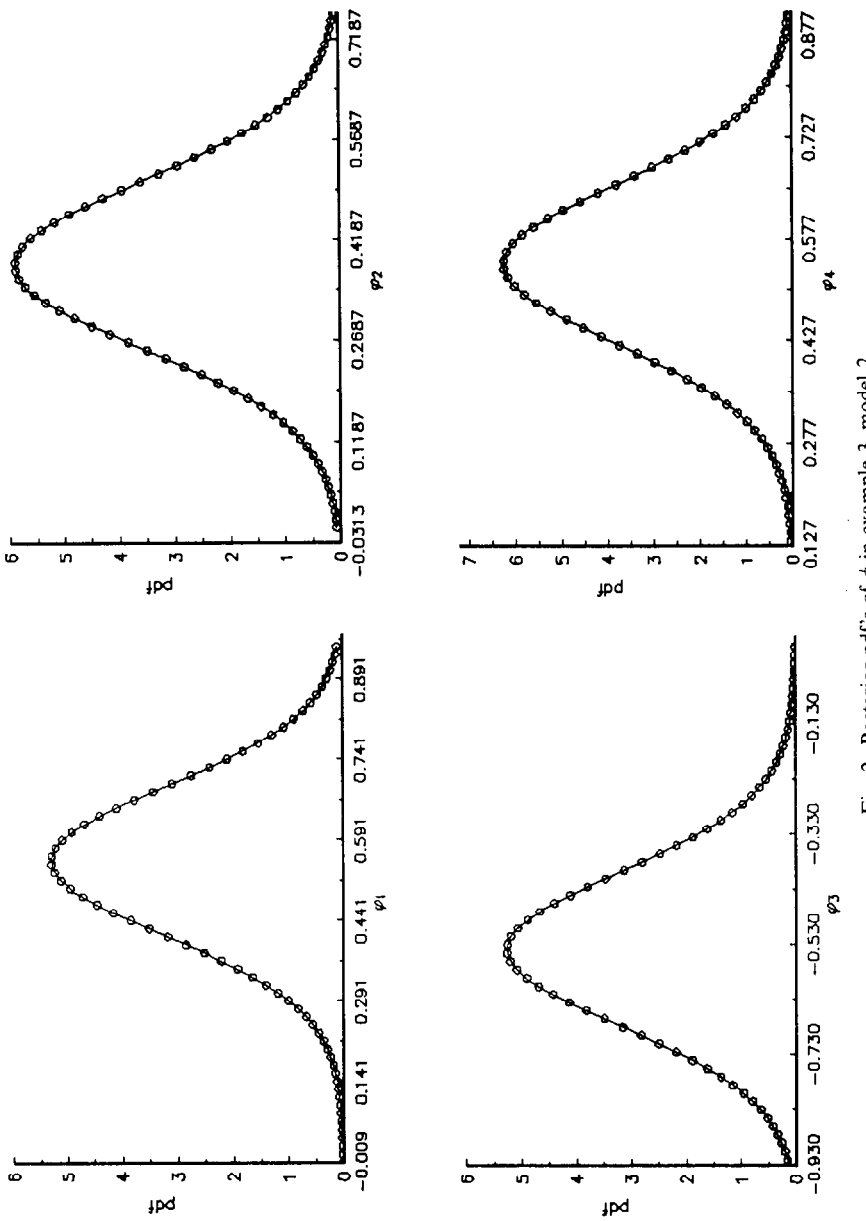

Fig. 2. Posterior pdf's of ϕ in example 3, model 2.

6. Concluding remarks

In this paper we have developed a practical and straightforward Bayesian approach to regression models with autocorrelated errors. It is shown that the proposed approach can readily deal with high-order autoregressive processes without requiring an importance sampling function or other tuning constants. Based on this work, it should be possible to analyze other models with correlated errors, for example, the seemingly unrelated regression model. Another interesting extension would be to relax the conditioning on the initial observations. We plan on providing some results on these extensions in future work.

Appendix: Data set used in example 3

KWH	*CNST*	*PCI*	*PE*	*PG*	*CDD*	*HDD*
− 6.538	1.000	2.997	− 3.149	− 1.494	2.000	515.0
− 6.733	1.000	3.019	− 3.095	− 1.392	62.00	213.0
− 6.746	1.000	3.054	− 3.102	− 1.202	564.0	0.000
− 6.638	1.000	3.059	− 3.152	− 1.339	55.00	409.0
− 6.458	1.000	2.996	− 3.166	− 1.442	8.000	812.0
− 6.671	1.000	3.076	− 3.100	− 1.331	57.00	281.0
− 6.669	1.000	3.096	− 3.115	− 1.125	693.0	0.000
− 6.531	1.000	3.105	− 3.164	− 1.342	108.0	564.0
− 6.418	1.000	3.105	− 3.176	− 1.450	0.000	652.0
− 6.668	1.000	3.139	− 3.091	− 1.299	105.0	112.0
− 6.623	1.000	3.111	− 3.125	− 1.155	594.0	0.000
− 6.522	1.000	3.126	− 3.066	− 1.322	54.00	402.0
− 6.387	1.000	3.108	− 3.110	− 1.451	0.000	625.0
− 6.603	1.000	3.110	− 3.063	− 1.343	124.0	169.0
− 6.632	1.000	3.098	− 3.078	− 1.206	430.0	0.000
− 6.543	1.000	3.110	− 3.078	− 1.341	79.00	343.0
− 6.470	1.000	3.078	− 3.033	− 1.435	0.000	603.0
− 6.674	1.000	3.079	− 2.888	− 1.324	80.00	144.0
− 6.636	1.000	3.087	− 2.880	− 1.172	537.0	0.000
− 6.542	1.000	3.084	− 2.873	− 1.299	94.00	376.0
− 6.416	1.000	3.064	− 2.895	− 1.414	0.000	735.0
− 6.618	1.000	3.113	− 2.863	− 1.276	19.00	275.0
− 6.666	1.000	3.128	− 2.838	− 1.116	467.0	0.000
− 6.533	1.000	3.137	− 2.764	− 1.237	62.00	406.0
− 6.455	1.000	3.088	− 2.741	− 1.338	24.00	494.0
− 6.605	1.000	3.110	− 2.734	− 1.245	181.0	131.0
− 6.584	1.000	3.130	− 2.753	− 1.139	705.0	0.000
− 6.513	1.000	3.149	− 2.769	− 1.221	302.0	168.0
− 6.431	1.000	3.096	− 2.793	− 1.319	14.00	461.0
− 6.620	1.000	3.117	− 2.799	− 1.241	37.00	194.0
− 6.600	1.000	3.146	− 2.806	− 1.116	694.0	0.000
− 6.552	1.000	3.214	− 2.818	− 1.216	176.0	92.00
− 6.438	1.000	3.131	− 2.786	− 1.306	46.00	286.0
− 6.614	1.000	3.141	− 2.749	− 1.190	313.0	51.00
− 6.571	1.000	3.132	− 2.776	− 1.198	740.0	0.000

− 6.471	1.000	3.143	− 2.825	− 1.322	177.0	399.0
− 6.352	1.000	3.120	− 2.870	− 1.327	10.00	617.0
− 6.593	1.000	3.086	− 2.902	− 1.354	221.0	71.00
− 6.560	1.000	3.081	− 2.839	− 1.083	847.0	0.000
− 6.491	1.000	3.113	− 2.894	− 1.211	137.0	215.0
− 6.456	1.000	3.061	− 2.875	− 1.243	18.00	271.0
− 6.617	1.000	3.021	− 2.731	− 1.249	160.0	105.0
− 6.634	1.000	3.057	− 2.455	− 1.080	712.0	0.000
− 6.606	1.000	3.058	− 2.504	− 1.195	108.0	214.0
− 6.551	1.000	3.008	− 2.515	− 1.183	36.00	330.0
− 6.708	1.000	3.009	− 2.568	− 1.140	351.0	41.00
− 6.631	1.000	3.030	− 2.579	− 1.073	943.0	0.000
− 6.616	1.000	3.025	− 2.641	− 1.154	101.0	202.0
− 6.538	1.000	2.997	− 2.513	− 1.140	13.00	516.0
− 6.723	1.000	3.014	− 2.506	− 1.063	132.0	75.00
− 6.674	1.000	3.015	− 2.493	− 0.9528	740.0	0.000
− 6.631	1.000	3.021	− 2.483	− 0.9287	176.0	322.0
− 6.590	1.000	2.996	− 2.409	− 0.8698	12.0	335.0

References

Albert, J. and S. Chib, 1991, Bayesian inference for binary and polychotomous response data, Journal of the American Statistical Association, forthcoming.

Carlin, B.P. and N.G. Polson, 1991, Inference for nonconjugate Bayesian models using the Gibbs sampler, Canadian Journal of Statistics 19, 399–405.

Chib, S., 1992, Bayes inference in the Tobit censored regression model, Journal of Econometrics 51, 79–99.

Devroye, L., 1986, Non-uniform random variate generation (Springer-Verlag, New York, NY).

Fomby, T.B. and D.K. Guilkey, 1978, On choosing the optimal level of significance of the Durbin-Watson test and the Bayesian alternative, Journal of Econometrics 8, 203–214.

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

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

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

Geweke, J., 1992, Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds., Bayesian statistics 4 (Oxford University Press, New York, NY).

Harvey, A.C., 1990, The econometric analysis of time series, 2nd ed. (MIT Press, Cambridge, MA).

Judge, G.G., W.E. Griffiths, R.C. Hill, and T.C. Lee, 1985, The theory and practice of econometrics, 2nd ed. (Wiley, New York, NY).

Judge, G.G., W.E. Griffiths, R.C. Hill, and T.C. Lee, 1988, Introduction to the theory and practice of econometrics, 2nd ed. (Wiley, New York, NY).

Keener, R.W., J. Kmenta, and N.C. Weber, 1991, Estimation of the covariance matrix of the least-squares regression coefficients when the disturbance covariance matrix is of unknown form, Econometric Theory 7, 22–45.

Kennedy, P. and D. Simons, 1991, Fighting the teflon factor: Comparing classical and Bayesian estimators for autocorrelated errors, Journal of Econometrics 48, 15–27.

McCulloch, R.E. and R.S. Tsay, 1991, Bayesian analysis of autoregressive time series via the Gibbs sampler, Manuscript.

Ramanathan, R., 1989, Introductory econometrics with applications (Harcourt Brace Jovanovich, Orlando, FL).

Richard, J.F., 1975, Bayesian analysis of the regression model when the disturbances are generated by an autoregressive process, in: A. Aykac and C. Brumat, eds., New developments in the application of Bayesian methods (North-Holland, Amsterdam).
Ripley, B.D., 1987, Stochastic simulation (Wiley, New York, NY).
Tanner, T.A. and W.H. Wong, 1987, The calculations of posterior distributions by data augmentation, Journal of the American Statistical Association 82, 528–549.
Tierney, L, 1991, Markov chains for exploring posterior distributions, Manuscript.
Zeger, S.L. and M.R. Karim, 1991, Generalized linear models with random effects: A Gibbs sampling approach, Journal of the American Statistical Association 86, 79–86.
Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York, NY).
Zellner, A. and G.C. Tiao, 1964, Bayesian analysis of the regression model with autocorrelated errors, Journal of the American Statistical Association 59, 763–778.