TECHNICAL NOTE

BAYESIAN ESTIMATION OF A SIMPLE SIMULTANEOUS EQUATION MODEL USING GIBBS SAMPLING

Mansor H. Ibrahim

Assistant Professor, Department of Economics, Kulliyyah of Economics and Management Sciences, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia.

Abstract

This technical note illustrates the estimation of a simple simultaneous equation model using Gibbs sampling. The results of the estimation show that Gibbs sampling can be an alternative approach for estimating the parameters of the model in a Bayesian setting.

1.   Introduction

This note is concerned with the estimation of Haavelmo's well-known simultaneous equation model using Gibbs sampling. Our purpose is two-fold. First, we illustrate the Gibbs sampling estimation approach, which could be a viable alternative estimation technique. Thus, we are not concerned with the comparative performance of alternative estimation methods. Second, we show how we can apply the Metropolis-Hasting algorithm in the Gibbs iteration process.

The model investigated consists of a consumption equation and an identity, namely,

$$(1) \qquad c_t = \beta + \alpha y_t + u_t$$

$$(2) \qquad y_t = c_t + z_t \qquad (t = 1, 2, ..., T)$$

where c_t is consumers' expenditure per capita, y_t is disposable income per capita, and z_t is exogenous investment expenditure per capita. All variables are in constant dollars. β is the constant term and α is the marginal propensity to consume assumed to be between 0 and 1. u_t is an error term normally distributed with mean zero and variance σ^2.

Chetty (1974) has estimated the above model using the Bayesian approach based on a numerical integration method (Simpson's rule). The cumbersome task of deriving marginal posterior distributions, however, may be avoided by applying the Gibbs sampling approach to the model. Specifically, the estimation is based on simulating samples from full conditional distributions, which are generally easy to derive and simulate from. Also, the method possesses very appealing theoretical properties and the joint density of the draws will converge to the true density as the number of draws becomes large (Gelfand and Smith, 1990). Lastly, the method can be readily applied to different specifications of the models, for example, models with

autoregressive error terms (Chetty, 1974) and with random coefficients (Tsurumi and Shiba, 1982).

The plan of the paper is as follows. In the next section, we briefly discuss the Gibbs Sampling estimation technique. In section 3, the full conditional distributions of the model are derived. Section 4 presents the results of the estimation using both simulated and empirical data. For the purpose of comparison, we also estimate the model using Full Information Maximum Likelihood Estimation (FIML). Finally, section 5 contains our concluding remarks.

2.   Gibbs Sampling: A Brief Overview

Gibbs sampling is a technique that allows us to indirectly generate a sample of parameters of interest from a distribution, without having to calculate the density which may be very complicated. The central idea is quite simple. Let $f(\theta_1, \theta_2, ..., \theta_N \mid y)$ be a joint posterior density function of the parameters of interest θ_i, $i = 1, ..., N$, where y denotes the data. The Gibbs sampler then generates samples from the joint density of all parameters in the model. To begin the process, starting values $\theta_{i,0}$ are assigned to each parameter i. Then, the algorithm iterates as follows:

(a)   Set $t = 1$

(b)   Sample $\theta_{1,t}$ from $f(\theta_1 \mid \theta_{2,t-1}, ..., \theta_{N,t-1}, y)$.

(c)   Sample $\theta_{2,t}$ from $f(\theta_2 \mid \theta_{1,t}, ..., \theta_{N,t-1}, y)$.

.

.

.

(d)   Sample $\theta_{N,t}$ from $f(\theta_N \mid \theta_{2,t}, ..., \theta_{N-1,t}, y)$.

(e)   Set $t = t + 1$ and return to (b)

where $f(\theta_i \mid .)$, $i = 1, ..., N$, are full conditional density functions.

Under very mild regularity conditions, the joint distribution of $(\theta_{1,t}, \ldots, \theta_{N,t})$ converges in distribution to the joint distributions of the parameters. The samples of the parameters can be obtained by independently repeating the cycle above M times. Alternatively, we may simulate the parameters iteratively for $T + M$ cycles, where T is a large enough number. The last M draws are then used as a basis for inference. That is, based on the M generated draws, posterior moments of any function, $g(\theta)$, as well as the marginal density of a component of θ can be calculated easily (see Casella and George, 1992; Gelfand et al. 1990; and Chib, 1992).

3. Full Conditional Distributions

Given a diffuse prior, $p(\alpha, \beta, \sigma^2) \propto \sigma^2$, the posterior density function of the model specified in (1) and (2) is:

(3) $$p(\alpha, \beta, \sigma^2 \mid c, y) \propto \frac{|1-\alpha|^T}{(\sigma^2)^{\frac{T}{2}+1}} \exp\left\{ -\frac{1}{2\sigma^2} \sum (c_t - \beta - \alpha y_t)^2 \right\}$$

where T is the total number of observations. The full conditional distribution can be easily shown to be:

$$p(\alpha \mid \beta, \sigma^2 c, y) \propto |1-\alpha|^T N(\bar{a}, \bar{A})$$

(4) $$\bar{a} = \frac{\sum c_t y_t - \beta \sum y_t}{\sum y_t^2}, \quad \bar{A} = \sigma^2 (\sum y_t^2)^{-1}$$

and

$$p(\beta \mid \alpha, \sigma^2 c, y) \sim N(\bar{b}, \bar{B})$$

(5) $$\bar{b} = \bar{c} - \alpha \bar{y}, \quad \bar{B} = \frac{\sigma^2}{T}, \quad \bar{c} = \sum \frac{c_t}{T}, \quad \bar{y} = \sum \frac{y_t}{T};$$

and lastly

$$(6) \qquad p(\sigma^2 | \alpha, \beta, c, y) \sim IG\left(\frac{T}{2}, \frac{\sum (c_t - \beta - \alpha y_t)^2}{2} \right)$$

In Gibbs sampling (4)–(6) are used for simulating the parameters of the simultaneous model (1)–(2). Note that (5) and (6) are easily simulated. However, in (4), we have an additional term attached to a standard normal distribution. Since the marginal propensity to consume, α, is theoretically between zero and one, an acceptance-rejection approach may be applied. In the estimation stage, however, we do not enforce α to be between 0 and 1. In this method, α is generated from $N(\bar{a}, \bar{A})$ and u from $U(0,1)$. Then if $u \leq |1-\alpha|^T$, α is accepted as a value generated from (4). If not, u and α are rejected and another sample of (u, α) is generated and tested for acceptance. Unfortunately, the acceptance-rejection method applied to this case is very inefficient. The draw is rarely accepted. Indeed, because α is less than unity but greater than zero, $|1-\alpha|^T$ will become very small and draws are rejected too often.

However, the Metropolis-Hasting algorithm can be applied to (4) (see Hasting, 1970, and Tierney, 1991, for details on this algorithm). In this case we specify $q(\alpha^*, \alpha^{**}) = N(\bar{a}, \bar{A})$. α^* is the current value of α, and α^{**} is the candidate for the next value generated from $N(\bar{a}, \bar{A})$. The acceptance probability $P(\alpha^*, \alpha^{**})$ can then be written as:

$$(7) \qquad P(\alpha^*, \alpha^{**}) = \min\left\{ \frac{w(\alpha^{**})}{w(\alpha^*)}, 1 \right\}$$

In our case, $w(\alpha)$ reduces to $|1-\alpha|^T$. Thus, the acceptance probability

becomes:

$$(8) \qquad P(\alpha^*, \alpha^{**}) = \min\left\{ \frac{\left|1 - \alpha^{**}\right|^T}{\left|1 - \alpha^*\right|^T}, 1 \right\}$$

This means that we set the next value of α, $\alpha^{t+1} = \alpha^{**}$ with probability P and $\alpha^{t+1} = \alpha^*$ with probability $(1 - P)$. The only problem that this approach might have is that α will persistently stay at the same value. However, in our case, we do not encounter this problem when we perform the simulations, the results of which are reported in the next section.

4.   Estimation Results

We estimate the simultaneous equation above using simulated as well as empirical data. We iteratively draw the parameters from (4)–(6) for 2500 cycles and keep the values of the last 2000 cycles as draws of the parameters. OLS estimation of (1) is used to initialize the values of the parameters.

Table 1 reports the estimation results using simulated data.[1] Specifically, z_t is simulated independently from a normal distribution with mean and standard deviation equal to 5 and 2 respectively. The parameters of the model, $\{\beta, \alpha, \sigma^2\}$, are $\{5, 0.6, 1\}$. The results of the estimation are presented in Table 1, where we compare the Gibbs estimates to the FIML estimates.[2] Note that Gibbs sampling performs well in estimating the coefficient estimates, especially when the sample

Table 1: Parameter Estimates Using Simulated Data
(Actual values: $\beta = 5$, $\alpha = 0.6$, $\sigma^2 = 1$, and n = sample size)

Parameters	$n = 20$		$n = 200$		$n = 500$	
	GS	FIML	GS	FIML	GS	FIML
α	0.5503	0.5647	0.5992	0.5964	0.6001	0.5994
	(0.0412)	(0.0589)	(0.0130)	(0.0179)	(0.0077)	(0.0094)
β	6.6527	6.2504	5.0397	5.1083	4.9583	4.9763
	(1.1699)	(1.6112)	(0.3356)	(0.4544)	(0.1997)	(0.2333)
σ^2	0.6901	0.5680	1.3555	1.3604	1.0107	1.0101
	(0.3138		(0.1637)		(0.1755)	

Note: The point estimates of the parameters for Gibbs sampling (GS) are the sample mean. The numbers in parentheses are the standard deviations of the sample draws for GS and standard errors for FIML.

size is large. As the sample size becomes larger, the estimated coefficients are closer to the actual values. The same also holds true for FIML. Thus, Gibbs sampling may provide a useful alternative for the estimation of the model.

We also estimate the model using a data set from Chetty (1974, 362, Table 1). The estimation results, using Gibbs sampling as well as FIML, are reported in Table 2. For the purpose of comparison, we also present the results obtained by Chetty (1974). Note that the marginal propensity to consume out of income, α, is of the same order of magnitude. The point estimates of α are 0.688 (Gibbs sampling), 0.6767 (FIML), and 0.660 (Chetty). The standard deviations of the estimated αs are also small. In addition, Gibbs sampling seems to be relatively

Table 2: Parameter Estimates

Parameters	Chetty, 1974	Gibbs Sampling	FIML
α	0.660 (0.0632	0.668 (0.0361)	0.6767 (0.0348)
β	111.589 (11.386)	115.920 (17.239)	111.665 (16.994)
σ^2	–	93.762	73.87

Note: See note of Table 1. For comparison, we took the squareroot of Chetty's estimated variances of the parameters.

efficient in estimating b. In Chetty's, the mean value of β is 111.589 while FIML estimation of β is 111.665. Similarly, Gibbs sampling's estimated value of β is 115.720.

Overall, it may be concluded from this technical exercise that Gibbs sampling is capable of providing relatively efficient estimates of the parameters in the model, especially when the number of observations is large. Also, use of this method may be possible when dealing with much larger models than can be done by numerical integration.

5.   Conclusion

In this note, we estimated a simple simultaneous model using the Gibbs sampling estimation. The model was first formulated by Haavelmo (1947) and analyzed in a Bayesian setting, using numerical integration methods, by Chetty (1974). The results of estimation using Gibbs sampling show that it can be an alternative approach for

estimating the parameters of the model in a Bayesian setting. The Gibbs sampling algorithm is easy to code and thereby avoids the difficulty of deriving marginal posterior distributions directly. Therefore, it provides a useful alternative to the maximum likelihood estimation methods. The method should prove useful for more elaborate simultaneous equation models.

Endnotes

1. The estimation results for Gibbs are estimated as follows: $E(x) = (1/T)\Sigma x$ and standard deviation $s_x = [(1/T)\Sigma(x - E(x))^2]^{1/2}$. In this case, x is a simulated sample of the parameters.

2. Since our focus is technically on the applicability of the Gibbs sampling, we do not go into details as to the question of which method of estimation is efficient. In fact, it may not be fair to compare Gibbs sampling with FIML, as noted by an anonymous referee. For comparative analyses of different alternative estimation methods, see Zellner (1997).

References

Casella, G. and E. George. "Explaining the Gibbs Sampler." *American Statistician* 46 (1992): 167–74.

Chib, S. "Bayes Regression for the Tobit Censored Regression Model." *Journal of Econometrics* 51 (1992): 79–99.

Chetty, V. K. "Bayesian Analysis of Haavelmo's Models." In *Studies in Bayesian Econometrics and Statistics In Honor of Leonard J. Savage*, edited by S.E. Foenberg And A. Zellner, Amsterdam: North-Holland, 1974: 359-81.

Gelfand, A.E. and A. F. M. Smith. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85 (1990): 398–409.

————, S. E. Hills, A. Racine-Poon and A. F. M. Smith. "Illustration of

Bayesian Inference in Normal Data Models Using Gibbs Sampling." *Journal of the American Statistical Association* 85 (1990): 972–84.

Haavelmo, T. "Methods of Measuring the Marginal Propensity to Consume." *Journal of the American Statistical Association* 42 (1947): 105–22.

Hasting, W. K. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrica* 57 (1970): 97–109.

Tierney, L. "Markov Chains for Exploring Posterior Distributions." *Technical Report 560*, University of Minnesota: School of Statistics, 1991.

Tsurumi, H. and T. Shiba. "A Bayesian Analysis of a Random Coefficient Model In a Simple Keynesian System." *Journal of Econometrics* 18 (1982): 239–49.

Zellner, A. *Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers*. United Kingdom: Edward Elgar, 1997.