THE KOREAN ECONOMIC REVIEW
Volume 13, Number 1, Summer 1997.

UNCONDITIONAL ESTIMATION OF TIME-VARYING-PARAMETER MODELS: A GIBBS-SAMPLING APPROACH*

CHUNG-KI MIN**

This study addresses the issue of unconditional estimation of regression models with time-varying parameters. Using a data augmentation in which unobserved random coefficients are treated as missing data, procedures for the Gibbs sampler are developd. Several examples are presented to illustrate how the Gibbs-sampling procedures perform in practice.

I. INTRODUCTION

Regression models with time-varying parameters have received a great deal of attention in econometrics. One simple form of the time-varying-parameter (TVP) models is:

$$y_t = x_t'\beta_t + u_t, \qquad u_t \sim \text{iid } N(0, \sigma^2) \tag{1}$$

$$\beta_t = \beta_{t-1} + \eta_t, \qquad \eta_t \sim \text{iid } N(0, \delta^2 I) \tag{2}$$

where a $(k \times 1)$ coefficient vector β_t follows a vector random walk process. This TVP model contains a hyperparameter δ which determines the variability of regression coefficients $\{\beta_t, t=1, \cdots, T\}$ over time. For example, $\delta=0$ implies constant regression coefficients over time, i. e., $\beta_t=\beta_0$ for all t. When the hyperparameter δ is known, this model can easily be estimated and be used for forecasting problems with an application of the Kalman filtering algorithm. However, since the hyperparameter δ is unknown in most econometric applications, previous studies in the literature have conditioned δ on chosen values.[1] For example, the maxi-

* I would like to thank Siddhartha Chib, Edward I. George, Robert E. McCulloch, Mike West, Herman K. van Dijk, Arnold Zellner, and two anonymous referees for their helpful comments. Of course, all remaining errors belong to me.

** Hankuk University of Foreign Studies.

[1] Previous studies used models in which a hyperparameter δ^{*2} represented the ratio of the variance

mum likelihood approach is conditioned on the value of δ which maximizes a concentrated likelihood function of δ (Cooley and Prescott(1976), Doan et al. (1983), and Harvey(1978)). Using Bayesian methods, Liu and Hanssens(1981) have derived a conditional posterior density for regression coefficients using the modal value of the marginal posterior density for δ; that is, $p(\beta_T \mid \hat{\delta}, D)$ where $\hat{\delta}$ is the mode and D denotes the data available for estimation. However, unless the conditional posterior densities are insensitive to changes in δ, the conditional analysis may be misleading with a choice of an inappropriate value of δ. Rather, it would be desirable to estimate the marginal posterior densities for parameters by integrating out the hyperparameter δ.

The Bayesian approach to estimating marginal posterior densities of the TVP model requires integrations which are analytically intractable. Previous efforts to approximate marginal posterior densities include: (1) analytic approximations (Tierney and Kadane(1986)); (2) importance sampling method (Chib et al.(1990) and Zellner and Rossi(1982)); and (3) quantile integration method (Johnson (1992)). However, these methods are practically impossible to apply when a hyperparameter space is high-dimensional.[2] More recently, a Markov-Chain Monte Carlo simulation method, known as the Gibbs sampler, has been developed for Bayesian analysis by Gelfand and Smith(1990) and others. The Gibbs sampler calculates marginal and joint densities using random samples from their full conditional densities. As shown in what follows, although marginal posterior densities of the TVP model are analytically intractable, the model's conditional posterior densities are well-known distributions from which it is easy to draw random samples. Therefore, the Gibbs sampler would be another approach to approximating the marginal posterior densities for the TVP model.

The purpose of this paper is to develop Gibbs sampling procedures which can approximate the marginal posterior densities for an extended version of the above TVP model (1) and (2). Other numerical methods such as the Simpson's rule are intractable for the extended model because its hyperparameter space is high-dimensional. The Gibbs sampling procedures work via a data augmentation in which unobserved random coefficient vectors, $\beta_1, \cdots, \beta_T$, are treated as missing data. It is shown that the Gibbs sampler performed well for data generated from regression and autoregressive models with time-varying parameters, in the sense that the Gibbs sampler converged and the Gibbs-estimated marginal posterior densities of parameters were close to their conditional posterior densities

---

of β_t to the variance of y_t, σ^2: i. e., $\eta_t \sim$ iid $N(0, \delta^{*2}\sigma^2 I)$. Since a high correlation between δ^* and σ causes an identification problem, the above specification (2) is employed in this study. I am grateful to Mike West for this suggestion.

[2] This study develops Gibbs sampling procedures for an extended version of the TVP model, in which the hyperparameter space is high-dimensional: i. e., $\beta_t = \beta_{t-1} + \eta_t$, $\eta_t \sim$ iid $N(0, \Delta)$ where Δ is a $k \times k$ matrix.

evaluated at the value of the hyperparameter δ used in the data generation. In order to estimate parameters more precisely, this paper extends the Gibbs-sampling procedures to use multiple observations in each time period. In addition, applications to optimal prediction problems are illustrated using squared and absolute error loss functions.

The rest of this paper is organized as follows. Section Ⅱ explains the problem with which this paper deals. In section Ⅲ, a Gibbs sampling procedure is developed for a TVP model with a single observation in each period and is applied to simulated data. Then, an extension of the Gibbs sampling procedure for the TVP model with multiple observations in each period is presented in section Ⅳ. In section Ⅴ, applications of the Gibbs sampler to optimal prediction problems are illustrated. Concluding remarks are made in the final section.

Ⅱ. THE PROBLEM

In order to use the TVP model, (1) and (2), for forecasting future values, we need to estimate the marginal posterior density for β_T, the last period's coefficient vector. The marginal posterior density for β_T can be obtained by integrating out σ and δ in the joint posterior density for $(\beta_T, \sigma, \delta)$: i. e.,

$$\begin{aligned} p(\beta_T \mid D_T) &= \int p(\beta_T, \sigma, \delta \mid D_T) d\delta \\ &= \int p(\beta_T \mid \sigma, \delta, D_T) \cdot p(\sigma \mid \delta, D_T) \cdot p(\delta \mid D_T) d\delta \end{aligned} \tag{3}$$

where D_T denotes the information available at time T, i. e., $D_T = \{y_T, \cdots, y_1, x_T, \cdots, x_1\}$. It is well known that the conditional posterior density for β_T, $p(\beta_T \mid \sigma, \delta, D_T)$, is a multivariate normal density. However, the problem arises with $p(\delta \mid D_T)$ which is too complicated for analytic integration. To show this, we first define a joint density for $(y_T, y_{T-1}, \cdots, y_{k+1})$, conditioned on (δ, σ) and D_k, as the product of one-step-ahead predictive densities:

$$\begin{aligned} & p(y_T, \cdots, y_{k+1} \mid \delta, \sigma, D_k, x_T, \cdots, x_{k+1}) \\ &\quad = \Pi_{t=k+1}^{T} p(y_t \mid \delta, \sigma, y_{t-1}, \cdots, y_1, x_t, \cdots, x_1) \\ &\quad \propto \{\Pi_{t=k+1}^{T} (\sigma^2 + x_t' V_{t-1} x_t)\}^{-\frac{1}{2}} \exp\{ - \sum_{t=k+1}^{T} (y_t - x_t'\hat{\beta}_{t-1})^2 / 2(\sigma^2 + x_t' V_{t-1} x_t)\} \\ &\quad \sim \Pi_{t=k+1}^{T} N\{x_t'\hat{\beta}_{t-1}, (\sigma^2 + x_t' V_{t-1} x_t)\} \end{aligned} \tag{4}$$

where $\hat{\beta}_{t-1} = E(\beta_{t-1} \mid \delta, \sigma, D_{t-1}) = E(\beta_t \mid \delta, \sigma, D_{t-1})$, $V_{t-1} = \mathrm{Cov}(\beta_t \mid \delta, \sigma, D_{t-1})$ and k is the dimension of a column vector β_t. In evaluating the joint density, the first k observations were used to compute $\hat{\beta}_k$ and V_k which were required for defining a conditional predictive density of y_{k+1}, $p(y_{k+1} \mid y_k, \cdots, y_1, \delta, \sigma, x_{k+1}, \cdots, x_1)$. Employing a prior density for (δ, σ) given D_k, $p(\delta, \sigma \mid D_k)$, we can derive a posterior density for δ as follows:

$$p(\delta \mid D_T) \propto \int p(\delta, \sigma \mid D_k) \cdot p(y_T, \cdots, y_{k+1} \mid \delta, \sigma, D_k, x_T, \cdots, x_{k+1}) d\sigma. \tag{5}$$

Unfortunately, the posterior density for δ resulting from the above integration is not a well-known distribution in general, and it is not possible to perform the integration required in (3) analytically.

In this study an indirect approach to the integration problem is employed for marginal posterior densities. The Gibbs sampling methods use only full conditional densities in approximating joint and marginal densities. As shown in the following section, the full conditional posterior densities in the TVP model (1) and (2) are well-defined and it is easy to draw random samples from them. For example, even though the marginal density $p(\delta \mid D_T)$ is of a complicated form, a conditional density for δ, e. g., $p(\delta \mid \beta_T, \cdots, \beta_0, D_T)$, is an inverted gamma density. Therefore, the Gibbs sampler would be able to approximate the marginal posterior densities for the TVP models.

Ⅲ. GIBBS SAMPLER FOR MODELS WITH A SINGLE OBSERVATION IN EACH PERIOD

3.1 Description of the Gibbs Sampler

The Gibbs sampler is a Monte Carlo method for approximating joint and marginal distributions by sampling from conditional distributions. This method is well discussed by Casella and George(1992), Gelfand and Smith(1990) and Geman and Geman(1984), among others. Further, the applications of the Gibbs sampler to various problems are found in Carlin et al.(1992), Blattberg and George (1991), Chib(1993), McCulloch and Rossi(1994), and the references cited therein. A basic idea of the Gibbs sampler is as follows. Let θ_1 and θ_2 be two random variables, possibly random vectors. Suppose their full conditional distributions are known and denoted by $p(\theta_1 \mid \theta_2)$ and $p(\theta_2 \mid \theta_1)$, respectively. Given an arbitrary starting value $\theta_2^{(0)}$, draw $\theta_1^{(1)}$ from $p(\theta_1 \mid \theta_2^{(0)})$ and $\theta_2^{(1)}$ from $p(\theta_2 \mid \theta_1^{(1)})$. Then repeat the drawing using $\theta_2^{(1)}$ as a new starting value. After K such iterations, we would arrive at $(\theta_1^{(K)}, \theta_2^{(K)})$. Geman and Geman(1984) have shown that under regularity conditions, the joint and marginal distributions of $(\theta_1^{(K)}, \theta_2^{(K)})$ converge to the joint and marginal distributions of θ_1 and θ_2: i. e., as $K \rightarrow \infty$, $(\theta_1^{(K)}, \theta_2^{(K)}) \rightarrow (\theta_1, \theta_2) \sim p(\theta_1, \theta_2)$ and $\theta_i^{(K)} \rightarrow \theta_i \sim p(\theta_i)$ for $i = 1, 2$.

The marginal densities and their moments of θ_1 and θ_2 may be approximated by repeating the Gibbs sampling iteration $K + N$ times and then using the drawn values in the last N iterations, where K is to be chosen large enough so that the Gibbs sampler has converged and N is to be chosen to give sufficient precision to the empirical distributions of interest. For example, the mean of the marginal distribution of θ_1 might be approximated by $\Sigma_{i=1}^N \theta_1^{(K+i)}/N$, or by $\Sigma_{i=1}^N E(\theta_1 | \theta_2^{(K+i)})/N$ using conditional expectations. And the marginal distribution of θ_1 might be approximated by the empirical distribution of $(\theta_1^{(K+1)}, \cdots, \theta_1^{(K+N)})$, or by $\Sigma_{i=1}^N E(\theta_1 | \theta_2^{(K+i)})/N$ using the information about conditional distributions.

Concerning the convergence issue, this study follows a practical suggestion made in McCulloch and Rossi(1994). We plot the estimates of posterior densities over Gibbs iterations. If these estimated posterior densities show little variation with additional Gibbs iterations, we may conclude that the Gibbs sampler has converged to the posterior densities. We also conduct an analysis of the sensitivity of estimated posterior distributions to various, widely dispersed starting values, e. g., $\theta_2^{(0)}$. Theoretical discussions on the convergence issue can be found in Geweke(1992), McCulloch and Rossi(1994), Tierney(1994) and Zellner and Min (1995). Further, Zellner and Min(1995) have developed convergence criteria which determine not only whether the Gibbs sampler has converged but also whether it has converged to a correct result.[3]

3.2 Full Conditional Posterior Densities

The following extended version of the TVP model, (1) and (2), is considered in what follows:

$$y_t = x_t'\beta_t + u_t, \qquad u_t \sim \text{iid } N(0, \sigma^2) \tag{6}$$

$$\beta_t = \beta_{t-1} + \eta_t, \qquad \eta_t \sim \text{iid } N(0, \Delta) \tag{7}$$

where Δ is a diagonal matrix with elements $(\delta_1^2, \delta_2^2, \cdots, \delta_k^2)$. If all of the diagonal elements are equal, i. e., $\delta_1 = \cdots = \delta_k = \delta$, then it reduces to the specification in (2), i. e., $\Delta = \delta^2 I$.

This extended TVP model can be expressed as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} x_1' & & & 0 \\ & x_2' & & \\ & & \ddots & \\ 0 & & & x_T' \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix} \tag{8}$$

[3] It is illustrated in Zellner and Min(1995) that the Gibbs sampler can be caught in trouble in several cases. Since the TVP model doesn't belong to those cases, the application of the convergence criteria proposed in Zellner and Min(1995) is not reported in this paper.

$$\begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ -I & I & 0 & \cdots & 0 \\ 0 & -I & I & & \vdots \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & & -I & I \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_T \end{bmatrix} = \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \beta_0 + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \\ \eta_T \end{bmatrix} \tag{9}$$

or, using matrix notations,

$$y_t = X\gamma + u, \qquad u \sim N(0, \sigma^2 I_T) \tag{10}$$

$$A\gamma = J\beta_0 + \eta, \qquad \eta \sim N(0, I_T \otimes \Delta) \tag{11}$$

where the definitions of y, X, γ, u, A, J and η are obvious, and $\otimes$ is the Kronecker product. This TVP model contains parameters $(\beta_0, \sigma, \Delta)$ and unobserved random variables $\gamma' = \{\beta_1', \cdots, \beta_T'\}$.

The Gibbs sampling procedures developed below work via a data augmentation in which the unobservable random variables γ are added to the distributional setup (Tanner and Wong(1987)). The full conditional posterior densities for $(\beta_0, \gamma, \sigma, \Delta)$ may be expressed as $p(\beta_0, \gamma \mid \sigma, \Delta, D_T)$ and $p(\sigma, \Delta \mid \beta_0, \gamma, D_T)$. Then, random drawings from the full conditional posterior densities can be implemented in a simple way using the following relationships:

$$p(\beta_0, \gamma \mid \sigma, \Delta, D_T) = p(\beta_0 \mid \sigma, \Delta, D_T) \cdot p(\gamma \mid \beta_0, \sigma, \Delta, D_T) \tag{12}$$

$$p(\sigma, \Delta \mid \beta_0, \gamma, D_T) = p(\sigma \mid \beta_0, \gamma, D_T) \cdot p(\Delta \mid \sigma, \beta_0, \gamma, D_T) \tag{13}$$

That is, we can obtain a random sample from $p(\beta_0, \gamma \mid \sigma, \Delta, D_T)$ by first drawing β_0^* from $p(\beta_0 \mid \sigma, \Delta, D_T)$ and then drawing γ^* from $p(\gamma \mid \beta_0^*, \sigma, \Delta, D_T)$. Similarly, we first draw σ^* from $p(\sigma \mid \beta_0, \gamma, D_T)$ and then draw Δ^* from $p(\Delta \mid \sigma^*, \beta_0, \gamma, D_T)$. Since $p(\sigma \mid \beta_0, \gamma, D_T)$ is independent of β_0 and $p(\Delta \mid \sigma, \beta_0, \gamma, D_T)$ is of σ and D_T, we complete the Gibbs sampler by specifying $p(\beta_0 \mid \sigma, \Delta, D_T)$, $p(\gamma \mid \beta_0, \sigma, \Delta, D_T)$, $p(\sigma \mid \gamma, D_T)$ and $p(\Delta \mid \beta_0, \gamma)$.

1. Priors:

$$p(\beta_0 \mid \sigma, \Delta) \propto \text{constant} \tag{14}$$

$$p(\gamma \mid \beta_0, \sigma, \Delta) \sim N\{A^{-1}J\beta_0, A^{-1}(I_T \otimes \Delta)A^{-1'}\} \tag{15}$$

$$p(\sigma \mid \gamma) \propto 1/\sigma \tag{16}$$

$$p(\Delta) \propto |\Delta|^{-1/2} \tag{17}$$

where the conditional prior for γ, (15), is obtained from (11): i. e., $\gamma = A^{-1} J\beta_0 + A^{-1}\eta$.[4)]

2. Conditional Posterior Densities[5)]:

$$p(\beta_0 \mid \sigma, \Delta, D_T) \sim N(\hat{\beta}_0, V) \tag{18}$$

$$p(\gamma \mid \beta_0, \sigma, \Delta, \gamma, D_T) \sim N(\hat{\gamma}, W) \tag{19}$$

$$p(\sigma \mid \gamma, D_T) \propto \frac{1}{\sigma^{T+1}} \exp\{-(y - X\gamma)'(y - X\gamma)/2\sigma^2\} \tag{20}$$

$$p(\delta_i \mid \beta_0, \gamma) \propto \frac{1}{\delta_i^{T+1}} \exp\{-\sum_{t=1}^{T} (\beta_{it} - \beta_{it-1})^2/2\delta_i^2\} \text{ for } i = 1, \cdots, k \tag{21}$$

where $\beta_0 = V \times (J'A^{-1\prime}X'Q^{-1}y), Q = (\sigma^2 I + XA^{-1}(I_T \otimes \Delta)A^{-1\prime}X'), V = (J'A^{-1\prime}X'Q^{-1}XA^{-1}J)^{-1}$, $\hat{\gamma} = W \times \{X'y/\sigma^2 + A'(I_T \otimes \Delta)^{-1} J\beta_0\}$, and $W = \{X'X/\sigma^2 + A'(I_T \otimes \Delta)^{-1} A\}^{-1}$.

3. Gibbs Sampler:

With the full conditional densities derived above, a Gibbs sampling procedure can be designed as follows:

(1) Initialize the values of σ and Δ, e. g., $\sigma^{(0)}$ and $\Delta^{(0)}$.
(2) Repeat steps (a)−(d) $K+N$ times. For $i = 1, \cdots, K+N$,
  (a) Sample $\beta_0^{(i)}$ from $p(\beta_0 \mid \sigma^{(i-1)}, \Delta^{(i-1)}, D_T)$ which is a multivariate normal, as in (18).
  (b) Sample $\gamma^{(i)}$ from $p(\gamma \mid \beta_0^{(i)}, \sigma^{(i-1)}, \Delta^{(i-1)}, D_T)$ which is a multivariate normal, as in (19).
  (c) Sample $\sigma^{(i)}$ from $p(\sigma \mid \gamma^{(i)}, D_T)$ which is an inverted gamma (IG) density, as in (20).

[4] The difffuse priors used for β_0, σ and Δ are improper. However, it is well known that the posteriors become proper with enough sample information. Although this study employs diffuse priors to minimize the effects of prior information, informative proper priors could be employed without difficulty.

[5] See Appendix for the detailed derivation of the conditional posterior densities.

(d) Sample $\Delta^{(i)}$ from $p(\Delta \mid \beta_0^{(i)}, \gamma^{(i)})$ which is an IG density, as in (21).

3.3 Example 1

The Gibbs sampling procedure proposed above is applied to a simulated data set. The objective is to illustrate how the Gibbs sampling procedure performs in practice. Data are generated according to the following model,[6] with $\delta = 1$, $\sigma = 1$ and z_t's being drawn independently from a uniform distribution with an interval (0, 10) for $t = 1, 2, \cdots, 20$:[7]

$$y_t = x_t'\beta_t + u_t, \qquad u_t \sim N(0, \sigma^2) \tag{22}$$

$$\beta_t = \beta_{t-1} + \eta_t, \qquad \eta_t \sim N(0, \delta^2 I_2) \tag{23}$$

where $x_t' = (1 \; z_t)$ and $\beta_t = (\beta_{ct} \, \beta_{zt})'$, a 2×1 coefficient vector. While only one simulated case is presented in this section, the Gibbs sampler has performed similarly well for other data generated from multiple regression and autoregressive models.

The Gibbs sampler converged quite well according to the plots of the estimates of posterior densities over Gibbs iterations. And, the Gibbs-estimated marginal posterior densities for regression coefficients were close to the parameters' conditional posterior densities evaluated at the value of δ used for generating the data. Therefore, I may conclude that the Gibbs sampler performed well in estimating the marginal posterior densities.

The Gibbs-estimated marginal posterior densities in Figure 1 were obtained by averaging conditional posterior densities, e. g., $\Sigma_{i=1}^{19,000} \; p(\beta_{cT} \mid \delta^{(i)}, \sigma^{(i)}, D_T)/19,000$ where $\delta^{(i)}$ and $\sigma^{(i)}$ are the drawn values in the 1,001st through 20,000th Gibbs-sampler iterations. The modes of the Gibbs-estimated marginal posterior densities for β_{cT} and β_{zT} are -2.9 and -11.68, respectively, while the values of (β_{cT}, β_{zT}) used for the data generation are $(-3.37, -11.27)$. The Gibbs-estimated marginal density for β_{zT} has fatter tails than the conditional posterior density evaluated at $\delta = 1$, the value used for the data generation, although the differences in the tail areas are too small to be shown in the figure.

However, it seems that precise estimation of δ and σ is not possible when only one observation is available in each period. A substantial portion of σ tends

---

[6] The same value of δ is used for the intercept and the regression coefficient. Therefore, the Gibbs sampler will use the following conditional posterior density for δ: $p(\delta \mid \beta_0, \gamma) \propto \frac{1}{\delta^{Tk+1}} \exp\{-\sum_{i=1}^{k}\sum_{t=1}^{T}(\beta_{it} - \beta_{it-1})^2/2\delta^2\}$, a special form of equation (21).

[7] Using $\alpha_{-1} = 0$ and $\beta_{-1} = 1$ for starting values, I generated 21 observations and discarded the first observation.

to be absorbed into δ. In other words, a part of a transient shock to y is estimated as a change in the coefficient vector β which can vary over time. As expected, the Gibbs sampler overestimated δ and underestimated σ, implying that the importance of the persistent effects was overestimated.[8] Note that η_t in the TVP model represents the persistent shocks to the system, while u_t represents the transient shocks. The proportion of the variance of η_t, i. e., δ^2, in the total variance, i. e., $(\sigma^2+\delta^2)$, had a mean of 0.915 and a standard deviation of 0.185, ranging between 0.019 and 1. Remember that the model used $\sigma=1$ and $\delta=1$ for the data generation, with the proportion equal to 0.5. More precise inference about δ and σ would be possible when we use multiple observations in each time period, because the availability of time series of cross-sections provides much more scope for investigating the nature of time-varying parameters (Harvey(1978)). The Gibbs sampling procedure is extended for the TVP model with multiple observations in the next section.

Ⅳ. GIBBS SAMPLER FOR MODELS WITH MULTIPLE OBSERVATIONS IN EACH PERIOD

Introducing subscript $j(=1, 2, \cdots, M)$ to represent individual data-generating units, we express the TVP model with multiple observations as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \gamma + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix} \tag{24}$$

$$A\gamma = J\beta_0 + \eta, \qquad \eta \sim N(0, I_T \otimes \Delta) \tag{25}$$

where $u_j \sim N(0, \sigma^2 I_T)$ and u_j's are mutually independent for all j.

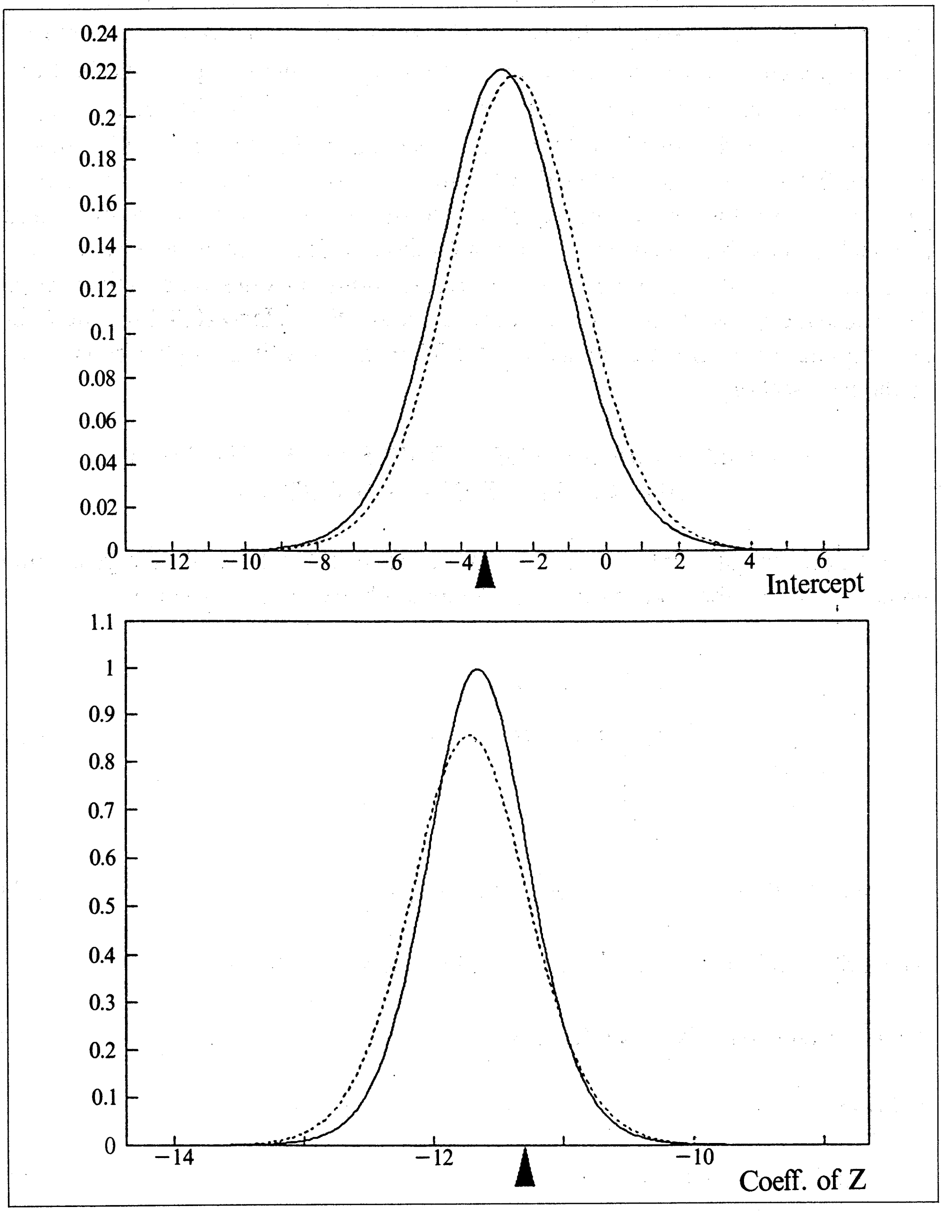
4.1 Full Conditional Posterior Densities

1. Priors:

$$p(\beta_0 \mid \sigma, \Delta) \propto \text{constant} \tag{26}$$

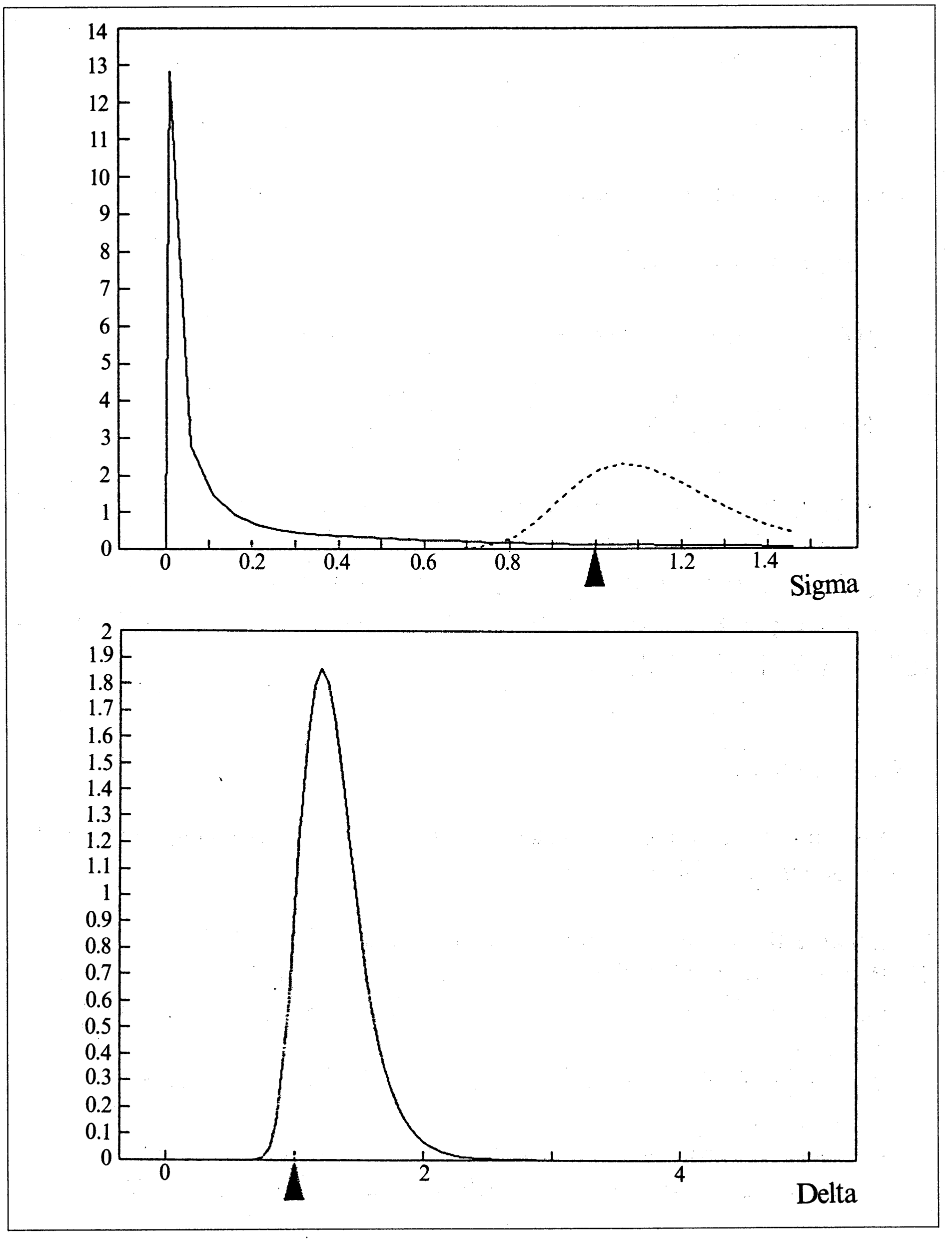
[8] It is not the problem with the Gibbs sampler, but with the nature of the model itself. However, precise estimates can be obtained using multiple observations in each period, which is discussed in the next section.

[Figure 1] Marginal and Conditional Posterior Densities for β_{cT}, β_{zT}, σ and δ (Example 1)



Solid lines are the plots of the Gibbs-estimated marginal posterior densities, and dotted lines are those of the conditional posterior densities evaluated at the value of the hyperparameter used for data generation, i.e., $\delta = 1$. Arrows show the values used in generating the data.

[Figure 1] (continued)



$$p(\gamma \mid \beta_0, \sigma, \Delta) \sim N\{A^{-1}J\beta_0,\ A^{-1}J\beta_0,\ A^{-1}(I_T \otimes \Delta)A^{-1\prime}\} \tag{27}$$

$$p(\sigma \mid \gamma) \propto 1/\sigma \tag{28}$$

$$p(\Delta) \propto |\Delta|^{-1/2} \tag{29}$$

2. Conditional Posterior Densities:

$$p(\beta_0 \mid \sigma, \Delta, D_T) \sim N(\hat{\beta}_0, V) \tag{30}$$

$$p(\gamma \mid \beta_0, \sigma, \Delta, D_T) \sim N(\hat{\gamma}, W) \tag{31}$$

$$p(\sigma \mid \gamma, D_T) \propto \frac{1}{\sigma^{MT+1}} \exp\{-\sum_{j=1}^{M} (y_j - X_j\gamma)'(y_j - X_j\gamma)/2\sigma^2\} \tag{32}$$

$$p(\delta_i \mid \beta_0, \gamma) \propto \frac{1}{\delta_i^{T+1}} \exp\{-\sum_{t=1}^{T} (\beta_{it} - \beta_{it-1})^2/2\delta_i^2\} \text{ for } i=1, \cdots, k \tag{33}$$

where $\hat{\beta}_0 = V \times (\Sigma_{j=1}^{M} J'A^{-1\prime}X_j'Q_j^{-1}y_j)$, $Q_j = (\sigma^2 I + X_j A^{-1}(I_T \otimes \Delta)A^{-1\prime}X_j')$, $V = \{J' A^{-1\prime}(\Sigma_{j=1}^{M} X_j'Q_j^{-1}X_j)A^{-1}J\}^{-1}$, $\hat{\gamma} = W \times \{\Sigma_{j=1}^{M} X_j'y_j/\sigma^2 + A'(I_T \otimes \Delta)^{-1}J\beta_0\}$, and $W = \{\Sigma_{j=1}^{M} X_j'X_j/\sigma^2 + A'(I_T \otimes \Delta)^{-1}A\}^{-1}$.

3. Gibbs Sampler:

With the full conditional densities derived above, a Gibbs sampling procedure can be designed as follows:

(1) Initialize the values of σ and Δ, e. g., $\sigma^{(0)}$and $\Delta^{(0)}$.
(2) Repeat steps (a)−(d) $K+N$ times. For $i=1, \cdots, K+N$,
 (a) Sample $\beta_0^{(i)}$ from $p(\beta_0 \mid \sigma^{(i-1)}, \Delta^{(i-1)}, D_T)$ which is a multivariate normal, as in (30)
 (b) Sample $\gamma^{(i)}$ from $p(\gamma \mid \beta_0^{(i)}, \sigma^{(i-1)}, \Delta^{(i-1)}, D_T)$ which is a multivariate normal, as in (31).
 (c) Sample $\sigma^{(i)}$ from $p(\sigma \mid \gamma^{(i)}, D_T)$ which is an IG density, as in (32)
 (d) Sample $\Delta^{(i)}$ from $p(\Delta \mid \beta_0^{(i)}, \gamma^{(i)})$ which is an IG density, as in (33).

4.2 Example 2

The TVP model with multiple observations assumes the same regression coefficients for all observations in the same period. Therefore, two more observations for each period were generated using the same values of β_t as in Example 1. Taking the values of β_t from Example 1, I generated y_t using (22) with $\sigma=1$ and z_t being drawn independently from a uniform distribution with an interval (0, 10).

It is shown in Figure 2 that the Gibbs-estimated marginal posterior densities for β_{cT} and β_{zT} are very close to the conditional posterior densities evaluated at $\delta=1$, the value used for data generation. Comparing with Figure 1, we can see that use of multiple observations produced more precise estimates of the marginal posterior densities for β_{cT} and β_{zT} than use of single observations did. Further, Figure 2 shows that use of multiple observations produced accurate estimates for σ and δ. The Gibbs-estimated densities for σ and δ are skewed to the right and look like inverted gamma densities. And their modes are 1.1 for both σ and δ, while the value used for data generation is 1 for both parameters. The proportion of the variance of η_t, i. e., δ^2, in the total variance, i. e., $\sigma^2+\delta^2$, in the total variance, i. e., $(\sigma^2+\delta^2)$ had a mean of 0.503 and a standard deviation of 0.104, ranging between 0.151 and 0.866. Remember that the model used $\sigma=1$ and $\delta=1$ for the data generation, with the proportion equal to 0.5. It suggests that the multiple-observation model be used if we are interested in estimating the relative importance of permanent effects as compared to the one of transient effects.

V. PREDICTION

The evaluation of marginal predictive densities for the TVP model involves integrations which are impossible to perform analytically for the same reasons explained in Section 2. By employing the Gibbs sampler, the marginal predictive density for y_{T+1}, $p(y_{T+1} \mid x_{T+1}, D_T)$, would be approximated by an average of conditional predictive densities, i. e., $\hat{p}(y_{T+1} \mid x_{T+1}, D_T)=\Sigma_{i=1}^{N} p(y_{T+1} \mid \beta_T^{(i)}, \sigma^{(i)}, \Delta^{(i)}, x_{T+1}, D_T)/N$ where $p(y_{T+1} \mid \beta_T^{(i)}, \sigma^{(i)}, \Delta^{(i)}, x_{T+1}, D_T) \sim N\{x_{T+1}'\beta_T^{(i)}, (\sigma^{(i)2}+x_{T+1}'\Delta^{(i)}x_{T+1})\}$ and $\beta_T^{(i)}$, $\sigma^{(i)}$ and $\Delta^{(i)}$ are random draws via the Gibbs sampler.[9]

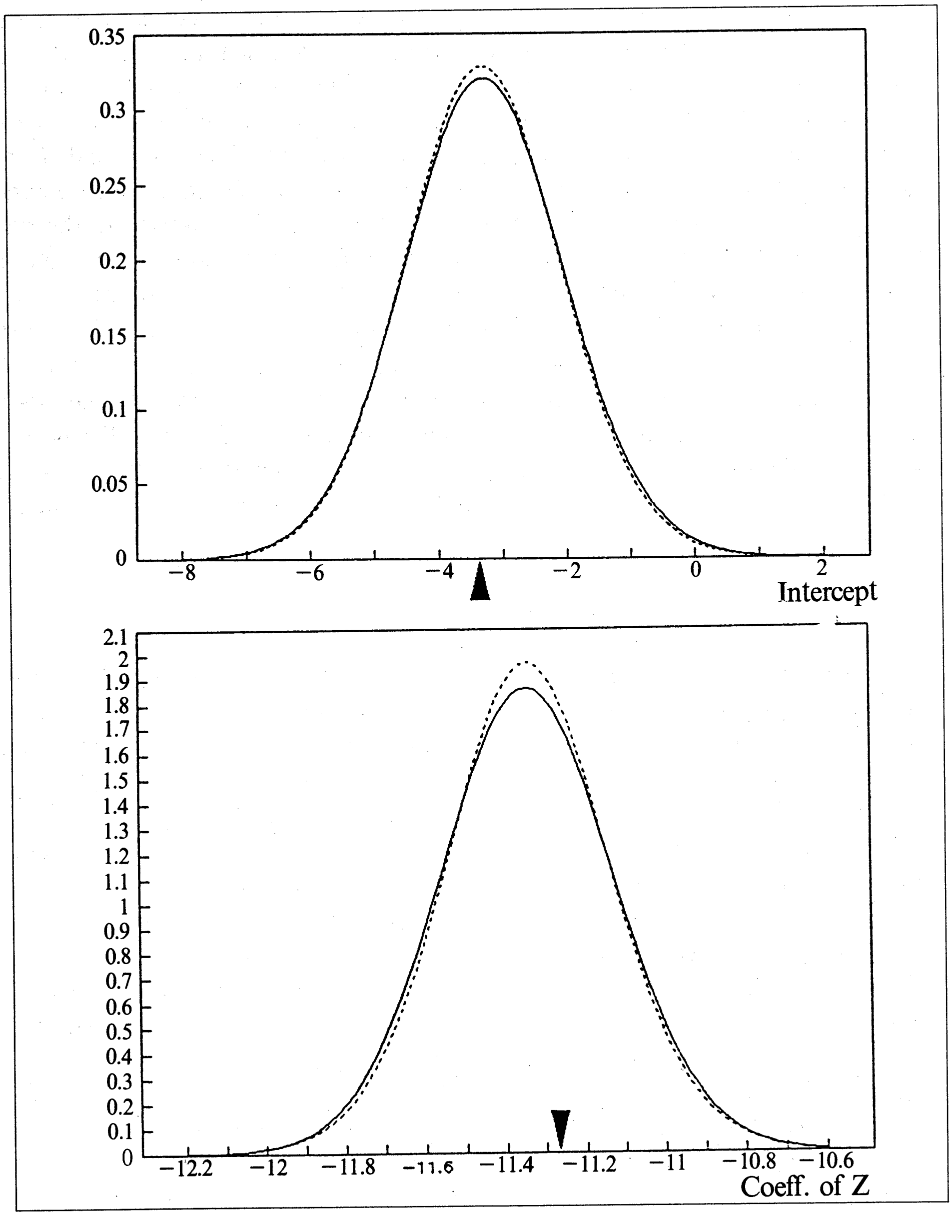
Figure 3 shows marginal and conditional predictive densities when $x_{T+1}=1$ in Examples 1 and 2. The marginal predictive densities are computed using random draws via the Gibbs sampler: i. e., $p(y_{T+1} \mid x_{T+1}=1, D_T)$ is approximated by $\Sigma_{i=1}^{19,000} p(y_{T+1} \mid \beta_T^{(i)}, \sigma^{(i)}, \delta^{(i)}, x_{T+1}=1, D_T)/19{,}000$ where $\beta_T^{(i)}$, $\sigma^{(i)}$ and $\delta^{(i)}$ are the Gibbs draws. As expected from the Gibbs-estimates of the posterior densities, the Gibbs sampler produced predictive densities which are close to the conditional predictive densities evaluated at $\delta=1$, the value used for data generation.

It is well known that the marginal predictive mean is the optimal forecast for squared error loss functions such as $L(y_{T+1}, \tilde{y}_{T+1})=c(y_{T+1}-\tilde{y}_{T+1})^2$ where $c>0$. The marginal predictive mean can be calculated easily by (34).

---

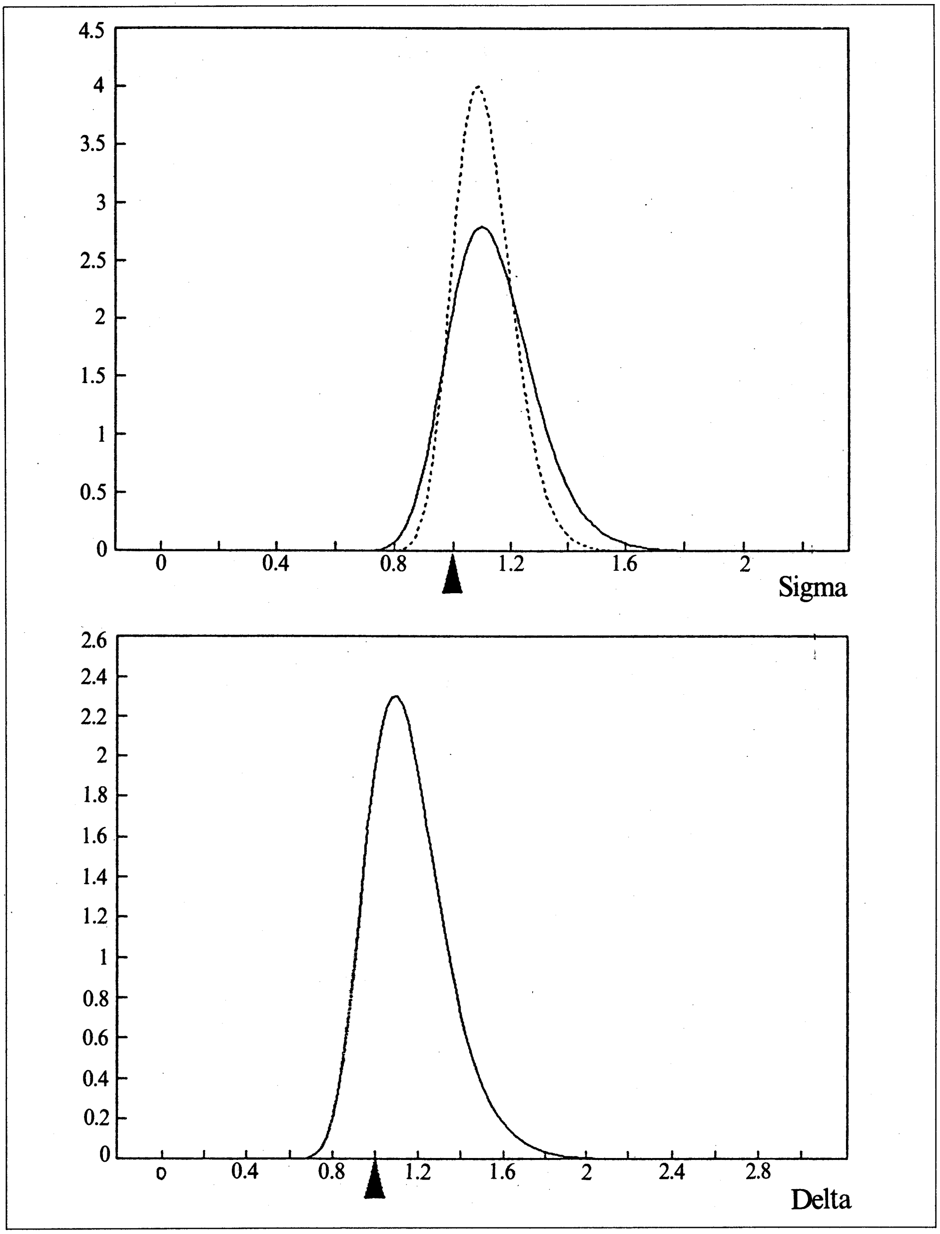
[9] Since $y_{T+1}=x_{T+1}'\beta_{T+1}+u_{T+1}=x_{T+1}'\beta_T+(x_{T+1}'\eta_{T+1}+u_{T+1})$, it follows that $y_{T+1} \sim N\{x_{T+1}'\beta_T, (\sigma^2+x_{T+1}'\Delta x_{T+1})\}$ when β_T, σ, Δ and x_{T+1} are given.

[Figure 2] Marginal and Conditional Posterior Densities for β_{cT}, β_{zT}, σ and δ (Example 2)

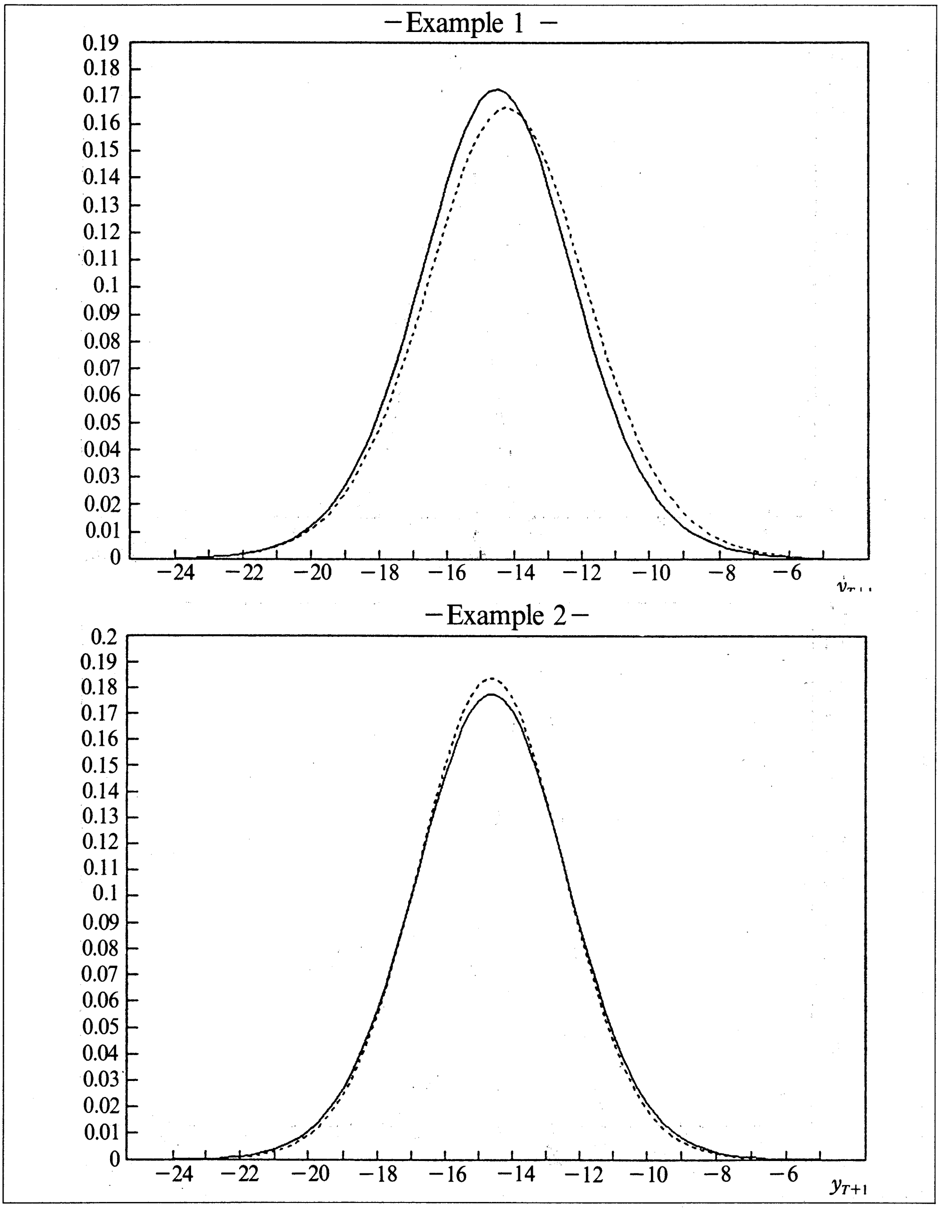


Solid lines are the plots of the Gibbs-estimated marginal posterior densities, and dotted lines are those of the conditional posterior densities evaluated at the value of the hyperparameter used for data generation, i.e., $\delta = 1$. Arrows show the values used in generating the data.

[Figure 2] (continued)



[Figure 3] Marginal and Conditional Predictive Densities for y_{T+1}, when $x_{T+1} = 1$



Solid lines are the plots of the Gibbs-estimated marginal predictive densities, and dotted lines are those of the conditional predictive densities evaluated at the value of the hyperparameter used for data generation, i. e., $\delta = 1$.

$$\hat{E}(y_{T+1} \mid x_{T+1}=1, D_T) = \frac{1}{N} \sum_{i=1}^{N} E(y_{T+1} \mid \beta_T^{(i)}, \sigma^{(i)}, \delta^{(i)}, x_{T+1}, D_T)$$
$$= \frac{1}{N} \sum_{i=1}^{N} x_{T+1}' \beta_T^{(i)} \tag{34}$$

Thus, we would be able to approximate the optimal forecast via the Gibbs sampler. For the cases of Examples 1 and 2, the optimal forecasts of y_{T+1} when $x_{T+1} = 1$ are -14.52 and -14.62, respectively.

The Gibbs sampler is also useful in making optimal predictions for other loss functions. For example, with an absolute error loss function $L(y_{T+1}, \tilde{y}_{T+1}) = | y_{T+1} - \tilde{y}_{T+1} |$, the optimal forecast might be approximated by the median of random draws from $p(y_{T+1} \mid \beta_T^{(i)}, \sigma^{(i)}, \delta^{(i)}, x_{T+1}=1, D_T)$. The medians of 19,000 draws from $p(y_{T+1} \mid \beta_T^{(i)}, \sigma^{(i)}, \delta^{(i)}, x_{T+1}=1, D_T)$ were -14.53 in Example 1 and -14.59 in Example 2.

Further, the Gibbs sampling procedures can be easily extended to multiperiod-ahead forecasting problems. Suppose we want a k-period-ahead forecast from an autoregressive model of order one, $y_t = x_t' \beta_t + u_t$ where $x_t' = (1\ y_{t-1})$. The joint predictive density for $(y_{T+k}, \cdots, y_{T+1})$ evaluated in period T is:

$$p(y_{T+k}, \cdots, y_{T+1} \mid D_T) = \iiint p(y_{T+k} \mid y_{T+k-1}, \cdots, y_{T+1}, \beta_T, \sigma, \delta, D_T) \times \cdots$$
$$\times\ p(y_{T+1} \mid \beta_T, \sigma, \delta, D_T) \times p(\beta_T, \sigma, \delta \mid D_T)\, d\beta_T\, d\sigma\, d\delta \tag{35}$$

where $p(y_{T+j} \mid y_{T+j-1}, \cdots, y_{T+1}, \beta_T, \sigma, \delta, D_T) \sim N\{x_{T+j}' \beta_T, (\sigma^2 + j\delta^2 x_{T+j}' x_{T+j})\}$ since $y_{T+j} = x_{T+j}' \beta_{T+j} + u_{T+j} = x_{T+j}' \beta_T + x_{T+j}' (\eta_{T+j} + \cdots, \eta_{T+1}) + u_{T+j}$. Using Gibbs-drawn values of $(\beta_T, \sigma, \delta)$ from $(\beta_T, \sigma, \delta \mid D_T)$, we can sample y_{T+1} from $p(y_{T+1} \mid \beta_T, \sigma, \delta, D_T)$, y_{T+2} from $p(y_{T+2} \mid y_{T+1}, \beta_T, \sigma, \delta, D_T)$, and so on. Then, the k-period-ahead predictive density may be approximated by

$$\hat{p}(y_{T+k} \mid D_T) = \frac{1}{N} \sum_{i=1}^{N} p(y_{T+k} \mid y_{T+k-1}^{(i)}, \beta_T^{(i)}, \sigma^{(i)}, \delta^{(i)}, D_T) \tag{36}$$

where $y_{T+k-1}^{(i)}$, $\beta_T^{(i)}$, $\sigma^{(i)}$ and $\delta^{(i)}$ are Gibbs-drawn values.

VI. CONCLUSIONS

This study has developed Gibbs sampling procedures to estimate the marginal posterior densities for regression models with time-varying parameters. The Gibbs sampler has performed satisfactorily for simulated data, in the sense that the Gibbs sampler converged and the Gibbs-estimated marginal posterior densities of parameters were close to their conditional posterior densities evaluated at the values of hyperparameters used for data generation.

In order to consider the effects of probable specification errors on the performance of the Gibbs sampler, several simulated data were generated using large variances, e. g., using $\sigma^2 = 25$ instead of $\sigma^2 = 1$ for the model in Example 1. The Gibbs sampler has also performed well for these data.

APPENDIX

DERIVATION OF THE FULL CONDITIONAL POSTERIOR DENSITIES

1. TVP Model with a Single Observation in Each Period

The TVP model considered in this study may be expressed using matrix notations as follows:

$$y = X\gamma + u, \qquad u \sim N(0, \sigma^2 I_T) \tag{37}$$

$$A\gamma = J\beta_0 + \eta, \qquad \eta \sim N(0, I_T \otimes \Delta) \tag{38}$$

where the definitions of y, X, γ, u, A, J and η are given in the text.

1) $p(\beta_0 \mid \sigma, \Delta, D_T)$: Equation (38) can be rewritten as follows:

$$\gamma = A^{-1} J\beta_0 + A^{-1}\eta \tag{39}$$

Substituting (39) into (37), we obtain

$$\begin{aligned} y &= XA^{-1}J\beta_0 + (XA^{-1}\eta + u) \\ &= X^* \beta_0 + u^* \end{aligned} \tag{40}$$

where $u^* \sim N(0, Q)$ and $Q = (\sigma^2 I + XA^{-1}(I_T \otimes \Delta)A^{-1\prime}X')$. With a diffuse prior for β_0 given σ and i. e., $p(\beta_0 \mid \sigma, \Delta) \propto$ constant, the conditional posterior density for β_0 is

$$p(\beta_0 \mid \sigma, \Delta, D_T) \sim N(\hat{\beta}_0, V) \tag{41}$$

where $\hat{\beta}_0 = V \times (X^{*\prime} Q^{-1} y) = V \times (J' A^{-1\prime} X' Q^{-1} y)$ and $V = (J' A^{-1\prime} X' Q^{-1} X A^{-1} J)^{-1}$.

2) $p(\gamma \mid \beta_0, \sigma, \Delta, D_T)$: From (39), we know that

$$p(\gamma \mid \beta_0, \sigma, \Delta) \sim N\{A^{-1}J\beta_0,\ A^{-1}(I_T \otimes \Delta)A^{-1\prime}\} \tag{42}$$

On combining (42) with the likelihood function for (37) via Bayes' Theorem, the conditional posterior density for γ is

$$p(\gamma \mid \beta_0, \sigma, \Delta, D_T) \sim N(\hat{\gamma}, W) \tag{43}$$

where $\hat{\gamma} = W \times (X'y/\sigma^2 + A'(I_T \otimes \Delta)^{-1}J\beta_0\}$ and $W = \{X'X/\sigma^2 + A'(I_T \otimes \Delta)^{-1}A\}^{-1}$.

3) $p(\sigma \mid \gamma, D_T)$:

$$\begin{aligned} p(\sigma \mid \gamma, D_T) &\sim p(\sigma \mid \gamma) \times p(y \mid \sigma, \gamma) \\ &\sim p(\sigma \mid \gamma) \times \frac{1}{\sigma^T} \exp\{-(y - X\gamma)'(y - X\gamma)/2\sigma^2\} \end{aligned} \tag{44}$$

Employing a diffuse prior for σ given γ, i. e., $p(\sigma \mid \gamma) \propto 1/\sigma$, we obtain

$$\begin{aligned} p(\sigma \mid \gamma, D_T) &\propto \frac{1}{\sigma^{T+1}} \exp\{-(y - X\gamma)'(y - X\gamma)/2\sigma^2\} \\ &\sim \text{Inverted Gamma} \end{aligned} \tag{45}$$

4) $p(\Delta \mid \beta_0, \gamma)$: Given β_0 and γ, the information about Δ is contained in (38) only. Therefore,

$$\begin{aligned} p(\Delta \mid \beta_0, \gamma) &\propto p(\Delta) \times p(\beta_0, \gamma \mid \Delta) \\ &\propto p(\Delta) \times \frac{1}{|\Delta|^{T/2}} \exp\{-(A\gamma - J\beta_0)'(I_T \otimes \Delta)^{-1}(A\gamma - J\beta_0)/2\} \\ &\propto p(\Delta) \times \frac{1}{|\Delta|^{T/2}} \exp\{-\sum_{t=1}^{T}(\beta_t - \beta_{t-1})'\Delta^{-1}(\beta_t - \beta_{t-1})/2\} \\ &\propto p(\delta_1, \cdots, \delta_k) \times \Pi_{1=1}^{k} \left[\frac{1}{\delta_i^T} \exp\{-\sum_{t=1}^{T}(\beta_{it} - \beta_{it-1})^2/2\delta_i^2\} \right] \end{aligned}$$

where β_{it} is the ith element of β_t. Employing a diffuse prior for $(\delta_1, \cdots, \delta_k)$, i. e., $p(\delta_1, \cdots, \delta_k) \propto 1/(\delta_1 \cdots \delta_k)$,[10] we obtain the following posterior density for δ_i ($i = 1, \cdots, k$):

[10] In Liu and Hanssens (1981), a locally uniform prior was used for the hyperparameters, claiming that precise choice of a prior for the hyperparameter is not critical for a moderate-sized sample.

$$p(\delta_i \mid \beta_0, \gamma) \propto \frac{1}{\delta_i^{T+1}} \exp\{-\sum_{t=1}^{T}(\beta_{it} - \beta_{it-1})^2/2\delta_i^2\}$$
$$\sim \text{Inverted Gamma} \tag{47}$$

In a special case of $\delta_1 = \cdots = \delta_k = \delta$, the following posterior density for δ is obtained using a diffuse prior $p(\delta) \propto 1/\delta$:

$$p(\delta \mid \beta_0, \gamma) \propto \frac{1}{\delta^{Tk+1}} \exp\{-\sum_{i=1}^{k}\sum_{t=1}^{T}(\beta_{it} - \beta_{it-1})^2/2\delta^2\}$$
$$\sim \text{Inverted Gamma} \tag{48}$$

2. TVP Model with Multiple Observations in Each Period

The following model has been considered in this study:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} \gamma + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix} \tag{49}$$

$$A\gamma = J\beta_0 + \eta, \qquad \eta \sim N(0, I_T \otimes \Delta) \tag{50}$$

where $j(=1, 2, \cdots, M)$ represents data-generating units, and u_j's are mutually independent for all j with $u_j \sim N(0, \sigma^2 I_T)$.

1) $p(\beta_0 \mid \sigma, \Delta, D_T)$: Equation (50) can be rewritten as follows:

$$\gamma = A^{-1}J\beta_0 + A^{-1}\eta \tag{51}$$

Substituting (51) into (49), we obtain

$$y_j = X_j A^{-1} J\beta_0 + (X_j A^{-1}\eta + u_j)$$
$$= X_j^* \beta_0 + u_j^* \tag{52}$$

where $u_j^* \sim N(0, Q_j)$ and $Q_j = \sigma^2 I + X_j A^{-1}(I_T \otimes \Delta)A^{-1\prime}X_j'$. With a diffuse prior for β_0 given σ and Δ, i. e., $p(\beta_0 \mid \sigma, \Delta) \propto$ constant, the conditional posterior density for β_0 is

$$p(\beta_0 \mid \sigma, \Delta, D_T) \sim N(\hat{\beta}_0, V) \tag{53}$$

where $\hat{\beta}_0=V\times(\Sigma_{j=1}^{M} X_j^{*\prime}Q_j^{-1}y_j)=V\times(\Sigma_{j=1}^{M} J'A^{-1\prime}X_j'Q_j^{-1}y_j)$ and $V=\{J'A^{-1}(\Sigma_{j=1}^{M} X_j'Q_j^{-1}X_j)A^{-1}J\}^{-1}$.

2) $p(\gamma\mid\beta_0,\sigma,\Delta,D_T)$: From (51), we know that

$$p(\gamma\mid\beta_0,\sigma,\Delta)\sim N\{A^{-1}J\beta_0,\ A^{-1}(I_T\otimes\Delta)A^{-1\prime}\} \tag{54}$$

On combining (54) with the likelihood function for (49) via Bayes' Theorem, the conditional posterior density for γ is

$$p(\gamma\mid\beta_0,\sigma,\Delta,D_T)\sim N(\hat{\gamma},W) \tag{55}$$

where $\hat{\gamma}=W\times\{\Sigma_{j=1}^{M} X_j'\,y_j/\sigma^2+A'(I_T\otimes\Delta)^{-1}J\beta_0\}$ and $W=\{\Sigma_{j=1}^{M} X_j'\,X_j/\sigma^2+A'(I_T\otimes\Delta)^{-1}A\}^{-1}$.

3) $p(\sigma\mid\gamma,D_T)$:

$$\begin{aligned}p(\sigma\mid\gamma,D_T)&\sim p(\sigma\mid\gamma)\times p(y\mid\sigma,\gamma)\\&\sim p(\sigma\mid\gamma)\times\frac{1}{\sigma^{MT}}\ \exp\{-\sum_{j=1}^{M}(y_j-X_j\gamma)'(y_j-X_j\gamma)/2\sigma^2\}\end{aligned} \tag{56}$$

Employing a diffuse prior for σ given γ, i. e., $p(\sigma\mid\gamma)\propto 1/\sigma$, we obtain

$$\begin{aligned}p(\sigma\mid\gamma,D_T)&\propto\frac{1}{\sigma^{MT+1}}\ \exp\{-\sum_{j=1}^{M}(y_j-X_j\gamma)'(y_j-X_j\gamma)/2\sigma^2\}\\&\sim\text{Inverted Gamma}\end{aligned} \tag{57}$$

4) $p(\Delta\mid\beta_0,\gamma)$: Given β_0 and γ, the information about Δ is contained in (50) only. Therefore,

$$\begin{aligned}p(\Delta\mid\beta_0,\gamma)&\propto p(\Delta)\times p(\beta_0,\gamma\mid\Delta)\\&\propto p(\Delta)\times\frac{1}{|\Delta|^{T/2}}\ \exp\{-(A\gamma-J\beta_0)'(I_T\otimes\Delta)^{-1}(A\gamma-J\beta_0)/2\}\\&\propto p(\Delta)\times\frac{1}{|\Delta|^{T/2}}\ \exp\{-\sum_{t=1}^{T}(\beta_t-\beta_{t-1})'\Delta^{-1}(\beta_t-\beta_{t-1})/2\}\\&\propto p(\delta_1,\cdots,\delta_k)\times\Pi_{i=1}^{k}\left[\frac{1}{\delta_i^T}\ \exp\{-\sum_{t=1}^{T}(\beta_{it}-\beta_{it-1})^2/2\delta_i^2\}\right]\end{aligned}$$

Employing a diffuse prior for $(\delta_1,\cdots,\delta_k)$, i. e., $p(\delta_1,\cdots,\delta_k)\propto 1/(\delta_1\cdots\delta_k)$, we obtain the following posterior density for δ_i $(i=1,\cdots,k)$:

$$p(\delta_i \mid \beta_0, \gamma) \propto \frac{1}{\delta_i^{T+1}} \exp\{-\sum_{t=1}^{T}(\beta_{it} - \beta_{it-1})^2/2\delta_i^2\}$$
$$\sim \text{Inverted Gamma} \tag{59}$$

In a special case of $\delta_1 = \cdots = \delta_k = \delta$, the following posterior density for δ is obtained using a diffuse prior $p(\delta) \propto 1/\delta$:

$$p(\delta \mid \beta_0, \gamma) \propto \frac{1}{\delta^{Tk+1}} \exp\{-\sum_{i=1}^{k}\sum_{t=1}^{T}(\beta_{it} - \beta_{it-1})^2/2\delta^2\}$$
$$\sim \text{Inverted Gamma} \tag{60}$$

REFERENCES

Blattberg, R. C., and E. I. George(1991), "Shrinkage Estimation of Price and Promotional Elasticities: Seemingly Unrelated Equations", *Journal of the American Statistical Association*, 86, 304-315.

Carlin, B. P., N. G. Polson, and D. S. Stoffer(1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling", *Journal of the American Statistical Association*, 87, 493-500.

Casella, G, and E. I. George(1992), "Explaining the Gibbs Sampler", *American Statistician*, 46, 167-174.

Chib, S.(1993), "Bayes Regression with Autoregressive Errors: A Gibbs Sampling Approach", *Journal of Econometrics*, 58, 275-294.

Chib, S., R. C. Tiwari, and S. R. Jammalamadaka(1990), "On Bayes Smoothing in a Time Varying Regression Model", in S. Geisser, J. Hodges, S. J. Press and A. Zellner, eds., *Bayesian Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard*, North-Holland, Amsterdam.

Cooley, T. F., and E. C. Prescott(1976), "Estimation in the Presence of Stochastic Parameter Variation", *Econometrica*, 44, 167-184.

Doan, T., R. Litterman, and C. Sims(1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Reviews*, 3, 1-144.

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.

Geman, S., and D. Geman(1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Geweke, J.(1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation 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, 169-188.

Harvey, A. C.(1978), "The Estimation of Time-Varying Parameters from Panel Data"(with discussion), *Annales de L'Insee*, 30/31, 203-226.

Johnson, V. E.(1992), "A Technique for Estimating Marginal Posterior Densities in Hierarchical Models Using Mixtures of Conditional Densities", *Journal of the American Statistical Association*, 87, 852-860.

Liu, L. M., and D. M. Hanssens(1981), "A Bayesian Approach to Time-Varying Cross-Sectional Regression Models", *Journal of Econometrics*, 15, 341-356.

McCulloch, R. E., and P. E. Rossi(1994), "An Exact Likelihood Analysis of the Multinomial Probit Model", *Journal of Econometrics*, 64, 207-240.

Tanner, M. A., and W. H. Wong(1987), "The Calculation of Posterior Distributions by Data Augmentation"(with discussion), *Journal of the American Sta-*

tistical Association, 82, 528-550.

Tierney, L.(1994), "Markov Chains for Exploring Posterior Distributions", (with discussion), *Annals of Statistics*, 22, 1701-1762.

Tierney, L., and J. B. Kadane(1986), "Accurate Approximations for Posterior Moments and Marginal Densities", *Journal of the American Statistical Association*, 81, 82-86.

Zellner, A., and C. Min(1995), "Gibbs Sampler Convergence Criteria", *Journal of the American Statistical Association*, 90, 921-927.

Zellner, A., and P. Rossi(1982), "Bayesian Analysis of Dichotomous Quantal Response Models", *Journal of Econometrics*, 25, 365-393.