STUDIES IN BAYESIAN
ECONOMETRICS AND STATISTICS

7

sbes

BAYESIAN AND LIKELIHOOD METHODS IN STATISTICS AND ECONOMETRICS

Essays in Honor of George A. Barnard

Edited by
SEYMOUR GEISSER
JAMES S. HODGES
S. JAMES PRESS
ARNOLD ZELLNER

North-Holland

Bayesian and Likelihood Methods in Statistics and Econometrics
S. Geisser, J.S. Hodges, S.J. Press and A. Zellner (Editors)

ON BAYES SMOOTHING IN A TIME VARYING REGRESSION MODEL

Siddhartha Chib, S. Rao Jammalamadaka, and Ram C. Tiwari
University of Missouri, University of California, and University of North Carolina

Abstract

This paper considers the Bayes smoothing problem for a time varying regression model with unknown noise variance parameters. Using the prediction error decomposition, an approach is developed that provides the exact pdf of the regression coefficients and the noise variance parameters. In addition, the paper provides a discussion, and an example, of Monte Carlo numerical integration procedures that can be used to generate estimates of the moments of the smoothed coefficients.

1. Introduction

In this paper we are concerned with the smoothing problem in a bivariate time varying regression (TVR) model when the variance parameters of the error terms are unknown. The model we consider is the following

$$y_t = x_t\beta_t + \epsilon_t, \qquad \beta_t = g_t\beta_{t-1} + \eta_t \qquad (t = 1, \ldots, n) \tag{1}$$

where ϵ_t are iid $N(0, \sigma^2)$, the η_t are iid $N(0, \lambda\sigma^2)$, $(\lambda, \sigma^2 > 0)$, and ϵ_t and η_s are uncorrelated for all t and s. We assume that $\{g_t\}$ is a fixed, nonstochastic sequence of scalars. We further assume that the two variance parameters, σ^2 and λ, are unknown.

An important issue in such models has been the choice of β_0, the initial value at start up time 0. We assume that the parameter β_0 is unknown with its uncertainty expressed by a normal distribution $N(\hat{\beta}_{0|0}, \sigma^2 R_{0|0})$, where the hyperparameters, $\hat{\beta}_{0|0}$ and $R_{0|0}$, are known.

The smoothing problem for models similar to (1) above have been analyzed by several authors including Anderson and Moore (1979), Chow (1983), and Engle and Watson (1988), using classical methods, and by Broemeling (1985) and Broemeling, Diaz and Yusoff (1985) using Bayesian methods.

In this paper, we provide a Bayes solution to the smoothing problem. Essentially the problem of smoothing is concerned with making inferences about the $n \times 1$ regresof total probability as:

$$\begin{aligned} \ell(\theta|y^n) &\propto f(y_1, \ldots, y_n|\theta) \\ &\propto f(y_1|\theta) f(y_2|y^1, \theta) \ldots f(y_n|y^{n-1}, \theta) \\ &\propto \left(\frac{1}{\sigma^2}\right)^{n/2} \prod_{t=1}^{n} [\tilde{f}_{t|t-1,\lambda}]^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2} \sum_{t=1}^{n} (y_t - \hat{y}_{t|t-1,\lambda})^2 / \tilde{f}_{t|t-1,\lambda}} \end{aligned} \tag{9}$$

where $\tilde{f}_{t|t-1,\lambda}$ and $\hat{y}_{t|t-1,\lambda}$ are produced by the KF, for each value of λ. It is clear from (9) that if we let the prediction error be denoted by $\tilde{y}_{t|t-1,\lambda}$ where

$$\tilde{y}_{t|t-1,\lambda} = y_t - \hat{y}_{t|t-1,\lambda}, \tag{10}$$

then the likelihood function is produced in terms of the distribution of the prediction errors. It is for this reason that (9) is referred to as the prediction error decomposition.

REMARK 1: In classical analysis, estimation of the parameter, θ, is accomplished by maximizing the likelihood function. The usual procedure is to fix β_0, and then adopt a Newton-Raphson algorithm to successively update the parameter values until the log-likelihood is maximized. Unless the likelihood function is well behaved, such a procedure can sometimes locate the wrong maxima, especially when the sample size is small.

3. The Smoothing Problem

3.1 Posterior of θ

In order to make inferences about the vector of regression parameters $\beta = (\beta_1, \ldots, \beta_n) : n \times 1$ when the variance parameter θ is unknown, we need to derive the marginal posterior of θ and the conditional posterior of β given y^n and θ. Then the unconditional posterior pdf of β is obtained by integrating out θ from the latter distribution, i.e.,

$$\pi(\beta|y^n) = \int_\theta \pi(\beta|y^n, \theta)\pi(\theta|y^n) d\theta \tag{11}$$

where $\pi(\theta|y^n)$ is the posterior pdf of θ specified in (2).

It is the purpose of this subsection to show that the posterior pdf of θ in (11) can be derived using the PED just described. It is, of course, necessary to specify a useful prior for the variance parameter θ that would allow the integration in (11) to be performed conveniently. It turns out that an informative Inverted Gamma pdf of σ^2 with known hyperparameters combines nicely with the likelihood function. However, as far as the second variance parameter, λ, is concerned no useful prior seems available. The best practical strategy appears to be to use some locally uniform prior pdf of λ. We prefer the latter Bayesian method to the alternative of specifying λ and ignoring the prior uncertainty about its value.

In the light of the previous comments, we assume that σ^2 and λ are *a priori* independent, with the prior of σ^2 an informative, Inverted Gamma$(\frac{v^*}{2}, \frac{\delta^*}{2})$ density with shape parameter $\frac{v^*}{2}$ and scale parameter $\frac{\delta^*}{2}$, and that of λ specified in general form as $\pi(\lambda)$, where $\pi(\lambda)$ is possibly some locally uniform pdf. Thus, the joint prior density of (σ^2, λ) is given by

$$\pi(\sigma^2, \lambda) \; \propto \; (\sigma^2)^{\frac{v^*}{2}+1} e^{-\frac{\delta^*}{2\sigma^2}} \cdot \pi(\lambda), \quad \sigma^2, \lambda > 0, \tag{12}$$

v^* and $\delta^* > 0$ known. If we combine this prior with the observed likelihood function in (9) and apply (2), then using (10) the posterior pdf of (σ^2, λ) is

$$\pi(\sigma^2, \lambda | y^n) \; \propto \; \pi(\sigma^2, \lambda) \ell_{obs}(\sigma^2, \lambda | y^n)$$

$$\propto \; \pi(\lambda) \left[\prod_{t=1}^{n} (\tilde{f}_{t|t-1,\lambda})^{-\frac{1}{2}} \right] (1/\sigma^2)^{\frac{v^*+n}{2}+1} e^{-\frac{1}{2\sigma^2}(\delta^* + \sum_{t=1}^{n} (\tilde{y}_{t|t-1,\lambda})^2 / \tilde{f}_{t|t-1,\lambda})} \tag{13}$$

where the normalizing constant is $1/\int \pi(\theta)\ell_{obs}(\theta|y^n)d\theta$. Observe that the posterior pdf in (13) can be broken up into two pieces as

$$\pi(\sigma^2, \lambda | y^n) = \pi(\sigma^2 | y^n, \lambda) \cdot \pi(\lambda | y^n),$$

where the conditional posterior pdf of σ^2 given λ and y^n is an updated Inverted Gamma
$(\frac{v^*+n}{2}, \frac{\delta^{**}_\lambda}{2})$ density with $\delta^{**}_\lambda = \delta^* + \sum_{t=1}^{n} \tilde{y}^2_{t|t-1,\lambda} / \tilde{f}_{t|t-1,\lambda}$.

The second piece, the marginal posterior of λ given y^n, is obtained by integrating out σ^2 from (13). Thus we have the following.

THEOREM 1: *Let the prior of (σ^2, λ) be as in (12). Then the conditional posterior distribution of σ^2 given y^n and λ is*

$$\sigma^2 | y^n, \lambda \sim \textit{Inverted Gamma}\left(\frac{v^*+n}{2}, \frac{\delta^{**}_\lambda}{2} \right), \tag{14}$$

and the marginal posterior pdf of λ given y^n is

$$\pi(\lambda | y^n) \; \propto \; \pi(\lambda) \left[\prod_{t=1}^{n} \tilde{f}^{-1/2}_{t|t-1,\lambda} \right] \left(\delta^* + \sum_{t=1}^{n} (\tilde{y}_{t|t-1,\lambda})^2 / \tilde{f}_{t|t-1,\lambda} \right)^{(-v^*+n)/2} \tag{15}$$

REMARK 2: If we wish to obtain the posterior pdf's when the prior of σ^2 is diffuse and is given by Jeffreys pdf: $\pi(\sigma^2) \propto 1/\sigma^2$, all that has to be done is to set $v^* = 0$ and $\delta^* = 0$ in (12)-(15).

REMARK 3: Unfortunately the posterior of λ in (15), for any prior $\pi(\lambda)$, does not belong to a known parametric family of densities. Nonetheless, (15) can be analyzed numerically, and the normalizing constant determined using Monte Carlo integration procedures. The inverse of the normalizing constant, as stated before, is the integral of the RHS of (15) over λ. In any given application, it will be necessary to approximate $\pi(\lambda|y^n)$ by a parametric density. Approximating pdf's that are symmetric are probably not good choices in small samples since in the example discussed in the next section, the pdf of $\lambda|y^n$ is shown to be skewed.

3.2 The Conditional Smoothing Problem

We now derive the posterior distribution of β given y^n and θ using a vector-matrix representation of (1) that allows us to appeal to known results from Bayes analysis of linear models. The vector of smoothed estimates of β given θ is then the posterior mean of this distribution.

If we let $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)' : n \times 1$, $\eta = (\eta_1, \eta_2, \ldots, \eta_n)' : n \times 1$, and $X = diag(x_1, \ldots, x_n) : n \times n$, equation (1) can be expressed in vector-matrix form as

$$y^n = X\beta + \varepsilon, \quad \varepsilon \sim N_n(0, \sigma^2 I_n) \tag{16}$$

with the evolution equation of the regression parameters summarized as

$$H\beta = e\beta_0 + \eta, \qquad \eta \sim N_n(0, \lambda\sigma^2 I_n), \tag{17}$$

where the matrix H is given in the simple form:

$$H = \begin{bmatrix} 1 & 0 & & 0 \\ -g_2 & 1 & & 0 \\ 0 & -g_3 & 1 & 0 \\ \cdot & & & \cdot \\ \cdots & & & \cdots \\ 0 & 0 & -g_n & 1 \end{bmatrix} : n \times n,$$

and $e = (g_1, 0, \ldots, 0)$ is an $n \times 1$ vector. Under this representation of the TVR model, it is relatively straightforward to derive the posterior pdf of β given y^n and θ. Equation (17), along with the assumption that the distribution of β_0 is $\beta_0|\theta \sim N_k(\hat{\beta}_{0|0}, \sigma^2 R_{0|0})$ imply that the prior of β given θ is

$$\beta|\theta \sim N_n(\beta^* = H^{-1}e\hat{\beta}_{0|0}, A_\theta^{*-1} = \sigma^2 \tilde{A}_\lambda^{*-1})$$

where $\tilde{A}_\lambda^{*-1} = H^{-1}eR_{0|0}e'H^{-1\prime} + \lambda H^{-1}H^{-1\prime}$. Consequently using standard calculations from Bayes analysis of linear models, we have that the distribution of β given y^n and θ is multivariate normal with mean vector $\hat{\beta}_\lambda : n \times 1$, and variance matrix $\sum_\theta : n \times n$, i.e.,

$$\beta | y^n, \theta \sim N_n(\hat{\beta}_\lambda, \Sigma_\theta) \tag{18}$$

where the posterior mean is the matrix weighted average

$$\hat{\beta}_\lambda = (\tilde{A}_\lambda^* + X'X)^{-1}(\tilde{A}_\lambda^*\beta^* + X'y^n), \tag{19}$$

and the posterior variance, which does not depend on y^n, is

$$\Sigma_\theta = \sigma^2(\tilde{A}_\lambda^* + X'X)^{-1} \equiv \sigma^2\Sigma_\lambda. \tag{20}$$

The smoothed estimate of β, given θ, is given by (19). This estimator of β is, optimal under the squared error loss function. Note that the smoothed value of β, i.e., $\hat{\beta}_\lambda$ does not depend on σ^2.

The marginal posterior density of β given λ and y^n can be obtained using (14), and is given by

$$\pi(\beta|y^n, \lambda) = \int_{\sigma^2>0} \pi(\beta|y^n, \sigma^2, \lambda)\pi(\sigma^2|y^n, \lambda)d\sigma^2,$$

where from (18), $\pi(\beta|y^n, \sigma^2, \lambda)$ is the pdf of a $N_n(\hat{\beta}_\lambda, \sum_\theta)$ distribution. Hence

$$\begin{aligned} \pi(\beta|y^n, \lambda) &\propto \int_{\sigma^2>0} (1/\sigma^2)^{\frac{v*+2n}{2}+1} e^{-\frac{1}{2\sigma^2}[(\beta-\hat{\beta}_\lambda)'\Sigma_\lambda^{-1}(\beta-\hat{\beta})+\delta_\lambda^{**}]} d\sigma^2 \\ &\propto [\delta_\lambda^{**} + (\beta - \hat{\beta}_\lambda)'\Sigma_\lambda^{-1}(\beta - \hat{\beta}_\lambda)]^{-(v^*+2n)/2} \end{aligned} \tag{21}$$

which is the kernel of a multivariate t density with mean $\hat{\beta}_\lambda$, dispersion $\tilde{\delta}_\lambda^{**} \sum_\lambda$, and $v^* + n$ degrees of freedom, where $\tilde{\delta}_\lambda^{**} \equiv \delta_\lambda^{**}/(v * +n)$.

3.3 The Unconditional Smoothing Problem

At this point, the marginal posterior of β can be derived using the results in Sections (3.1) and (3.2). Since the marginal posterior of λ, and the conditional posterior of β given λ, is available, the unconditional posterior of β is obtained as

$$\pi(\beta|y^n) = \int_\lambda \pi(\beta|y^n, \lambda)\pi(\lambda|y^n)d\lambda$$

where $\pi(\lambda|y^n)$ is given in (15) and $\pi(\beta|y^n, \lambda)$ in (21). Although the above integral cannot be performed analytically, we can state the following result.

THEOREM 2: *Let $\pi(\lambda|y^n)$ be as in (15). Then in the TVR model, the marginal posterior distribution of β is*

$$\pi(\beta|y^n) \propto \int \left[1 + \frac{1}{\tilde{\delta}^{**}_\lambda (v^* + n)} (\beta - \hat{\beta}_\lambda)' \Sigma_\lambda^{-1} (\beta - \hat{\beta}_\lambda)\right]^{-(v^*+2n)/2} \cdot \pi(\lambda|y^n) d\lambda. \quad (22)$$

REMARK 4: If one is interested in the pdf of some subvector of β, for example, the most current value β_n, the posterior is obtained as follows. If we denote the appropriate subcomponents of $\hat{\beta}_\lambda$ and $\sum_\lambda$ by $\hat{\beta}_{n,\lambda}$ and $\sum_{n,\lambda}$ respectively, then from (21) and Theorem 2, the pdf of β_n is

$$\pi(\beta_n|y^n) \propto$$
$$\int_\lambda \left[1 + \frac{1}{\tilde{\delta}^{**}_\lambda (v^* + n)} (\beta_n - \hat{\beta}_{n,\lambda})' \Sigma_{n,\lambda}^{-1} (\beta_n - \hat{\beta}_{n,\lambda})\right]^{-(1+v^*+n)/2} \cdot \pi(\lambda|y^n) d\lambda. \quad (23)$$

An important issue at this stage is how the pdf in (22), and also (23), should be analyzed. The most effective method is to use Monte Carlo (MC) integration procedures that have been recently applied by Zellner and Rossi (1984), Zellner, Van Dijk and Bauwens (1988), and Chib, Tiwari and Jammalamadaka (1988), among others. The general method is developed in Klock and van Dijk (1978), and Geweke (1987). Application of that methodology to the specific problem at hand can proceed as follows. Suppose we are interested in obtaining the unconditional moments of the elements of β, for example the mean and variance. We can use well known formulas to compute marginal moments from conditional moments. For example, to compute the unconditional mean of β, we have

$$\hat{\beta} = E^{\pi(\beta|y^n)}[\beta] = \int_\beta \beta \pi(\beta|y^n) d\beta$$
$$= \int_\lambda \left[\int_\beta \beta \pi(\beta|y^n, \lambda) d\beta\right] \pi(\lambda|y^n) d\lambda = \int_\lambda \hat{\beta}_\lambda \pi(\lambda|y^n) d\lambda. \quad (24)$$

This can be computed numerically as follows. Let $h(\lambda)$ be an *importance density* with the property that it closely approximates $\pi(\lambda|y^n)$. Further $h(\lambda)$ should be a pdf *from which it is relatively easy to simulate.* Let $\lambda_1, \ldots, \lambda_N$ be N random draws of λ from $h(\lambda)$ where N is a suitably large number. Then the MC estimate of (24) is

$$\hat{\beta} \simeq \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_{\lambda_i} \frac{\pi(\lambda_i|y^n)}{h(\lambda_i)}, \quad (25)$$

where $\hat{\beta}_{\lambda_i}$ is $\hat{\beta}_\lambda$ evaluated at $\lambda = \lambda_i$, and $\pi(\lambda_i|y^n)$ is the posterior of λ evaluated at $\lambda = \lambda_i$, i.e.,

$$\pi(\lambda_i|y^n) =$$

$$\frac{\pi(\lambda_i)(\Pi_{t=1}^n \tilde{f}_{t|t-1,\lambda_i}^{-1/2})(\delta^* + \sum_{t=1}^n \tilde{y}_{t|t-1,\lambda_i}^2/\tilde{f}_{t|t-1,\lambda_i})^{-(v^*+n)/2}}{\frac{1}{N}\sum_{i=1}^N \pi(\lambda_i)(\Pi_{t=1}^n \tilde{f}_{t|t-1,\lambda_i}^{-1/2})(\delta^* + \sum_{t=1}^n \tilde{y}_{t|t-1,\lambda_i}^2/\tilde{f}_{t|t-1,\lambda_i})^{-(v^*+n)/2}/h(\lambda_i)}. \quad (26)$$

Note that the denominator of (26) is the MC estimate of the normalizing constant of $\pi(\lambda|y^n)$.

Now the marginal covariance matrix of β is defined as the sum of the expectation of the conditional variance and the variance of the conditional expectation, i.e.,

$$\sum = V^{\pi(\beta|y^n)}[\beta] = \int_\lambda \frac{v^* + n}{(v^* + n - 2)} \tilde{\delta}_\lambda^{**} \Sigma_\lambda \pi(\lambda|y^n) d\lambda$$

$$+ \int_\lambda (\hat{\beta}_\lambda - \hat{\beta})(\hat{\beta}_\lambda - \hat{\beta})' \quad \pi(\lambda|y^n) d\lambda. \quad (27)$$

where $\hat{\beta}$ is defined in (24). Each of these terms may be approximated by sums as in (25).

When we consider the method described above, it is clear that we have to find a suitable importance density $h(\lambda)$. If the underlying pdf of λ were symmetric, it may be possible to let the importance density be a student t density with parameters suitably chosen. This strategy will not work when the pdf of λ is nonsymmetric, as in the example in the next section. Although we have not been able to test this assertion in a wide array of examples, we are of the opinion that the pdf of λ is nonsymmetric, unless the sample size is large. Thus, symmetric importance functions are unlikely to be adequate for this problem.

4. An Example

In this section we report some numerical results from a Monte Carlo experiment. The idea is to illustrate how the methods described in the paper work in practice.

The data for the experiment is generated according to the model

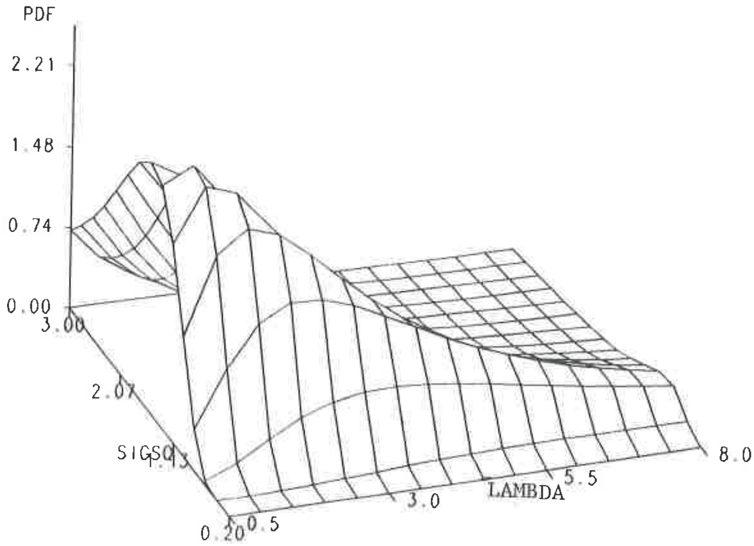
$$y_t = x_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0,1),$$

$$\beta_t = .5\beta_{t-1} + \eta_t, \quad \eta_t \sim N(0, 0.5), \quad t = 1, 2, \ldots, 15 \quad (28)$$

with $\beta_0 \sim N(.9, 1)$, and x_t distributed as iid Uniform(0, 1) random variables. In terms of the model of the paper, we have set $\hat{\beta}_{0|0} = .9$; $\sigma^2 = 1$; $\lambda = 0.5$; and $R_{0|0} = 1$. Fifteen observations are generated from model (28).

Using equation (13), and a prior of (σ^2, λ) that is proportional to $1/\sigma^2$, the joint posterior pdf of σ^2, λ and λ is reproduced in Figure 1. It is clear that the joint pdf has the appropriate shape with most probability concentrated in the parameter space that generated the data.

Figure 1: Joint Posterior Density Function



The marginal pdf of λ, from equation (15) is shown in Figure 2. Also shown in Figure 2 is the pdf of a half normal distribution with parameter $\theta = .24$. This pdf has the form

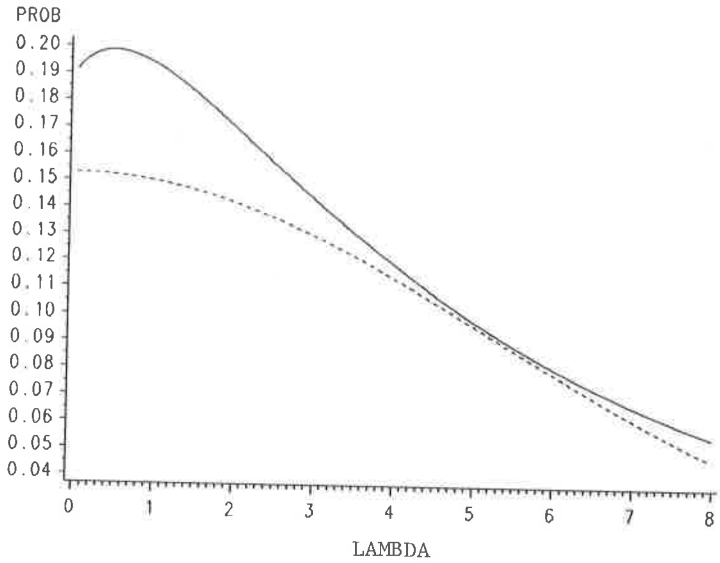
$$h(\lambda) = \frac{2\theta}{\pi} e^{-\theta^2 \lambda^2/\pi}, \qquad 0 \leq \lambda < \infty, \tag{29}$$

and appears to be an appropriate importance density.

Note that the pdf of λ is skewed with fairly heavy tails, and a pronounced peak in the region around $\lambda = 0.5$.

Using the half normal importance density specified in (29) we have computed the conditional and unconditional smoothed estimates of β as described in (25) and (26).

Figure 2: Posterior PDF of Lambda and Half Normal Approximation



Legend: $\pi(\lambda \mid y^n)$: ———————— ; $h(\lambda)$: – – – – –.

[The normalizing constant of $\pi(\lambda \mid y^n)$ is found by Monte Carlo integration. The importance function is the uniform density on $(0, 6)$ which was used to produce 500 random draws of λ.]

At the suggestion of a referee, the estimates are calculated from repeated samples, not from one Monte Carlo run. By keeping β_t and x_t fixed at the values in the previous examples, 30 samples of size 15 are generated from the model in (28). *The estimates reported in Table 1 are averages of the estimates from the* 30 *samples.* Also reported is the conditional variance of β (i.e., diag $\sigma^2 \sum_\lambda$) with $\sigma^2 = 1$ and $\lambda = 0.5$, and the unconditional variance of β using (27). The results are summarized in Table 1 which has seven columns. The first consists of the true values of the components of β, and second contains the filtered estimates using $\lambda = 0.5$. The third and fourth provide the conditional smoothed estimates of β (i.e., $\hat{\beta}_\lambda$), and the conditional variance of β. The last two columns contain the unconditional estimates of β and its variance.

In interpreting the evidence in Table 1, we should first note the satisfactory performance of the unconditional estimate. On graphing in Figure 3, the absolute

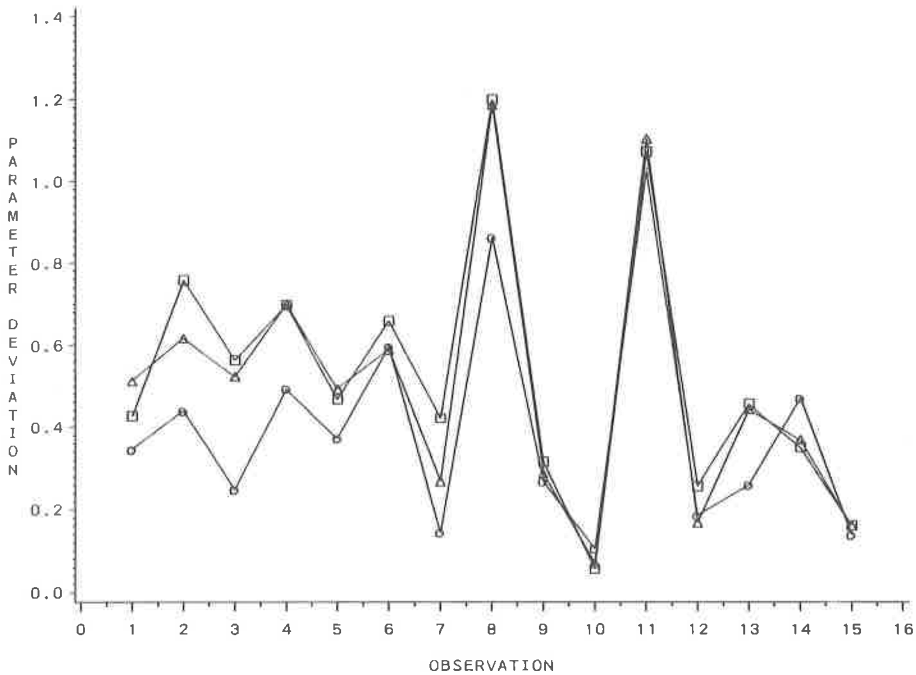
Table 1

obs	β	$\hat{\beta}_{t\|t,\lambda=0.5}$	$R_{t\|t,\lambda=0.5}$	$\hat{\beta}_{\lambda=0.5}$	$diag\sum_{\theta}$	$\hat{\beta}$	$diag\sum$
1	1.14560	0.72006	0.43161	0.63597	0.42027	0.80254	0.54337
2	-0.49108	0.26409	0.55852	0.12317	0.51788	-0.05372	1.20456
3	-0.71234	-0.15107	0.44494	-0.19072	0.42641	-0.46455	0.72449
4	-0.88160	-0.18723	0.47735	-0.18449	0.47137	-0.39093	0.85193
5	-0.57651	-0.11087	0.58570	-0.08649	0.57908	-0.20567	1.49513
6	0.58402	-0.07145	0.62570	-0.00162	0.61419	-0.00878	1.76890
7	0.37801	-0.04192	0.65443	0.11079	0.60574	0.23466	1.81598
8	1.47084	0.27726	0.50172	0.28875	0.46337	0.61048	0.84762
9	0.45402	0.14029	0.39500	0.16727	0.38702	0.18634	0.56644
10	0.22205	0.16462	0.52929	0.15194	0.52534	0.32790	1.22710
11	1.15013	0.08290	0.63229	0.05201	0.60979	0.12943	1.85377
12	-0.19034	0.06482	0.59747	-0.02284	0.56058	-0.00588	1.36721
13	-0.59954	-0.14579	0.48406	-0.15813	0.47502	-0.34160	0.83063
14	0.26274	-0.08539	0.61649	-0.10455	0.56152	-0.20503	1.58431
15	-0.24382	-0.08335	0.40660	-0.08335	0.40660	-0.10558	0.59966

Note: The estimates reported in the table are averages calculated from 30 samples of size 15. The number of replications from the half normal pdf in (29) is $N = 500$. The joint prior of λ and σ^2 is taken to be proportional to $1/\sigma^2$. Thus in the calculations, $\upsilon * = 0$ and $\delta^* = 0$.

deviations of the estimates from the true parameter, we can conclude that the unconditional estimate is, in general, closer to the true value than both the filtered and conditional estimates.

Figure 3: Absolute Parameter Deviations



Legend : 0 = $|\hat{\beta} - \beta|$; $\Delta = |\hat{\beta}_{\lambda=0.5} - \beta|$; $\square = |\hat{\beta}_{t|t,\lambda=0.5} - \beta_t|$

We also point out that the filtered and conditional estimates in the table above are computed using the value of λ that generated the data. When λ is unknown, as in most applications, these estimates can still be computed but with a value of λ that is likely to be incorrect. In such circumstances, it is more reasonable to bear the extra computational burden, and compute the unconditional estimate, than to use the conditional estimate based on a misspecified value of λ. To emphasize this point, we have provided in Table 2, the sampling mean squared error (MSE) of the conditional estimate using two values of λ, i.e., $\lambda = 0.5$ and $\lambda = 5$, and the MSE of the unconditional estimate. The MSE of the optimal estimate, $\beta_{\lambda=0.5}$, is, of course, smaller than that of $\hat{\beta}$. However, the MSE of the conditional estimate using an incorrect value of λ is larger than that of $\hat{\beta}$. Although the results below are based on a limited sampling investigation, we confirmed that similar results are obtained in other examples. To conserve space, those results are not reported.

Table 2: Sampling Mean Squared Error of the Estimates

Computed from 30 samples of size 15

obs	$\hat{\beta}_{\lambda=0.5}$	$\hat{\beta}_{\lambda=5}$	$\hat{\beta}$
1	0.42341	0.70180	0.53698
2	0.44656	0.81684	0.45471
3	0.41962	1.02542	0.63801
4	0.58891	1.06727	0.71121
5	0.28103	1.04921	0.46559
6	0.37140	1.52418	0.66745
7	0.10165	0.51738	0.19686
8	1.50176	1.14964	1.17638
9	0.25482	0.79170	0.46704
10	0.13425	1.57139	0.60643
11	1.25848	1.55346	1.34032
12	0.12108	1.60701	0.72236
13	0.31260	1.01436	0.58318
14	0.19325	0.85835	0.41076
15	0.23863	1.14518	0.70925

5. Conclusion

This paper considers the Bayes smoothing problem in the TVR model and presents several exact results that can be put to use in applications. It is shown that Monte-Carlo numerical procedures can be implemented relatively easily to compute the results in the paper. We are confident that the multiple covariate case can be handled along the lines described in the paper with only a little extra difficulty. In future work, we plan to describe the analysis to cover the multiple covariate case.

Finally, we point out that the results in this paper can be directly used to find the exact distribution of the coefficients in the filtering problem. After noticing from (7) that the posterior distribution of $\beta_t|t, \lambda$ is univariate-t, λ can be integrated out using the results in Section 3. We suppress the details.

Acknowledgments

We are grateful to the anonymous referees for many useful comments on an earlier draft of the paper. We would also like to thank Benjaporn Thangkasemvanthana for providing assistance with the numerical computations reported in the paper.

Partial support for Professor Chib's work was provided by a University of Missouri Summer Research Fellowship.

References

Anderson, B.D., and J.B. Moore (1979). *Optimal Filtering*, New Jersey: Prentice Hall, Inc.

Broemeling, L. (1985). *Bayesian Analysis of Linear Models*, New York: Marcel Dekker.

Broemeling, L., J. Diaz and A.M. Yusoff (1985). "Some Bayesian Solutions for Problems of Adaptive Estimation in Linear Dynamic Systems," *Communications in Statistics: Theory and Methods,* Vol. 14, No. 2, 401-418.

Chib, S., R.C. Tiwari, and S.R. Jammalamadaka (1988). "Bayes Prediction in Regressions with Elliptical Errors," *Journal of Econometrics*, Vol. 38, 349-360.

Chow, G.C. (1983). *Econometrics*, New York: McGraw Hill.

Engle, R.F. and M.W. Watson (1988). "The Kalman Filter: Applications to Forecasting and Rational-Expectations Models," in *Advances in Econometrics*, T.F. Bewley, ed., New York: Fifth World Congress, Cambridge University Press.

Geweke, J. (1987). "Bayesian Inference in Econometric Models Using Monte Carlo Integration", mimeo.

Harvey, A.C. (1981). *Time Series Models*, Oxford: Phillip Allan.

Klock, T. and H. van Dijk (1978). "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte-Carlo", *Econometrica*, Vol. 46, 1-19.

Schweppe, F.C. (1965). "Evaluation of Likelihood Functions for Gaussian Signals", *I.E.E.E. Transactions on Information Theory*, Vol. 11, 61-70.

Zellner, A., L. Bauwens, and H.K. Van Dijk (1988). "Bayesian Specification Analysis and Estimation of Simultaneous Equation Models Using Monte Carlo Methods," *Journal of Econometrics*, Vol. 38, 39-72.

Zellner, A., and P.E. Rossi (1984). "Bayesian Analysis of Dichotomous Quantal Response Models," *Journal of Econometrics*, 25, 365-393.