Journal of Statistical
Theory & Methods
1995, Vol. 6, pp. 29~39

BAYESIAN ESTIMATION PROCEDURES IN MULTIPROCESS DISCOUNT NORMAL MODEL[1]

Joong Kweon Sohn, Sang Gil Kang and Heon Joo Kim[2]

Abstract A model used in the past may be altered at will in modeling for the future. For this situation, the multiprocess dynamic model provides a general framework. In this paper we consider the multiprocess discount normal model with parameters having a time dependent non-linear structure. This model has nice properties such as insensitivity to outliers and quick reaction to abrupt changes of pattern.

Keywords : Multiprocess Dynamic Model, Normal Discount Bayesian Model, Nonlinear Model.

1. Introduction

Dynamic systems have been used by communications and control engineers to the state of a system as it evolves through time since the works of Kalman(1960). Kalman(1960) developed an recursive estimation procedure for the state variables of a linear dynamic system. Ho and Lee(1964) studied the dynamic linear model with Bayesian framework. Duncan and Horn(1972) introduced the Kalman filter by relating the dynamic linear model to random β regression theory using the time varying random parameters as state variables. Harrison and Stevens(1976) summarized the foundations of Bayesian forecasting as the parametric or statespace model, the probabilistic information on model parameters, the sequential model definition which describes the dynamic behavior of model parameters and some uncertainty in choosing the underlying model from a number of discrete alternatives. Ameen and Harrison(1985) developed normal discount Bayesian models in order to overcome some practical disadvantages of dynamic

---

[1] The Present Studies were Supported by the Basic Research Institute Program, Ministry of Education, 1994, Project No. BSRI-94-1403.
[2] Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea.

linear models. West, Harrison, and Migon(1985) developed the dynamic generalized linear model for applications in non-linear, non-normal time series and regression problems. Migon and Harrison(1985) proposed a descriptive model related television advertising to consumer awareness.

The multiprocess dynamic linear model was developed by Harrison and Stevens (1971, 1976) for the time series that contain outliers and are subject to abrupt changes in pattern. Smith and West(1983) and Smith, Gordon, Knapp and Trimble(1983) described a related monitoring procedure for detecting various forms of kidney failure in renal transplant patients. West and Harrison(1986) studied the method of model monitoring and adapting to structural changes in the time series. Bolstad(1986) presented Harrison-Stevens forecasting algorithm and the multiprocess dynamic linear model. Bolstad(1988) developed the multiprocess dynamic generalized linear model. Bolstad(1995) developed the multiprocess dynamic poisson model for estimating and forecasting a poisson random variable with a time-varying parameter. Whittaker and Fruhwirth-Schnatter(1994) used to a triangular multiprocess Kalman filter for detecting bacteriological growth in *routine monitoring of feedstuff. In this paper, we develop multiprocess discount* normal model with non-linear structure by incorporating the perturbation index variable which determines the perturbation distribution. In Section 2, we develop the recursive estimation for the multiprocess discount normal model with parameter non-linearities. In Section 3, we study the proposed recursive estimation for the generalized exponential growth model by using Monte Carlo *simulation study.*

2. Recursive Estimation of the Multiprocess Discount Normal Model

In this section, we are concerned with the mutiprocess discount normal model. We encounter several models which depend on parameter non-linearities in applications. These models with parameter non-linearities can be written in the following form.

$$\text{Observation equation : } y_t = H_t(\beta_t) + w_t$$
$$\text{Evolution equation : } \beta_t = g_t(\beta_{t-1}) + r_t ,$$

where $H_t(\cdot)$ is a known non-linear regression function, $g_t(\cdot)$ is a known non-linear vector evolution function, w_r and r_t are error terms. In these models, we encounter difficulties with determining the posterior distribution of β_t given Y_t

since $H_t(\cdot)$ and / or $g_t(\cdot)$ are non-linear function of β_t and β_{t-1}, respectively. Thus we suggest the linearization technique with discount matrix. The multiprocess dynamic model is like the dynamic model in that the parameter vector on subject to perturbation. However, in the multiprocess model the distribution of the perturbation depends on the perturbation index random variable at that time. The sequences of perturbation index variables are independent of each other and each can be considered to be the outcome of a single multinomial trial with known prior probabilities. The prior probabilities do not have to remain constant over time. This allows prior knowledge by the forecaster into the model, hence the forecasting system is very flexible. This multiprocess dynamic model is expressed as follows.

Let I_t be the perturbation index variable at time t.

$$P(I_t = j) = \pi_t^{(j)} \quad \text{for} \quad j = 1,2,\cdots,k.$$

When $I_t = j$,

$$\beta_t = g_t(\beta_{t-1}) + r_t,$$

where $g_t(\cdot)$ is a known non-linear vector evolution function, r_t is the perturbation vector, which is normally distributed with mean vector O and known variance-covariance matrix $R_t^{(j)}$. The variance-covariance matrix depends on the perturbation index variable I_t and can change over time. The observation equation is given by

$$y_t = H_t(\beta_t) + w_t,$$

where $H_t(\cdot)$ is a known non-linear regression function mapping the n-vector β_t to the real line and observation error, w_t are independent normal distribution with mean O and variance W_t.

Ameen and Harrison(1985) introduced normal discount Bayesian models to overcome some practical disadvantages associated with the dynamic linear model. In practical problems, modelling the parameter change by introducing a perturbation may not be appropriate. Instead of updating the parameter variance matrix by adding the perturbation variance matrix, the normal discount Bayesian model updates it by pre and post multiplication by a discount matrix. This also has the same effect of increasing the variances, and in many cases modellers and forecasters have a more intuitive feel for the appropriate discount matrix than for a perturbation variance matrix.

The assumptions of the dynamic discount normal Bayesian model are the same as those for the dynamic linear model, except that if previous posterior conditional distribution is

$$(\beta_{t-1}|Y_{t-1}) \sim N(\hat{\beta}_{t-1}, V_{t-1})$$

then

$$(\beta_t|Y_{t-1}) \sim N(G_t\hat{\beta}_{t-1}, B_tG_tV_{t-1}G_t'B_t),$$

where G_t is the known matrix of dynamic coefficient at time t and B_t is the discount matrix, a diagonal matrix of discount facor. The effect is similar to that of adding a perturbation in the dynamic linear model. The mean of the subsequent prior distribution is unchanged, and the variance matrix has been inflate to allow for increased uncertainty. However, the variance matrix inflation is multiplicative instead of additive, and this produces some slight differences.

The multiprocess extension of this model allows the discount matrix to have one of k possible values $B_t^{(1)},\cdots,B_t^{(k)}$, depending on the value of the discount index variable I_t. The discount index variables $\{I_t\}$ are an independent sequence of multinomial random trials with known prior probabilities $P(I_t = j) = \pi_t^{(j)}$ which may change over time.

2.1 Recursive Estimation

The initial conditions for the recursive estimation at time t require the k posterior conditional distributions $(\beta_{t-1}|I_{t-1} = i, Y_{t-1}) \sim N(\beta_{t-1}^{(i)}, V_{t-1}^{(i)})$ and the posterior probability $q_{t-1}^{(i)} = P(I_{t-1} = i|Y_{t-1})$. The notation $Y_{t-1} = y_{t-1},\cdots,y_1$ denotes all the observation up to and including y_{t-1}.

For the structure of parameter non-linearities, we suggest the linearization technique. Various linearization techniques have been developed for dynamic non-linear models. The most straightforward and easily interpreted approach is the one that is based on the use of first order Taylor series approximations to the non-linear regression and evolution functions. This requires the assumptions that both $H_t(\cdot)$ and $g_t(\cdot)$ be differentiable functions of their vector arguments. A taylor series expansion of the evolution functions about the estimate of β_{t-1}, $\beta_{t-1}^{(i)}$ gives

$$g(\beta_{t-1}) = g_t(\beta_{t-1}^{(i)}) + G_t(\beta_{t-1} - \beta_{t-1}^{(i)})$$
$$+ \text{ quadratic and higher order terms in elements of } (\beta_{t-1} - \beta_{t-1}^{(i)}),$$

where G_t is the known $n \times n$ matrix derivative of the evolution matrix evaluated at the estimate $\beta_{t-1}^{(i)}$,

$$G_t = [\frac{\partial g_t(\beta_{t-1})}{\partial \beta_{t-1}}']_{\beta_{t-1}=\hat{\beta}_{t-1}^{(i)}}$$

Assuming that terms other than the linear term are negligible, the linearized expression of the evolution equation becomes

$$\begin{aligned} \beta_t &\propto g_t(\hat{\beta}_{t-1}^{(i)}) + G_t(\beta_{t-1} - \hat{\beta}_{t-1}^{(i)}) + r_t \\ &= h_t + G_t \beta_{t-1} + r_t \end{aligned} \tag{2.1}$$

where $h_t = g_t(\beta_{t-1}^{(i)}) - G_t \beta_{t-1}^{(i)}$ is also known.

Proceeding to the observation equation, similar ideas apply. The non-linear regression function is linearized about the expected value $a_t = h_t + G_t \beta_{t-1}^{(i)}$ for β_t,

$$H_t(\beta_t) = H_t(a_t) + F_t'(\beta_t - a_t)$$
$$+ \text{ quadratic and higher order terms in elements of } (\beta_t - a_t),$$

where F_t is the known n-vector derivative of H_t evaluated at the prior mean a_t,

$$F_t = \left[\frac{\partial H_t(\beta_t)}{\partial \beta_t}\right]_{\beta_t = a_t}.$$

Assuming the linear term dominates the expansion, the non-linear regression function is linearized as follows.

$$\begin{aligned} y_t &= H_t(\beta_t) + w_t \\ &\propto f_t + F_t'(\beta_t - a_t) + w_t, \end{aligned} \tag{2.2}$$

where $f_t = H_t(a_t)$.

Assuming (2.1) and (2.2) as an adequate approximation to the model, it follows immediately that the usual multiprocess dynamic linear model applies in the evolution and observation equations.

(1) Evolution Step

In this step, each of these k distributions is updated to time t conditional on $I_t = j$, the perturbation index variable at time t being equal to j for $j = 1,2,\cdots,k$. At time t the prior, one-step forecast and joint distributions for each I_{t-1} and I_t are given as follows.

By using the discount matrix and evolution equation (2.1), the prior distribution of β_t given $I_{t-1}=i, I_t=j$ and Y_{t-1} is

$$(\beta_t | I_{t-1}=i, I_t=j, Y_{t-1}) \sim N(\hat{\beta}_t^{(i,j)}, C_t^{(i,j)}), \tag{2.3}$$

where $\beta_t^{(i,j)} = h_t + G_t \beta_{t-1}^{(i)}$ and $C_t^{(i,j)} = B_t^{(j)} G_t V_{t-1}^{(i)} G_t' B_t^{(j)}$.

By using the prior distribution (2.3) and the observation equation (2.2), the one-step forecast distribution of y_t given $I_{t-1}=i, I_t=j$ and Y_{t-1} is

$$(y_t | I_{t-1}=i, I_t=j, Y_{t-1}) \sim N(f_t + F_t' (\hat{\beta}_t^{(i,j)} - a_t), F_t' C_t^{(i,j)} F_t).$$

Then the joint distributioin of β_t and y_t given $I_{t-1}=i, I_t=j$ and Y_{t-1} is

$$\begin{pmatrix} \beta_t \\ y_t \end{pmatrix} | I_{t-1}=i, I_t=j, Y_{t-1} \Bigg) \sim N\left[\begin{pmatrix} \hat{\beta}_t^{(i,j)} \\ f_t + F_t' (\hat{\beta}_t^{(i,j)} - a_t) \end{pmatrix}, \begin{pmatrix} C_t^{(i,j)} & C_t^{(i,j)} F_t \\ F_t' C_t^{(i,j)} & F_t' C_t^{(i,j)} F_t + W_t \end{pmatrix} \right]. \tag{2.4}$$

(2) Updating Step

In this step, we consider updating the prior distribution of parameter given observation y_t.

By using the standard normal theory for joint distribution of β_t and y_t in the evolution step, (2.4), the posterior distribution of β_t given $I_{t-1}=i$, $I_t=j$ and Y_t is

$$(\beta_t | I_{t-1}=i, I_t=j, Y_t) \sim N(\hat{\hat{\beta}}_t^{(i,j)}, V_t^{(i,j)}), \tag{2.5}$$

where

$$\hat{\hat{\beta}}_t^{(i,j)} = \hat{\beta}_t^{(i,j)} + C_t^{(i,j)} F_t (F_t' C_t^{(i,j)} F_t)^{-1} (y_t - (f_t + F_t' (\beta_t^{(i,j)} - a_t)))$$

and

$$V_t^{(i,j)} = C_t^{(i,j)} - C_t^{(i,j)} F_t (F_t' C_t^{(i,j)} F_t)^{-1} F_t' C_t^{(i,j)}.$$

To complete the development of the recursive estimation, we need to determine the posterior probabilities of the perturbation indices given the present observation. This probability is called the posterior index probability. Using Bayes theorem, we have

$$P_t^{(i,j)} = P(I_{t-1} = i, I_t = j \mid Y_t)$$
$$= \pi_t^{(j)} q_{t-1}^{(i)} \frac{P(y_t \mid I_{t-1} = i, I_t = j, Y_{t-1})}{P(y_t \mid Y_{t-1})}.$$

(3) Collapsing Step

To proceed to time $t+1$, we need to remove the dependence of the joint posterior $P(\beta_t \mid Y_t)$ on possible models obtained at time $t-1$. If we evaluate $P(\beta_t \mid Y_t)$ to time $t+1$ directly, the mixture will expand to k^3 components for β_{t+1}, dependently on all possible combinations of $I_{t+1} = k$, $I_t = j$ and Y_t. However, the principle of approximating such mixtures by assuming that the effects of different models at $t-1$ are negligible for time $t+1$ applies. Thus, in moving to $(t+1)$, $(k \times k)$ component mixture $P(\beta_t \mid Y_t)$ will be collapsed over possible models at $t-1$. Posterior distribution of β_t given $I_t = j$ and Y_t and mean vector and variance-covariance matrix of β_t given $I_t = j$ and Y_t at time t are given as follows.

By using the posterior index probability at time t, the posterior distribution of β_t given $I_t = j$ and Y_t is

$$f(\beta_t \mid I_t = j, Y_t) = \sum_{i=1}^{k} (q_t^{(j)})^{-1} P_t^{(i,j)} f(\beta_t \mid I_{t-1} = i, I_t = j, Y_t),$$

where $q_t^{(j)} = \sum_{i=1}^{k} P_t^{(i,j)}$.

By using the technique of approximation of mixture, the mean vector and variance-covariance matrix of β_t given $I_t = j$ and Y_t are

$$\hat{\beta}_t^{(j)} = \sum_{i=1}^{k} (q_t^{(j)})^{-1} P_t^{(i,j)} \hat{\beta}_t^{(i,j)} \qquad (2.6)$$

and

$$V_t^{(j)} = \sum_{i=1}^{k} (q_t^{(j)})^{-1} P_t^{(i,j)} [V_t^{(i,j)} + (\hat{\beta}_t^{(i,j)} - \hat{\beta}_t^{(j)})(\hat{\beta}_t^{(i,j)} - \hat{\beta}_t^{(j)})'],$$

respectively. We are now the same position as when we started the recursive estimation, so we are ready to repeat the process when the next observation becomes available.

2.2 Forecast Distributions

At time t, the distributions required by the forecaster are the distributions of

$(\beta_t | I_t = i, I_{t+1} = j, Y_t)$ and $(y_{t+1} | I_t = i, I_{t+1} = j, Y_t)$. At time t, the forecast distributions for each $I_t = i$ and $I_{t+1} = j$ are given as follows.

The forecast distribution of β_{t+1} given $I_t = i, I_{t+1} = j$ and Y_t is

$$(\beta_{t+1} | I_t = i, I_{t+1} = j, Y_t) \sim N(\hat{\beta}_{t+1}^{(i,j)}, C_{t+1}^{(i,j)}),$$

where $\beta_{t+1}^{(i,j)} = h_{t+1} + G_{t+1}\beta_t^{(i)}$ and $C_{t+1}^{(i,j)} = B_{t+1}^{(j)} G_{t+1} V_t^{(i)} G_{t+1}' B_{t+1}^{(j)}$.

The forecast distribution of y_{t+1} given $I_t = i, I_{t+1} = j$ and Y_t is

$$(y_{t+1} | I_t = i, I_{t+1} = j, Y_t) \sim N(f_{t+1} + F_{t+1}' (\hat{\beta}_{t+1}^{(i,j)} - a_{t+1}), F_{t+1}' C_{t+1}^{(i,j)} F_{t+1} + W_{t+1}).$$

The unconditional forecast distribution of y_{t+1} given $I_t = i, I_{t+1} = j$ and Y_t is

$$P(y_{t+1} | Y_t) = \sum_{i=1}^{k} \sum_{j=1}^{k} q_t^{(i)} \pi_{t+1}^{(j)} P(y_{t+1} | I_t = i, I_{t+1} = j, Y_t).$$

3. Monte Carlo Simulation Study

In this section, we study the performance of the Bayesian estimation proposed in Section 2 via Monte Carlo simulation for the multiprocess discount normal model.

We consider a member of the generalized exponential growth models by Gamerman and Migon(1991). Let y_t, $t = 1, 2, \cdots, n$, be a time series of interest. The model is defined as

$$(y_t | \beta_t) \sim N(\beta_t, V(\beta_t)\sigma^2),$$

where $\mu_t = \beta_t^{\lambda}$ with non-linear evolution equations

$$\mu_t = \mu_{t-1} + \gamma_{t-1} + w_{t1},$$
$$\gamma_t = \phi_{t-1}\gamma_{t-1} + w_{t2},$$
$$\phi_t = \phi_{t-1} + w_{t3}.$$

μ_t is the level, γ_t is the growth in the level and ϕ_t is the damping factor for the model. The non-linearity in the model is due to multiplicative effect of ϕ_t. The simulation study will be carried out with the following example on an artificially generated time series. The time series consists of 80 normally distributed random variables and the following change pattern. The time series data start with no change, but there is an outlier of at observation 12. At step 21, the growth change starts and to be continued at step 30. At step 31, no change starts. At step 51, the damping factor change starts and to be continued at step 50. At step 61, level change starts and to be continued at step 80, but there is an outlier of at observation 72.

The forecast and the actual observations are shown in Figure 3.1. The forecast errors are shown in Figure 3.2. From these figures, we suggest the following properties.

(i) The developed models give good estimates by using past data as well as presentdata when the time series is in a stable pattern.
(ii) The developed models are not sensitive when an outlier occurs.
(iii) The developed models react quickly when a change occurs, but when a changeoccurs, the forecast error is sligtly increasing.

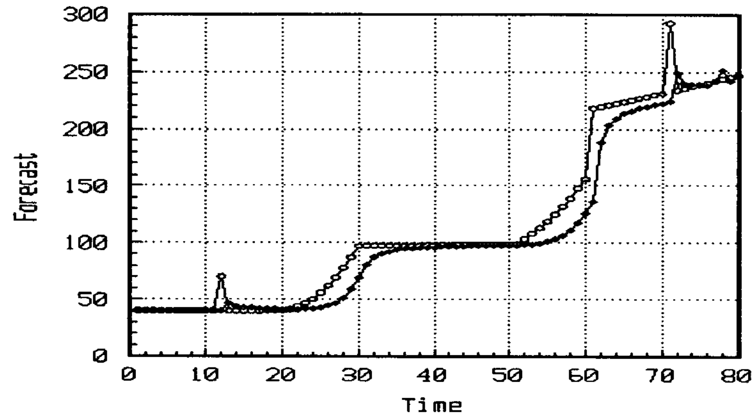

Figure 3.1 Observed(o) and Forecast(•) Value

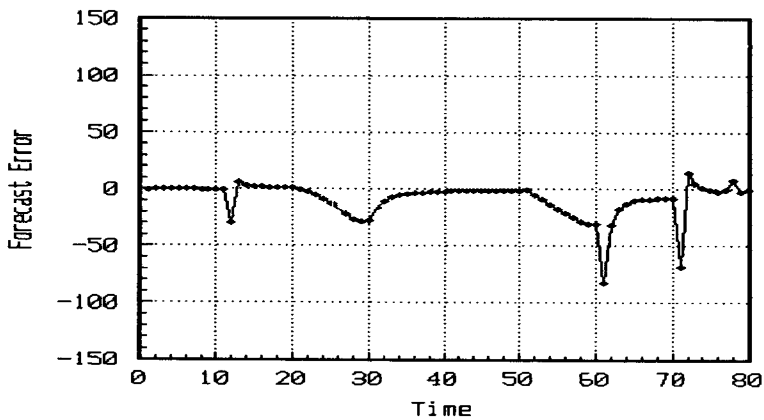

Figure 3.2 Forecast Error

References

Ameen, J. R. M. and Harrison, P. J. (1985), *Normal Discount Bayesian Models*, Bayesian Statistics 2, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, Amsterdam, North Holland.

Bolstad, W. M. (1986), Harrison-Stevens Forecasting and the Multiprocess Dynamic Model, *The American Statistician*, 40, 129-135.

Bolstad, W. M. (1988), Estimation in the Multiprocess Dynamic Generalized Linear Model, *Communications in Statistics: Theory and Method*, 17, 4179-4204.

Bolstad, W. M. (1995), The Multiprocess Dynamic Poisson Model, *Journal of the American Statistical Association*, 90, 227-232.

Duncan, B. and Horn, S. D. (1972), Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis, *Journal of the American Statistical Association*, 67, 815-821.

Gamerman, D. and Migon, H. S.(1991), Forecasting the number of AIDS cases in Brizil, *Statistician*, 40, 427-442.

Harrison, P. J. and Stevens, C. F. (1971), A Bayesian Approach to Short-term Forecasting, *Operations Research Quarterly*, 22, 341-362.

Harrison, P. J. and Stevens, C. F. (1976), Bayesian Forecasting, *Journal of the Royal Statistical Society* B, 38, 205-247.

Ho, Y. C. and Lee, R. C. K. (1964), A Bayesian Approach to Problems in Stochastic Estimation and Control, *IEEE Transactions on Automatic Control*, 33-39.

Kalman, R. E. (1960), A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering*, 95-108.

Migon, H. S. and Harrison, P. J. (1985), An Application of Non-linear Bayesian Forecasting to Television Advertising, I*n Bayesian Statistics 2*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, Amsterdam, North Holland.

Smith, A. F. M. and West, M. (1983), An Application of Multiprocess Kalman Filter to Renal Transplant Data, *Biometrics*, 39, 867-878.

Smith, A. F. M., West, M., Gordon, K., Knapp, M. and Trimble, I. C. (1983), Monitoring Kidney Transplant Patients, *The Statistician*, 32, 46-54.

West, M. and Harrison, P. J. (1986), Monitoring and Adaptation in Bayesian Forecasting Models, *Journal of the American Statistical Association*, 81, 741-750.

West, M., Harrison, P. J. and Migon, H. S. (1985), Dynamic Generalized Linear Models and Bayesian Forecasting, *Journal of the American Statistical Association*, 80, 73-97.

Whittaker, J. and Fruhwirth-Schnatter, S. (1994), A Dynamic Changepoint Model for Detecting the Onset of Growth in Bacteriological Infections, *Applied Statistics*, 43, 625-640.