Journal of Statistical
Data & Information Science Society
1999, Vol. 10, No. 2, pp. 329 ~ 340

Finite Population Prediction under Multiprocess Dynamic Generalized Linear Models

Dal Ho Kim[1] · Young Joon Cha[2] · Jae Man Lee[3]

Abstract

We consider a Bayesian forcasting method for the analysis of repeated surveys. It is assumed that the parameters of the superpopulation model at each time follow a stochastic model. We propose Bayesian prediction procedures for the finite population total under multiprocess dynamic generalized linear models. The multiprocess dynamic model offers a powerful framework for the modelling and analysis of time series which are subject to a abrupt changes in pattern. Some numerical studies are provided to illustrate the behavior of the proposed predictors.

Key Words and Phrases: Kalman filtering, Bayesian prediction, multiprocess dynamic models, finite population sampling.

1. Introduction

Often finite populations are subject to change in time. Repeated surveys carried out at regular time interval have the time series information. In such surveys, one has at one's disposal not only the current data, but also data from similar past experiments. Standard time series analysis methods have been applied to repeated survey data. For example, Blight and Scott(1973) and Scott and Smith(1974) have derived the estimates for the mean of a time dependent population using AR(1) model under the assumption that all the parameters of the model are known. Recently Rodrigues and Bolfarine(1987) considered the prediction of the population total in a finite population using a Bayesian approach based on the Kalman Filter

[1]Assistant Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea
[2]Professor, Department of Statistics, Andong National University, Kyungpook, 760-749, Korea
[3]Associate Professor, Department of Statistics, Andong National University, Kyungpook, 760-749, Korea

algorithm. Bolfarine(1988) replaced the normality assumption by the more general exponential family of distributions.

In this paper, we adapt multiprocess dynamic generalized linear models in the sense of Harrison and Stevens(1976) and Bolstad(1988) to finite population sampling problem. In Section 2, we develop the Bayesian prediction procedures under the multiprocess dynamic generalized linear model for repeated surveys. Here the model is only partially specified in terms of their first and second moments. In Section 3, specifically we devolop the Bayesian predictors under the multiprocess dynamic normal superpopulation model as well as the multiprocess dynamic Poisson superpopulation model, respectively. In Section 4, we investigate the the behavior of proposed predictors via Monte Carlo simulation studies.

2. Bayesian Prediction under Multiprocess Dynamic Models

2.1 Multiprocess Dynamic Models in Repeated Survey

Consider a finite population U with units labeled $1, \cdots, N$. Let y_i denote the value of a single characteristic attached to the unit i. It is considered in the sequel that y_i follows an univariate exponential family of distributions with superparameter η and a known scale parameter ϕ. That is, $f(y_i|\eta,\phi) = \exp[\phi(\eta t - a(\eta)) + c(y,\phi)]$. We select a sample S of size n from the finite population to get information about the finite population total $T = \sum_{i=1}^{N} y_i$. Let $D = \{y_i, i \in S\}$ denote the observed data from the finite population. Then the quantity T can be partitioned as $T = \sum_{i\in S} y_i + \sum_{i\notin S} y_i$. Thus predicting T is equivalent to predicting $\sum_{i\notin S} y_i$. Since

$$E[\sum_{i\notin S} y_i|D] = (N-n)E[\dot{a}(\eta)|D]$$

where $\dot{a}(\eta) = E[y_i|\eta,\phi]$, the Bayesian predictor of T(cf. Ericson(1969)) is given by

$$\widehat{T}_B = E[T|D] = n\overline{y} + (N-n)E[\dot{a}(\eta)|D] \tag{2.1}$$

where $\overline{y}$ is the sample mean. To generalize the above formulation to multiprocess dynamic models in repeated surveys, we replace y, T, D, η, n and $\widehat{T}_B$ by $y_t, T_t, D_t, \eta_t, n_t$ and $\widehat{T}_{D_t}$. Here $D_t = \{y_t, D_{t-1}\}$ represents all the relevant information set available at any time t.

A general Bayesian approach to the analysis of multiprocess dynamic generalized linear models was given by Bolstad(1988). Since y_i is assumed to have a sampling distribution in the exponential faimily, the density of y_t can be expressed as

$$f(y_t|\eta_t,\phi_t) = \exp[\phi_t(\eta_t y_t, a(\eta_t)) + c(y_t,\phi_t)] \tag{2.2}$$

where η_t is the natural continuous parameter of the distribution, ϕ_t is a known scale parameter. The mean and variance of y_t are given by $\mu_t = E(y_t|\eta_t, \phi_t) = \dot{a}(\eta_t)$ and $Var(y_t|\eta_t, \phi_t) = \phi_t^{-1}\ddot{a}(\eta_t)$, respectively. Following Bolstad(1988), we consider the multiprocess dynamic generalized linear model for y_t in the context of repeated survey with the following components.

(i) Observation model :

$$f(y_t|\eta_t, \phi_t) = \exp[\phi_t(\eta_t y_t - a(\eta_t)) + c(y_t, \phi_t)],$$

$$g(\eta_t) = \lambda_t = F_t'\theta_t. \tag{2.3}$$

(ii) Evolution equation :

$$\theta_t = G_t\theta_{t-1} + w_t \tag{2.4}$$

where θ_t is a n-dimensional state vector, F_t is a known n-dimensional regression vector, G_t is a known, $n \times n$ evolution matrix, $\lambda_t = F_t'\theta_t$ is a linear function of the state vector, $g(\eta)$ is a known, continuous and monotonic function mapping η to the real line and w_t is a n-vector of evolution errors whose distribution is specified by zero mean and known variance-covariance matrix $W_t^{(j)}$ which depend on the value of the perturbation index random variable at that time. The perturbation distribution at time t is allowed to take on one of k possibilities conditional on the current value of the perturbation index vairable. The perturbation index variables I_t are a sequence of independent multinomial trials with known prior probabilities $P(I_t = j) = \pi_t^{(j)}$ for $j = 1, \cdots, k$.

2.2 Bayesian Prediction Procedures

Given initial prior information D_0 at $t = 0$, the information set available at any time t is simply $D_t = \{y_t, D_{t-1}\}$ where y_t is observed value of the series at t. Here it is assumed that the initial prior for θ_0 is given as $(\theta_0|D_0) \sim (m_0, C_0)$ irrespective of possible models obtaining at any time where some prior moments m_0 and C_0 are known.

2.2.1 Evolution Step

In this step, evolving to time t, we find the prior distributions about θ_t and λ_t at both $t-1$ and t. For the estimation the initial conditions require that the first and second moments for posterior distribution of θ_{t-1} given $I_{t-1} = i$ and D_{t-1} is known at time t. This distribution is

$$(\theta_{t-1}|I_{t-1} = i, D_{t-1}) \sim (m_{t-1}^{(i)}, C_{t-1}^{(i)}). \tag{2.5}$$

Also at time t we require the posterior probability of perturbation index variable, $P_{t-1}^{(i)} = P(I_{t-1} = i|Y_{t-1})$, and posterior probability are known. By using the posterior distribution of θ_{t-1} and evolution equation (2.4), we obtain the moments of θ_t given $I_{t-1} = i, I_t = j$ and D_{t-1}. Thus the mean vector and variance-covariance matrix are

$$a_t^{(i,j)} = E[\theta_t|I_{t-1} = i, I_t = j, D_{t-1}] = G_t m_{t-1}^{(i)}$$

and

$$R_t^{(i,j)} = Var[\theta_t|I_{t-1} = i, I_t = j, D_{t-1}] = G_t C_{t-1}^{(i)} G_t' + W_t^{(j)},$$

respectively. Therefore the prior distribution for θ_t given $I_{t-1} = i, I_t = j$ and D_{t-1} is

$$(\theta_t|I_{t-1} = i, I_t = j, D_{t-1}) \sim (a_t^{(i,j)}, R_t^{(i,j)}). \tag{2.6}$$

By using the prior distribution (2.6) and the observation equation (2.3), we obtain the joint distribution of θ_t and λ_t. Since $E[\theta_t|I_{t-1} = i, I_t = j, D_{t-1}]$ and $V[\theta_t|I_{t-1} = i, I_t = j, D_{t-1}]$ is given in prior distribution (2.6), we only need to calculate $E[\lambda_t|I_{t-1} = i, I_t = j, D_{t-1}]$, $Var[\lambda_t|I_{t-1} = i, I_t = j, D_{t-1}]$ and $Cov[\theta_t, \lambda_t|I_{t-1} = i, I_t = j, D_{t-1}]$. Thus the moments are

$$f_t^{(i,j)} = E[\lambda_t|I_{t-1} = i, I_t = j, D_{t-1}] = F_t' a_t^{(i,j)}$$

and

$$q_t^{(i,j)} = Var[\lambda_t|I_{t-1} = i, I_t = j, D_{t-1}] = F_t' R_t^{(i,j)} F_t,$$

respectively. The covariance of θ_t and λ_t given $I_{t-1} = i, I_t = j$ and D_{t-1} is

$$Cov[\theta_t, \lambda_t|I_{t-1} = i, I_t = j, D_{t-1}] = F_t' R_t^{(i,j)}.$$

Therefore the joint distribution of θ_t and λ_t given $I_{t-1} = i, I_t = j$ and D_{t-1} is

$$\begin{bmatrix} \theta_t \\ \lambda_t \end{bmatrix} |I_{t-1} = i, I_t = j, D_{t-1} \Bigg] \sim \left(\begin{bmatrix} a_t^{(i,j)} \\ f_t^{(i,j)} \end{bmatrix}, \begin{bmatrix} R_t^{(i,j)} & F_t' R_t^{(i,j)} \\ R_t^{(i,j)} F_t & q_t^{(i,j)} \end{bmatrix} \right). \tag{2.7}$$

By using the method of linear Bayes estimation, the moments of the conditional distribution of θ_t given λ_t , $I_{t-1} = i, I_t = j$ and D_{t-1} are directly obtained in (2.7). Therefore the distribution of θ_t given λ_t, $I_{t-1} = i, I_t = j$ and D_{t-1} is

$$\begin{aligned} &(\theta_t|\lambda_t, I_{t-1} = i, I_t = j, D_{t-1}) \\ &\sim \left[(a_t^{(i,j)} + R_t^{(i,j)} F_t (q_t^{(i,j)})^{-1} (\lambda_t - f_t^{(i,j)}), (R_t^{(i,j)} - R_t^{(i,j)} F_t (q_t^{(i,j)})^{-1} F_t' R_i^{(i,j)}) \right]. \end{aligned}$$

2.2.2 Updating Step

In this step we update the prior distribution of the parameter given the observation y_t. Assume that the prior distribution $(\eta_t|I_{t-1} = i, I_t = j, D_{t-1})$ has the conjugate prior distribution CP $(r_t^{(i,j)}, s_t^{(i,j)})$. That is,

$$f(\eta_t|I_{t-1} = i, I_t = j, D_{t-1}) = \exp[c(r_t^{(i,j)}, s_t^{(i,j)}) + \eta_t r_t^{(i,j)} - s_t^{(i,j)} a(\eta_t)].$$

The parameters $r_t^{(i,j)}$ and $s_t^{(i,j)}$ are chosen to be consistent with the moments for λ_t in joint distribution (2.7). That is, $E[g(\eta_t)|I_{t-1} = i, I_t = j, D_{t-1})] = f_t^{(i,j)}$ and $Var[g(\eta_t)|I_{t-1} = i, I_t = j, D_{t-1})] = q_t^{(i,j)}$. The relationship between the moments of η_t and the moments of λ_t is called the guide relationship by West, Harrison, and Migon(1985).

Now the joint distribution of y_t and η_t is

$$\begin{aligned} &f(y_t, \eta_t|I_{t-1} = i, I_t = j, D_{t-1}) \\ &\quad = \exp[c(y_t, \phi_t) + c(r_t^{(i,j)}, s_t^{(i,j)}) + \eta_t(r_t^{(i,j)} + y_t\phi_t) - (s_t^{(i,j)} + \phi_t)a(\eta_t)] \end{aligned}$$

and the marginal distribution of y_t is

$$\begin{aligned} &f(y_t|I_{t-1} = i, I_t = j, D_{t-1}) \\ &\quad = \exp[c(y_t, \phi_t) + c(r_t^{(i,j)}, s_t^{(i,j)}) - c(r_t^{(i,j)} + \phi_t y_t, s_t^{(i,j)} + \phi_t)]. \end{aligned}$$

Thus posterior distribution of η_t given D_t is

$$\begin{aligned} &f(y_t|I_{t-1} = i, I_t = j, D_t) \\ &\quad = \exp[c(r_t^{(i,j)} + \phi_t y_t, s_t^{(i,j)} + \phi_t) + \eta_t(r_t^{(i,j)} + \phi_t y_t) - (s_t^{(i,j)} + \phi_t)a(\eta_t)], \end{aligned}$$

that is, the posterior distribution of η_t given $I_{t-1} = i, I_t = j$ and D_t is the conjugate CP$(r_t^{*(i,j)}, s_t^{*(i,j)})$ where $r_t^{*(i,j)} = r_t^{(i,j)} + \phi_t y_t$ and $s_t^{*(i,j)} = s_t^{(i,j)} + \phi_t$.

Now we find the moments of the posterior distribution of θ_t by using the moments of λ_t. Since

$$E[\theta_t|I_{t-1} = i, I_t = j, D_t] = E[E(\theta_t|\lambda_t, I_{t-1} = i, I_t = j, D_t)|I_{t-1} = i, I_t = j, D_t]$$

and

$$\begin{aligned} &Var[\theta_t|I_{t-1} = i, I_t = j, D_t] \\ &\quad = E[Var(\theta_t|\lambda_t, I_{t-1} = i, I_t = j, D_{t-1})|I_{t-1} = i, I_t = j, D_t] \\ &\qquad + Var[E(\theta_t|\lambda_t, I_{t-1} = i, I_t = j, D_{t-1})|I_{t-1} = i, I_t = j, D_t], \end{aligned}$$

we obtain the mean vector and variance-covariance matrix as

$$m_t^{(i,j)} = E[\theta_t | I_{t-1} = i, I_t = j, D_t] = a_t^{(i,j)} + R_t^{(i,j)} F_t (q_t^{(i,j)})^{-1} (f_t^{*(i,j)} - f_t^{(i,j)})$$

and

$$\begin{aligned} C_t^{(i,j)} &= Var[\theta_t | I_{t-1} = i, I_t = j, D_t] \\ &= R_t^{(i,j)} - R_t^{(i,j)} F_t (q_t^{(i,j)})^{-1} - (q_t^{(i,j)})^{-2} q_t^{*(i,j)}) F_t' R_t^{(i,j)} \end{aligned}$$

where $f_t^{*(i,j)} = E[g(\eta_t)|I_{t-1} = i, I_t = j, D_t]$ and $q_t^{*(i,j)} = Var[g(\eta_t)|I_{t-1} = i, I_t = j, D_t]$. Therefore when y_t is observed, the posterior distribution of θ_t given $I_{t-1} = i, I_t = j$ and D_t is

$$(\theta_t | I_{t-1} = i, I_t = j, D_t) \sim (m_t^{(i,j)}, C_t^{(i,j)}). \qquad (2.8)$$

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

$$\begin{aligned} P_t^{(i,j)} &= P(I_{t-1} = i, I_t = j | D_t) \\ &= P_{t-1}^{(i)} \pi_t^{(i)} \frac{\exp[\mathrm{c}(\mathrm{y_t}, \phi_t) + \mathrm{c}(\mathrm{r_t^{(i,j)}}, \mathrm{s_t^{(i,j)}}) - \mathrm{c}(\mathrm{r_t^{(i,j)}} + \phi_t \mathrm{y_t}, \mathrm{s_t^{(i,j)}} + \phi_t)]}{P(y_t | D_{t-1})}, \end{aligned}$$

for $i = 1, \cdots, k$ and $j = 1, \cdots, k$. The quantity $P(y_t | D_{t-1}$ is a constant of normalization such that $\sum_{i=1}^k \sum_{j=1}^k P_t^{(i,j)} = 1$. Hence the $P_t^{(i,j)}$ are all completely determined.

2.2.3 Collapsing Step

In moving to time $t+1$, we need to collapse over possible models at time $t-1$. By using the posterior index probabilities at time t, the posterior distribution of $(\theta_t | I_t = j, D_t)$ is represented as a k component mixtures of $(\theta_t | I_{t-1} = i, I_t = j, D_t)$. Thus the posterior distribution is

$$f(\theta_t | I_t = j, D_t) = \sum_{i=1}^{k} (P_t^{(i,j)})^{-1} P_t^{(i,j)} f(\theta_t | I_{t-1} = i, I_t = j, D_t), \qquad (2.9)$$

where $P_t^{(t)} = \sum_{i=1}^k P_t^{(i,j)}$. Also by using the method of approximation of mixture, the mean vector and variance-covariance matrix are

$$m_t^{(j)} = E[\theta_t | I_t = j, D_t] = \sum_{i=1}^{k} (P_t^{(i,j)})^{-1} P_t^{(i,j)} m_t^{(i,j)}$$

and

$$C_t^{(j)} = Var[\theta_t | I_t = j, D_t] = \sum_{i=1}^{k} (P_t^{(i,j)})^{-1} P_t^{(i,j)} [C_t^{(i,j)} + (m_t^{(i,j)} - m_t^{(j)})(m_t^{(i,j)} - m_t^{(j)})'],$$

respectively.

2.2.4 Prediction Step

In this step, we predict the population total by using the posterior distribution of η_t given D_t. Since $\widehat{T}_B = n\overline{y} + (N-n)E[a(\eta)|D]$, T_t is predicted by

$$\widehat{T}_{D_t} = n_t\overline{y}_t + (N_t - n_t)\widehat{a}(\eta_t), \tag{2.10}$$

where $\widehat{a}(\eta_t) = E[\dot{a}(\eta_t)|D_t]$ calculated from the posterior distribution of η_t and $\overline{y}_t$ is the mean of the sample S_t of size n_t selected at time t. Also at time t the posterior variance of T_t is

$$Var[T_t|D_t] = (N_t - n_t)^2 Var[\dot{a}(\eta_t)|D_t] + (N_t - n_t)E[\ddot{a}(\eta_t)/\phi_t|D_t]. \tag{2.11}$$

It is noted that prediction of T_t depends on the posterior moments of the function $\dot{a}(\eta_t)$ and $\ddot{a}(\eta_t)$. Also these moments are easily and explicitly computed for most exponential families and thus leading to explicit expressions for $\widehat{T}_{D_t}$.

At this point, we are in the same position as we were when we started the prediction procedure, so we are ready to repeat the prediction process when the time index updated from $t-1$ to t.

3. Multiprocess Dynamic Superpopulation Models

3.1 Normal Superpopulation Model

We are considered that the finite population at each time was generated according to a normal superpopulation model with known variances. More specifically, at time t, y_{it} is normal distributed with mean μ_t and variance V_t. That is, $y_{it} \sim N(\mu_t, V_t)$ with $\mu_t = \eta_t, i = 1, \cdots, N_t, t = 1, \cdots, T$. So for the observed sample, $\overline{y}_t$ is normal distributed, $f(\overline{y}_t|\eta_t) \sim N(\mu_t, V_t/n_t)$. Here $\dot{a}(\eta_t) = \eta_t = \mu_t$, $\phi_t = n_t/V_t$ and $\overline{y}_t = \sum_{i \in S} y_{it}/n_t$ stands for the sample mean at time t.

The dynamic model is obtained by $g(\eta_t) = \eta_t$ so that $\mu_t = \eta_t = \lambda_t = F_t'\theta_t$. We work in terms of the μ_t notation. At time $t-1$ the dynamic model is completely the posterior distribution of θ_{t-1} given $I_{t-1} = i$ and D_{t-1}, that is,

$$(\theta_{t-1}|I_{t-1} = i, D_{t-1}) \sim (m_{t-1}^{(i)}, C_{t-1}^{(i)}).$$

Now the prediction procedures under the normal multiprocess dynamic superpopulation model is follows:

(i) The joint distribution of μ_t and θ_t given $I_{t-1} = i, I_t = j$ and $D_t - 1$ is

$$\begin{bmatrix} \theta_t \\ \mu_t \end{bmatrix} | I_{t-1} = i, I_t = j, D_{t-1} \Bigg] \sim \left(\begin{bmatrix} a_t^{(i,j)} \\ f_t^{(i,j)} \end{bmatrix}, \begin{bmatrix} R_t^{(i,j)} & F_t' R_t^{(i,j)} \\ R_t^{(i,j)} F_t & q_t^{(i,j)} \end{bmatrix} \right) \tag{3.1}$$

where $a_t^{(i,j)} = G_t m_{t-1}^{(i)}$, $R_t^{(i,j)} = G_t C_{t-1}^{(i)} G_t' + W_t^{(j)}$, $f_t^{(i,j)} = F_t' a_t^{(i,j)}$ and $q_t^{(i,j)} = F_t' R_t^{(i,j)} F_t$.

(ii) When observing $\overline{y}_t$, the posterior distribution of μ_t given $I_{t-1} = i, I_t = j$ and D_t is

$$(\mu_t | I_{t-1} = i, I_t = j, D_t) \sim (f_t^{*(i,j)}, q_t^{*(i,j)}), \tag{3.2}$$

where $f_t^{*(i,j)} = f_t^{(i,j)} + \frac{q_t^{(i,j)}}{q_t^{(i,j)} + V_t/n_t} (\overline{y}_t - f_t^{(i,j)})$ and $q_t^{*(i,j)} = q_t^{(i,j)} - \frac{(q_t^{(i,j)})^2}{q_t^{(i,j)} + V_t/n_t}$ and the posterior distribution of θ_t given $I_{t-1} = i, I_t = j$ and D_t is

$$(\theta_t | I_{t-1} = i, I_t = j, D_t) \sim (m_t^{(i,j)}, C_t^{(i,j)}), \tag{3.3}$$

where $m_t^{(i,j)} = a_t^{(i,j)} + R_t^{(i,j)} F_t (f_t^{*(i,j)} - f_t^{(i,j)}) / q_t^{(i,j)}$ and $R_t^{(i,j)} = R_t^{(i,j)} - R_t^{(i,j)} F_t F_t' R_t^{(i,j)} (1 - q_t^{*(i,j)} / q_t^{(i,j)}) / q_t^{(i,j)}$.

(iii) The posterior distribution of μ_t given $I_t = j$ and D_t is

$$(\mu_t | I_t = j, D_t) \sim (f_t^{*(j)}, q_t^{*(j)}),$$

where $f_t^{*(j)} = \sum_{i=1}^{k} (P_t^{(i)})^{-1} P_t^{(i,j)} f_t^{*(i,j)}$ and $q_t^{*(j)} = \sum_{i=1}^{k} (P_t^{(i)})^{-1} P_t^{(i,j)} [q_t^{*(i,j)} + (f_t^{*(i,j)} - f_t^{*(j)})(f_t^{*(i,j)} - f_t^{*(j)})']$. And the posterior distribution of θ_t given $I_t = j$ and D_t is

$$(\theta_t | I_t = j, D_t) \sim (m_t^{(j)}, C_t^{(j)}), \tag{3.4}$$

where $m_t^{(j)} = \sum_{i=1}^{k} (P_t^{(i)})^{-1} P_t^{(i,j)} m_t^{(i,j)}$ and $C_t^{(j)} = \sum_{i=1}^{k} (P_t^{(i)})^{-1} P_t^{(i,j)} [C_t^{(i,j)} + (m_t^{(i,j)} - m_t^{(j)})(m_t^{(i,j)} - m_t^{(j)})']$.

(iv) The predictor of population total is

$$\begin{aligned} \widehat{T}_{D_t} &= n_t \overline{y}_t + (N_t - n_t) E[\dot{a}(\eta_t) | D_t] \\ &= n_t \overline{y}_t + (N_t - n_t) f_t^*, \end{aligned}$$

where $f_t^* = \sum_{j=1}^{k} (P_t^{(j)})^{-1} f_t^{*(j)}$. The posterior variance of population total is

$$Var[T_t | D_t] = (N_t - n_t)^2 Var[\dot{a}(\eta_t) | D_t] + (N_t - n_t) E[\ddot{a}(\eta_t)/\phi_t | D_t]$$

$$= (N_t - n_t)^2 q_t^* + (N_t - n_t)V_t,$$

where $q_t^* = \sum_{j=1}^{k} P_t^{(j)}[q_t^{*(j)} + (f_t^{*(j)} - f_t^*)(f_t^{*(j)} - f_t^*)']$.

3.2 Poisson Superpopulation Model

In this case, the random quantity y_{it} associated with unit i at time t have the Poisson distribution with superpopulation parameter $\pi_t, i = 1, \cdots, T$. Thus at time t, $\bar{y}_t$, the mean of the y values in the selected sample S_t, follows the expnontial sampling model with $\eta_t = \log(\pi_t)$, $\phi_t = n_t$ and $a(\eta_t) = \exp(\eta_t)$, $t = 1, \cdots, T$. The dynamic model is obtained by $g(\eta_t) = \eta_t$ so that $\eta_t = \lambda_t = F_t'\theta_t$ is given by $\eta_t = \lambda_t = \log(\pi_t) = F_t'\theta_t$ At time $t-1$ the dynamic model is completely the posterior distribution of θ_{t-1} given $I_{t-1} = i$ and D_{t-1}, that is,

$$(\theta_{t-1}|I_{t-1} = i, D_{t-1}) \sim (m_{t-1}^{(i)}, C_{t-1}^{(i)}).$$

Now the prediction procedures under the Poisson multiprocess dynamic superpopulation model is follows:

(i) The joint distribution of μ_t and θ_t given $I_{t-1} = i, I_t = j$ and $D_t - 1$ is the same as (3.1).

(ii) The conjugate prior,CP(r_t, s_t), is log-gamma form for $\eta_t = \log(\pi_t)$. Here using the mode and curvature of $(\eta_t|I_{t-1} = i, I_t = j, D_t)$ for $f_t^{(i,j)}$ and $(q_t^{(i,j)})^{-1}$ lead to $r_t^{(i,j)} = (q_t^{(i,j)})^{-1}$ and $s_t = (q_t^{(i,j)})^{-1}e^{-f_t^{(i,j)}}$. When observing $\bar{y}_t$, the posterior distribution of η_t given $I_{t-1} = i, I_t = j$ and D_t is

$$(\eta_t|I_{t-1} = i, I_t = j, D_t) \sim (f_t^{*(i,j)}, q_t^{*(i,j)}),$$

where $f_t^{*(i,j)} = \log(r_t^{*(i,j)}/s_t^{*(i,j)})$, $q_t^{*(i,j)} = 1/r_t^{*(i,j)}$, $r_t^{*(i,j)} = r_t{}^{(i,j)} + n_t\bar{y}_t$ and $s_t^{*(i,j)} = s_t{}^{(i,j)} + n_t$. Note that the posterior distribution of θ_t given $I_{t-1} = i, I_t = j$ and D_t is the same as (3.3).

(iii) The posterior distribution of θ_t given $I_t = j$ and D_t is the same as (3.4).

(iv) Since

$$f_t^{**(i,j)} = E[\dot{a}(\eta_t)|I_{t-1} = i, I_t = j, D_t] = r_t^{*(i,j)}/s_t^{*(i,j)}$$

and

$$q_t^{**(i,j)} = Var[\dot{a}(\eta_t)|I_{t-1} = i, I_t = j, D_t] = r_t^{*(i,j)}/(s_t^{*(i,j)})^2,$$

we have

$$(\dot{a}(\eta_t)|I_t = j, D_t) \sim (f_t^{**(j)}, q_t^{**(j)}),$$

where $f_t^{**(j)} = \sum_{i=1}^{k} (P_t^{(i)})^{-1} P_t^{(i,j)} f_t^{**(i,j)}$ and $q_t^{**(j)} = \sum_{i=1}^{k} (P_t^{(i)})^{-1} P_t^{(i,j)} [q_t^{**(i,j)} + (f_t^{**(i,j)} - f_t^{**(j)})(f_t^{**(j)} - f_t^{**(j)})']$. And so

$$(\dot{a}(\eta_t)|D_t) \sim (f_t^{**}, q_t^{**})$$

where $f_t^{**} = \sum_{j=1}^{k} P_t^{(j)} f_t^{**(j)}$ and $q_t^{**} = \sum_{j=1}^{k} P_t^{(j)} [q_t^{**(j)} + (f_t^{**(j)} - f_t^{**})(f_t^{**(j)} - f_t^{**})']$. The predictor of population total is

$$\begin{aligned} \widehat{T}_{D_t} &= n_t\bar{y}_t + (N_t - n_t)E[\dot{a}(\eta_t)|D_t] \\ &= n_t\bar{y}_t + (N_t - n_t)f_t^{**}. \end{aligned}$$

The posterior variance of population total is

$$\begin{aligned} Var[T_t|D_t] &= (N_t - n_t)^2 Var[\dot{a}(\eta_t)|D_t] + (N_t - n_t)E[\ddot{a}(\eta_t)/\phi_t|D_t] \\ &= (N_t - n_t)^2 q_t^{**} + (N_t - n_t)f_t^{**}. \end{aligned}$$

4. Simulation Studies

In order to illustrate the behavior of the proposed predictors, we consider the following specific model for our simulation studies.

$$\begin{aligned} y_t &= \mu_t + v_t, \\ \mu_t &= \mu_{t-1} + \beta_t + \delta\mu_t, \\ \beta_t &= \beta_{t-1} + \delta\beta_t. \end{aligned}$$

where the zero-mean, evolution errors $\delta\mu_t$ and $\delta\beta_t$ uncorrelated. This model can be rewritten in the form

$$\begin{aligned} y_t &= F_t'\theta_t + v_t, \\ \theta_t &= G_{t-1}\theta_{t-1} + w_t, \end{aligned}$$

where $F_t = [1 \;\; 0], G_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \theta_t' = [\mu_t, \beta_t], w_t' = [\delta\mu_t + \delta\beta_t, \delta\beta_t]$, and v_t and w_t are distributed with zero means.

The data from the normal superpopulation model is generated as follow. Staring with $\beta_0 = 0.0$, the β_t were generated with the errors $\delta\beta_t$ generated according to the normal disribution with zero mean and variance 1. But in this data there exist a slope change from $t = 20$ to $t = 30$. Next Staring with $\mu_0 = 0.0$, the μ_t were generated with the errors $\delta\mu_t$ generated according to the normal disribution with zero mean and variance 1. But in this data there exist a level change at time $t = 40$.

Next for each μ_t a population of size $N_t = 100$ with the errors v_t generated according to the normal disribution with zero mean and variance 1 were generated. From this population, a sample of size $n_t = 10$ was selected at random, without replacement, $t = 1, \cdots, 50$. Figure 1 represents the performance of predictor $\widehat{T}_{D_t}$. We can see that the sequences $\widehat{T}_{D_t}$ follows very closely the sequence T_t for the underlying change.

The data from the Poisson superpopulation model is generated as follow. Staring with $\beta_0 = 0.0$, the β_t were generated with the errors $\delta\beta_t$ generated according to the normal disribution with zero mean and variance 1. But in this data there exist a slope change from $t = 20$ to $t = 30$. Next Staring with $\mu_0 = 0.0$, the μ_t were generated with the errors $\delta\mu_t$ generated according to the normal disribution with zero mean and variance 1. But in this data there exist a level change at time $t = 40$. Next for each $\pi_t = e^{\eta_t}$ generated, a population of size $N_t = 100$ was generated. From this population, a sample of size $n_t = 10$ was selected at random, without replacement, $t = 1, \cdots, 50$. Figure 2 represents the performance of predictor $\widehat{T}_{D_t}$. We can see that the sequences $\widehat{T}_{D_t}$ follows quite closely the sequence T_t for the underlying change.

References

1. Blight, B. J. N. and Scott A. J. (1973), A Stochastic Model for Repeated Surveys, *Journal of the Royal Statistical Society, series B*, 35, 61-66.

2. Bolfarine, H. (1988). Finite Population Prediction under Dynamic Generlized Linear Models, *Communications in Statistics, Computation and Simulation*, Vol. 17, 187-208.

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

4. Ericson, W. A. (1969). Subjective Bayesian Models in Sampling Finite Populations, *Journal of the Royal Statistical Society*, B31, 195-233.

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

6. Rodrigue, J. and Bolfarine, H. (1987), A Kalman Filter Model for Single and Two-stage Repeated Surveys, *Statistics and Probability Letters*, 5, 299-303.

7. Scott, A. J. and Smith, T. M. F. (1974). Analysis of Repeated Surveys Using Time Series Methods, *Journal of the American Statistical Association*, Vol. 69, 674-678.

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

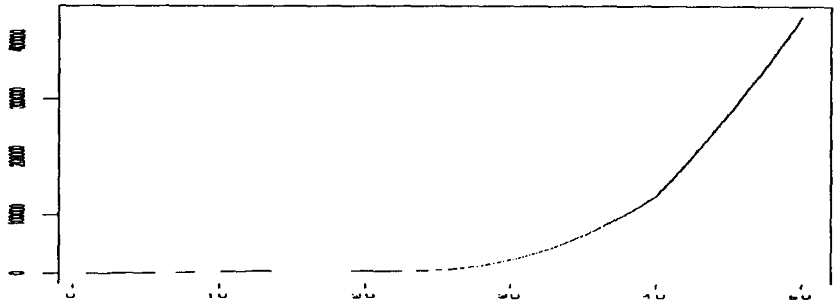

Figure 1: Population Total(solid line) and Predictor(dotted line): Normal Superpopulation Model

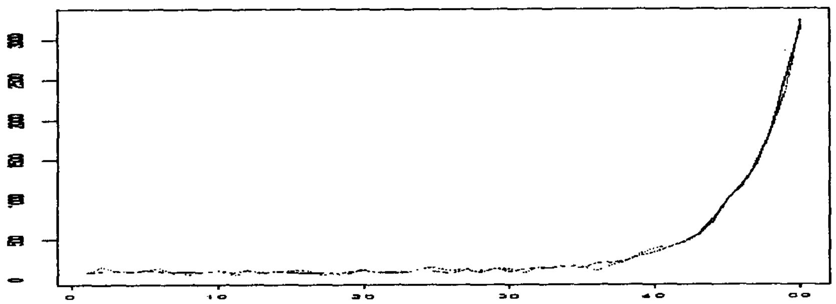

Figure 2: Population Total(solid line) and Predictor(dotted line): Poisson Superpopulation Model