Ann. Inst. Statist. Math.
Vol. 42, No. 3, 435–444 (1990)

BAYESIAN LINEAR PREDICTION IN FINITE POPULATIONS

HELENO BOLFARINE

Universidade de São Paulo, Instituto de Matematica e Estatistica, Caixa Postal 20570, Agencia Iguatemi, 01498 São Paulo, Brasil

(Received November 24, 1988; revised October 16, 1989)

Abstract. In this paper, Bayesian linear prediction of the total of a finite population is considered in situations where the observation error variance is parameter dependent. Connections with least squares prediction (Royall (1976, *J. Amer. Statist. Assoc.*, **71**, 657–664)) in mixed linear models (Theil (1971, *Principles of Econometrics*, Wiley, New York)), are established. Extensions to the case of dynamic (state dependent) superpopulation models are also proposed.

Key words and phrases: Bayes linear prediction, parameter dependent error variance, mixed linear model, dynamic superpopulation model.

1. Introduction

Let the finite population be denoted by $\mathscr{P} = \{1,..., N\}$. Associated with unit k of $\mathscr{P}$, there are $p + 1$ quantities $y_k, x_{k1},..., x_{kp}$, where all but y_k are known. The quantity y_k is considered to be a realization of a random variable $Y_k, k = 1,..., N$, but, since both are unknown, it is not distinguished between them. Relating the two sets of variables, $\boldsymbol{y}' = (y_1,..., y_N)$ and $\boldsymbol{X}' = (\boldsymbol{X}_1,..., \boldsymbol{X}_N)$, where $\boldsymbol{X}_k = (x_{k1},..., x_{kp})'$, $k = 1,..., N$, we consider the following conditional superpopulation model:

$$E[\boldsymbol{y}|\theta] = \boldsymbol{X}\theta \quad \text{and} \quad \mathrm{Var}\,[\boldsymbol{y}|\theta] = \boldsymbol{V}(\theta)\,. \tag{1.1}$$

With respect to the random vector θ, it is assumed that

$$E[\theta] = \mu \quad \text{and} \quad \mathrm{Var}\,[\theta] = \boldsymbol{\Omega}\,,$$

a known matrix. Note that the superpopulation model (1.1) above is more general than the superpopulation model considered by Smouse (1984), for example, in the sense that the covariance matrix of the error vector is allowed to depend on θ.

In order to gain information about the population total $T = \sum_{j=1}^{N} y_j$, a sample s of size n is selected from $\mathscr{P}$ according to some specified sampling plan. After s has been selected, we may write $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{V}(\theta)$, $\boldsymbol{\Omega}$ and $E[\boldsymbol{V}(\theta)] = \bar{\boldsymbol{V}}$ in the following obvious fashion:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_s \\ \boldsymbol{y}_r \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_s \\ \boldsymbol{X}_r \end{pmatrix}, \quad \boldsymbol{V}(\theta) = \begin{pmatrix} \boldsymbol{V}_s(\theta) & \boldsymbol{V}_{sr}(\theta) \\ \boldsymbol{V}_{rs}(\theta) & \boldsymbol{V}_r(\theta) \end{pmatrix},$$

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_s & \boldsymbol{\Omega}_{sr} \\ \boldsymbol{\Omega}_{rs} & \boldsymbol{\Omega}_r \end{pmatrix}, \quad \bar{\boldsymbol{V}} = \begin{pmatrix} \bar{\boldsymbol{V}}_s & \bar{\boldsymbol{V}}_{sr} \\ \bar{\boldsymbol{V}}_{rs} & \bar{\boldsymbol{V}}_r \end{pmatrix},$$

where the subscript s corresponds to the observed units in $\mathscr{P}$ and r corresponds to the units not in s.

The main object of this paper is to develop a distribution free Bayesian approach to predict the population total T under the superpopulation model (1.1). That is, in the class of the linear predictors $\hat{T}_L = a + \boldsymbol{h}'\boldsymbol{y}_s$, where a and the $n \times 1$ vector $\boldsymbol{h}$ are known, we seek to minimize the Bayes risk (total mean squared error) $E[\hat{T}_L - T]^2$, where the (unconditional) expectation operator is taken with respect to the joint distribution of $\boldsymbol{y}$ and θ. But the derivations will require only the first and second moments of those distributions.

In Section 2, the Bayesian linear predictor of the population total T is derived by using some results in Rao (1973). Relationships with other approaches are noted. In Section 3, an important connection with least squares theory in mixed linear models is established. By exploring this connection, a close relationship between prediction of T and estimation of θ is obtained which extends existing results in the literature. In Section 4, we make use of the connection with least squares prediction in mixed linear models to propose a linear Kalman filter version of the least squares predictor under dynamic (state dependent) superpopulation models.

2. General results

In this section, the Bayes linear predictor of T is derived under the superpopulation model (1.1). Some special cases and relationships to other approaches are discussed.

Let $\mathbf{1}_s$ and $\mathbf{1}_r$ be vectors of ones of dimensions n and $N - n$, respectively. After s has been selected we may write $T = \mathbf{1}_s'\boldsymbol{y}_s + \mathbf{1}_r'\boldsymbol{y}_r$ in such a way that predicting T is therefore equivalent to predicting $T_r = \mathbf{1}_r'\boldsymbol{y}_r$. Let $\boldsymbol{h}$ and $\tilde{\boldsymbol{h}}$ be known n-dimensional vectors. As emphasized before, we restrict ourselves to linear predictors of T, that is, predictors of the form

$$\hat{T}_L = a + \boldsymbol{h}'\boldsymbol{y}_s = \mathbf{1}_s'\boldsymbol{y}_s + \tilde{\boldsymbol{h}}'\boldsymbol{y}_s + a = \mathbf{1}_s'\boldsymbol{y}_s + \hat{\tau}\,, \tag{2.1}$$

where $\hat{\tau} = \tilde{\boldsymbol{h}}'\boldsymbol{y}_s + a$ is a predictor of T_r and seek to minimize

$$E[\hat{T}_L - T]^2 . \tag{2.2}$$

The lemma that follows provides an important link between the estimation of linear functions of θ and the prediction of T.

LEMMA 2.1. *Let $\hat{T}_L$ be any linear predictor of the form* (2.1) *above. Then,*

$$\begin{aligned} E[\hat{T}_L - T]^2 = {} & E\{[\hat{\tau} - \mathbf{1}_r'\bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{y}_s] - \mathbf{1}_r'[\boldsymbol{X}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{X}_s]\theta\}^2 \\ & + \mathbf{1}_r'\{\bar{\boldsymbol{V}}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\bar{\boldsymbol{V}}_{sr}\}\mathbf{1}_r . \end{aligned} \tag{2.3}$$

PROOF. After straightforward but lengthy algebraic manipulations, it can be shown that

$$\begin{aligned} E[\hat{T}_L - T]^2 &= E[a + \tilde{\boldsymbol{h}}'\boldsymbol{y}_s - \mathbf{1}_r'\boldsymbol{y}_r]^2 \\ &= \operatorname{Var}[(\tilde{\boldsymbol{h}}' - \mathbf{1}_r'\bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1})\boldsymbol{y}_s + a] \\ &\quad + \mathbf{1}_r'\{\bar{\boldsymbol{V}}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\bar{\boldsymbol{V}}_{sr}\}\mathbf{1}_r + \{E[\hat{\tau} - \mathbf{1}_r'\boldsymbol{X}_r\theta]\}^2 \\ &= \operatorname{Var}\{[\hat{\tau} - \mathbf{1}_r'\bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{y}_s] - \mathbf{1}_r'[\boldsymbol{X}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{X}_s]\theta\} \\ &\quad + \mathbf{1}_r'\{\bar{\boldsymbol{V}}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\bar{\boldsymbol{V}}_{sr}\}\mathbf{1}_r + \{E[\hat{\tau} - \mathbf{1}_r'\boldsymbol{X}_r\theta]\}^2 , \end{aligned} \tag{2.4}$$

from which the result follows.

As a direct consequence of Lemma 2.1, we have

THEOREM 2.1. *Let $\hat{T}_{L1}$ and $\hat{T}_{L2}$ be two linear predictors of the form* (2.1). *Then,*

$$E[\hat{T}_{L1} - T]^2 \le E[\hat{T}_{L2} - T]^2 , \tag{2.5}$$

if and only if

$$\begin{aligned} & E\{[\hat{\tau}_1 - \mathbf{1}_r'\bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{y}_s] - \mathbf{1}_r'[\boldsymbol{X}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{X}_s]\theta\}^2 \\ & \quad \le E\{[\hat{\tau}_2 - \mathbf{1}_r'\bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{y}_s] - \mathbf{1}_r'[\boldsymbol{X}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{X}_s]\theta\}^2 . \end{aligned} \tag{2.6}$$

Remark 2.1. The relevant point about Theorem 2.1 is that it establishes a relationship between the problem of predicting T and the problem of estimating a linear function of θ, which is typically an easier problem. Expressions (2.5) and (2.6) enable the construction of predictors of T given estimators of θ. Indeed, they provide generalizations to similar results in Bellhouse (1987), derived under a less general superpopulation model. See also Fuller (1970) and Rodrigues (1989). Also, (2.6) implies that $\hat{\tau}_1 - \mathbf{1}_r'\bar{V}_{rs}\bar{V}_s^{-1}\mathbf{y}_s$ has smaller expected risk than $\hat{\tau}_2 - \mathbf{1}_r'\bar{V}_{rs}\bar{V}_s^{-1}\mathbf{y}_s$ for estimating $\mathbf{1}_r'[\mathbf{X}_r - \bar{V}_{rs}\bar{V}_s^{-1}\mathbf{X}_s]\theta$.

Now, we restrict ourselves to linear unbiased predictors of T, that is, predictors satisfying

$$E[\hat{T}_L - T] = E[\hat{\tau} - \mathbf{1}_r'\mathbf{X}_r\theta] = 0 \,. \tag{2.7}$$

The theorem that follows next gives the Bayes linear unbiased predictor, $\hat{T}_L^*$, of T and its linear prediction variance.

THEOREM 2.2. *In the class of all linear predictors of T which satisfy* (2.7), *the Bayes linear predictor of T is given by*

$$\hat{T}_L^* = \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r'\{\mathbf{X}_r\hat{\hat{\theta}} + \bar{V}_{rs}\bar{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\hat{\theta}})\} \,, \tag{2.8}$$

where

$$\hat{\hat{\theta}} = (\mathbf{X}_s'\bar{V}_s^{-1}\mathbf{X}_s + \Omega^{-1})^{-1}(\mathbf{X}_s'\bar{V}_s^{-1}\mathbf{y}_s + \Omega^{-1}\mu)$$

and

$$\begin{aligned} E[\hat{T}_L^* - T]^2 = {} & \mathbf{1}_r'\{\bar{V}_r - \bar{V}_{rs}\bar{V}_s^{-1}\bar{V}_{sr}\}\mathbf{1}_r \\ & + \mathbf{1}_r'\{\mathbf{X}_r - \bar{V}_{rs}\bar{V}_s^{-1}\mathbf{X}_s\}\{\mathbf{X}_s'\bar{V}_s^{-1}\mathbf{X}_s + \Omega^{-1}\}^{-1} \\ & \cdot\{\mathbf{X}_r - \bar{V}_{rs}\bar{V}_s^{-1}\mathbf{X}_s\}'\mathbf{1}_r \,. \end{aligned} \tag{2.9}$$

PROOF. Following Rao ((1973), Subsection 4a.11.4) it follows from Theorem 2.1 that (2.2) above is minimized when μ is known by taking

$$\hat{T}_L^* = \mathbf{1}_s'\mathbf{y}_s + a^* + \tilde{\mathbf{h}}'^*\mathbf{y}_s = \mathbf{1}_s'\mathbf{y}_s + \hat{\tau}^* \,,$$

where

$$a^* = \mu'\{\mathbf{X}_r - \bar{V}_{rs}\bar{V}_s^{-1}\mathbf{X}_s\}'\mathbf{1}_r$$

and

$$\begin{aligned} \tilde{\mathbf{h}}'^* &= (\bar{V}_s + \mathbf{X}_s\Omega\mathbf{X}_s')^{-1}\mathbf{X}_s\Omega\{\mathbf{X}_r - \bar{V}_{rs}\bar{V}_s^{-1}\mathbf{X}_s\}'\mathbf{1}_r \\ &= \bar{V}_s^{-1}\mathbf{X}_s(\Omega^{-1} + \mathbf{X}_s'\bar{V}_s^{-1}\mathbf{X}_s)^{-1}\{\mathbf{X}_r - \bar{V}_{rs}\bar{V}_s^{-1}\mathbf{X}_s\}\mathbf{1}_r' \,. \end{aligned}$$

Moreover,

$$\begin{aligned}\operatorname{Var}\{[\hat{\tau}^* - \mathbf{1}_r'\bar{V}_{rs}\bar{V}_s^{-1}y_s] - \mathbf{1}_r'[X_r - \bar{V}_{rs}\bar{V}_s^{-1}X_s]\theta\} \\ = \mathbf{1}_r'\{X_r - \bar{V}_{rs}\bar{V}_s^{-1}X_s\}\{\Omega^{-1} + X_s'\bar{V}_s^{-1}X_s\}^{-1}\{X_r - \bar{V}_{rs}\bar{V}_s^{-1}X_s\}\mathbf{1}_r\,,\end{aligned}$$

from which the result follows.

Remark 2.2. If $V(\theta) = V$ is fixed and known, all the above results continue to hold, with $\bar{V}$ replaced by V.

In the case where $V(\theta) = V$ is constant and known, and μ is unknown, we have the following result

THEOREM 2.3. *If* $V(\theta) = V$ *is known and* μ *is unknown, then the Bayes least squares predictor of* T *and its linear prediction variance are given by* (2.1) *and* (2.2) *in Royall* (1976).

The proof follows directly from Rao ((1973), Subsection 4a.11.5), by following the same steps as in the proof of Theorem 2.2.

Remark 2.3. Predictor (2.8) and the linear prediction variance (2.9) may also be obtained (without the unbiasedness restriction (2.7) from (2.1) and (2.2) in Smouse (1984), by considering the superpopulation model where, marginally,

$$E[y] = X\mu$$

and

$$\begin{aligned}\operatorname{Var}[y] &= E\{\operatorname{Var}[y|\theta]\} + \operatorname{Var}\{E[y|\theta]\} \\ &= \bar{V} + X\Omega X'\,.\end{aligned}$$

So, predictor (2.8) is also a Bayes linear predictor within the class of all linear predictors. Note however that (2.8) can not be argued for by using a direct Bayesian analysis involving prior and posterior distributions, as was done, for example, in Royall and Pfefferman (1982) or Bolfarine *et al.* (1987).

Example 2.1. Suppose that the elements of the superpopulation model (1.1) are such that $X = \mathbf{1}_N$ and $V(\theta) = \theta^2 I$, where I is the identity matrix of dimension N. Suppose also that $E[\theta] = \mu$ and that $\operatorname{Var}[\theta] = \Omega$. Hence, $\bar{V} = (\Omega + \mu^2)I = \upsilon I$, where $\upsilon = \mu^2 + \Omega$. It then follows from (2.8) and (2.9) that the Bayes linear predictor of T is

$$\hat{T}_L^* = n\bar{y}_s + (N-n)\frac{(n\bar{y}_s/v + \mu/\Omega)}{n/v + 1/\Omega},$$

with linear prediction variance

$$E[\hat{T}_L^* - T]^2 = (N-n)v + (N-n)^2\frac{1}{n/v + 1/\Omega}.$$

3. A connection with least squares prediction in mixed linear models

Now we turn to the least squares (in mixed linear models) interpretation of predictor (2.8). Consider the following mixed linear model (Theil (1971))

$$(3.1) \qquad \begin{pmatrix} \mu \\ \boldsymbol{y}_s \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{X}_s \end{pmatrix}\theta + \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e}_s \end{pmatrix},$$

where $\boldsymbol{u}$ and $\boldsymbol{e}_s$ are independent zero mean vectors with covariance matrices $\boldsymbol{\Omega}$ and $\bar{\boldsymbol{V}}_s$, respectively, and $\boldsymbol{I}$ is the identity matrix of dimension p. After some algebraic manipulations, the Gauss-Markov estimator of θ may be written as

$$(3.2) \qquad \hat{\theta} = \mu + \boldsymbol{F}(\boldsymbol{y}_s - \boldsymbol{X}_s\mu),$$

where

$$\boldsymbol{F} = \boldsymbol{\Omega}\boldsymbol{X}_s'(\boldsymbol{X}_s\boldsymbol{\Omega}\boldsymbol{X}_s' + \bar{\boldsymbol{V}}_s^{-1})^{-1} = (\boldsymbol{X}_s'\bar{\boldsymbol{V}}_s\boldsymbol{X}_s + \boldsymbol{\Omega}^{-1})^{-1}\boldsymbol{X}_s'\bar{\boldsymbol{V}}_s^{-1}.$$

Moreover,

$$(3.3) \qquad \mathrm{Var}[\hat{\hat{\theta}}] = (\boldsymbol{I} - \boldsymbol{F}\boldsymbol{X}_s)\boldsymbol{\Omega}.$$

Now, from (2.5) and (2.6) and any given sample s, it follows by using the Gauss-Markov theorem for mixed linear models that the variance of $\hat{\tau} - \boldsymbol{1}_r'\bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{y}_s$ is minimum among the variance of all linear unbiased estimators of $\boldsymbol{1}_r'\{\boldsymbol{X}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{X}_s\}\theta$ if

$$\begin{aligned}\hat{\tau} &= \boldsymbol{1}_r'\bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\boldsymbol{y}_s + \boldsymbol{1}_r'\{\boldsymbol{X}_r - \bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}\bar{\boldsymbol{X}}_s\}\hat{\theta} \\ &= \boldsymbol{1}_r'\boldsymbol{X}_r\hat{\hat{\theta}} + \boldsymbol{1}_r'\boldsymbol{X}_r\bar{\boldsymbol{V}}_{rs}\bar{\boldsymbol{V}}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s\hat{\hat{\theta}}).\end{aligned}$$

Therefore, the least squares predictor of T under the superpopulation model (1.1) and the mixed model (3.1) for the observed sample and the prior information is equal to the Bayes linear predictor of T. The linear prediction variance of the least squares predictor follows from (2.4) and (3.3) and it is easily shown to be equal to (2.9).

Hence, the above connection between the two approaches provides an alternative way of obtaining the Bayes linear predictor (2.8). In the next section use will be made of the above connection to derive the Bayes dynamic linear predictor when sampling on successive occasions.

4. Dynamic Bayesian linear prediction

In this section, it is considered that the finite population is observed t times. At time j, the observation equation is denoted by

$$y_j = X_j\theta_j + e_j \,, \tag{4.1}$$

while the state equation is denoted by

$$\theta_j = G_j\theta_{j-1} + w_j \,, \tag{4.2}$$

where e_j and w_j are independent zero mean vectors for all j, $\mathrm{Var}[e_j] = V_j(\theta_j)$, $\mathrm{Var}[w_j] = W_j$ and G_j are known, $j = 1,\ldots,t$. The observation equation (4.1) describes the way data is generated at time j, while the state equation (4.2) describes how θ_j evolves through time.

It is of interest predicting the population total, T_t, at time t. After a sample s_j of size n_j is selected from $\mathscr{P}_j$ (the finite population at time j), $j = 1,\ldots,t$ we seek among the class of all linear predictors of the form $\hat{T}_{Lt} = a + \sum_{j=1}^{t} h_j' y_{s_j}$, the one minimizing

$$E[\hat{T}_{Lt} - T]^2 \,, \tag{4.3}$$

where h_j is a known vector of appropriate dimensions and y_{s_j} denotes the observed sample at time j. As before, the subscript r_j is used to denote, at time j, the unobserved part of $\mathscr{P}_j, j = 1,\ldots,t$. The recursive Kalman filter linear predictor of T_t is now described under the superpopulation model (4.1) and (4.2), by making use of the least squares connection described in the previous section. Some other approaches where the population mean itself is considered to follow a stochastic structure are considered in the pioneering papers by Blight and Scott (1973) and Scott and Smith (1974). In those papers, time series methods have successfully been applied to the analysis of repeated surveys from random or probability samples. As pointed out in Binder and Hidiroglou (1988), Blight and Scott (1973) gave a recursive formulation for the Bayes estimation of the population mean which is equivalent to those obtained by using the Kalman filter estimating algorithm.

Starting with the first iteration, each iteration thereafter may be described as follows. Suppose that we have reached time t, so that before observing $\boldsymbol{y}_{s_t}$, we have $\hat{\hat{\theta}}_{t-1}$ and $\boldsymbol{C}_{t-1}$, the Bayes linear estimator of θ_{t-1} and its linear variance. Hence, prior to observing $\boldsymbol{y}_{s_t}$ our state of knowledge (prior information) about θ_t may be expressed as

$$\mu_t = \boldsymbol{G}_t\hat{\hat{\theta}}_{t-1} \quad \text{and} \quad \boldsymbol{\Omega}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t + \boldsymbol{W}_t ,$$

both equations following from equation (4.2). Now, in order to adapt the least squares connection to the present context, we identify, from the previous section, μ with μ_t, $\boldsymbol{\Omega}$ with $\boldsymbol{\Omega}_t$, $\boldsymbol{y}_s$ with $\boldsymbol{y}_{s_t}$, θ with θ_t and $\boldsymbol{V}(\cdot)$ with $\boldsymbol{V}_t(\cdot)$. Therefore, by appealing to the least squares connection, we write the following mixed linear model for the observed sample and prior information at time t:

$$\text{(4.4)} \qquad \begin{pmatrix} \mu_t \\ \boldsymbol{y}_{s_t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_t \\ \boldsymbol{X}_{s_t} \end{pmatrix} \theta_t + \begin{pmatrix} \boldsymbol{u}_t \\ \boldsymbol{e}_{s_t} \end{pmatrix},$$

where the covariance matrices of the independent zero mean vectors $\boldsymbol{u}_t$ and $\boldsymbol{e}_{s_t}$ are $\boldsymbol{\Omega}_t$ and $\bar{\boldsymbol{V}}_{s_t}$, respectively. So, by making use of (3.2) and (3.3), it follows that the Gauss-Markov estimator of θ_t and its variance, under the mixed linear model (4.4), are given by

$$\text{(4.5)} \qquad \hat{\hat{\theta}}_t = \mu_t + \boldsymbol{F}_t(\boldsymbol{y}_{s_t} - \boldsymbol{X}_{s_t}\mu_t)$$

and

$$\text{(4.6)} \qquad \boldsymbol{C}_t = [\boldsymbol{I}_t - \boldsymbol{F}_t\boldsymbol{X}_{s_t}]\boldsymbol{\Omega}_t ,$$

where $\boldsymbol{F}_t = \boldsymbol{\Omega}_t\boldsymbol{X}'_{s_t}[\boldsymbol{X}_{s_t}\boldsymbol{\Omega}_t\boldsymbol{X}'_{s_t} + \bar{\boldsymbol{V}}_t]^{-1}$. Then, by using (2.8) at time t, it follows that the Bayes linear predictor of T_t at time t is

$$\text{(4.7)} \qquad \hat{T}^*_{Lt} = n_t\bar{y}_{s_t} + \boldsymbol{1}'_r\{\boldsymbol{X}_{r_t}\hat{\hat{\theta}}_t + \bar{\boldsymbol{V}}_{r_ts_t}\bar{\boldsymbol{V}}_{s_t}^{-1}(\boldsymbol{y}_{s_t} - \boldsymbol{X}_{s_t}\hat{\hat{\theta}})\} ,$$

where $\bar{y}_{s_t}$ is the sample mean at time t and with linear prediction variance given by

$$\text{(4.8)} \qquad E[\hat{T}^*_{Lt} - T]^2 = \boldsymbol{1}'_{r_t}\{\bar{\boldsymbol{V}}_{r_t} - \bar{\boldsymbol{V}}_{r_ts_t}\bar{\boldsymbol{V}}_{s_t}^{-1}\bar{\boldsymbol{V}}_{s_tr_t}\}\boldsymbol{1}_{r_t}$$
$$+ \boldsymbol{1}'_{r_t}\{\boldsymbol{X}_{r_t} - \bar{\boldsymbol{V}}_{r_ts_t}\bar{\boldsymbol{V}}_{s_t}^{-1}\boldsymbol{X}_{s_t}\}\boldsymbol{C}_t\{\boldsymbol{X}_{r_t} - \bar{\boldsymbol{V}}_{r_ts_t}\bar{\boldsymbol{V}}_{s_t}^{-1}\boldsymbol{X}_{s_t}\}'\boldsymbol{1}_{r_t} .$$

Remark 4.1. Note that predictor (4.7) is indeed a more general (and dynamic) version of the least squares predictor proposed by Royall (1976). Moreover, a less general version ($\boldsymbol{V}_t$ fixed and known) of predictor (4.7) was obtained by Bolfarine (1989) by making use of a dynamic multivariate normal superpopulation model. Therefore, predictor (5) in Bolfarine (1989), and the more particular predictor (12) in Bolfarine (1988) are indeed Bayes linear predictors, if the normality assumption is dropped. One difference between the two approaches is that the algorithm in Bolfarine (1988, 1989) updates the conditional error covariance matrix, whereas the present algorithm updates the unconditional error covariance matrix.

Example 4.1. Consider at time j, the observation equation

$$y_{jk} = \theta_j + e_{jk} , \tag{4.9}$$

where the e_{jk} are all independent, $E[e_{jk}] = 0$ and $\boldsymbol{V}_j(\theta_j) = \theta_j \boldsymbol{I}_j, k = 1,\ldots, N_j$, $j = 1,\ldots, t$. The state equation at time j is considered to be

$$\theta_j - \mu = \rho(\theta_{j-1} - \mu) + w_j , \tag{4.10}$$

where w_j are all independent (also independent of e_{jk}) with $E[w_j] = 0$ and Var $[w_j] = W_j$, a known and positive constant. Equations (4.9) and (4.10) characterize the well-known steady-state Poisson forecasting model, which is a special case of the model of Blight and Scott (1973). They assume an independent $AR(1)$ process on the error term $\boldsymbol{e}$, for each k. Now, by using (4.5)–(4.8), it follows that the Bayes linear predictor of the population total at time t, T_t, given $\hat{\hat{\theta}}_{t-1}$ and C_{t-1}, is

$$\hat{T}^*_{Lt} = n\bar{y}_{s_t} + (N_t - n_t)\hat{\hat{\theta}}_t ,$$

with prediction variance given by

$$E[\hat{T}^*_{Lt} - T]^2 = (N_t - n_t)\upsilon_t + (N_t - n_t)^2 C_t ,$$

where $\upsilon_t = E[\theta] = \bar{\theta}$, the mean level of the process,

$$\hat{\hat{\theta}} = \mu_t + \boldsymbol{F}_t(\boldsymbol{y}_{s_t} - \mathbf{1}_{s_t}\mu_t) = \mu_t + \frac{n_t/\upsilon_t}{n_t/\upsilon_t + 1/\Omega_t}(\bar{y}_{s_t} - \mu_t) ,$$

$$C_t = \frac{1}{n_t/\upsilon_t + 1/\Omega_t} ,$$

$$\Omega_t = \rho^2 C_{t-1} + W_t \quad \text{and} \quad \mu_t = \rho\hat{\hat{\theta}}_{t-1} + (1 - \rho)\bar{\theta} .$$

Acknowledgements

Helpful suggestions by a referee substantially improved the paper's presentation. Partial financial support from CNPq-Brasil is also acknowledged.

REFERENCES

Bellhouse, D. (1987). Model based inference in finite population sampling, *Amer. Statist.*, **41**, 260–262.

Binder, D. A. and Hidiroglou, M. A. (1988) Sampling in time, *Handbook of Statistics*, **6**, Sampling, (ed. P. R. Krishnaiah and C. R. Rao), North-Holland, Amsterdam.

Blight, B. G. N. and Scott, A. J. (1973). A stochastic model for repeated surveys, *J. Roy. Statist. Soc. Ser. B*, **35**, 61–66.

Bolfarine, H. (1988). Finite population prediction under dynamic generalized linear superpopulation models, *Comm. Statist. B—Simulation Comput.*, **17**, 187–208.

Bolfarine, H. (1989). Population variance prediction under dynamic normal superpopulation models, *Statist. Probab. Lett.*, **8**, 35–39.

Bolfarine, H., Pereira, C. A. B. and Rodrigues, J. (1987). Robust linear prediction in finite populations. A Bayesian perspective, *Sankhyā Ser. B*, **49**, 23–35.

Fuller, W. (1970). Simple estimators for the mean of skewed populations, Tech. Report U. S. Bureau of the Census.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Rodrigues, J. (1989). The coordinate free estimation in finite population sampling, *Statist. Probab. Lett.*, **7**, 293–295.

Royall, R. (1976). The least squares prediction approach to two stage sampling, *J. Amer. Statist. Assoc.*, **71**, 657–664.

Royall, R. and Pfefferman, D. (1982). Balanced samples and robust Bayesian inference in finite population sampling, *Biometrika*, **69**, 401–409.

Scott, A. J. and Smith, T. F. M. (1974). Analysis of repeated surveys using time series methods, *J. Amer. Statist. Assoc.*, **69**, 674–678.

Smouse, E. (1984). A note on Bayesian least squares inference for finite population models, *J. Amer. Statist. Assoc.*, **79**, 390–392.

Theil, H. (1971). *Principles of Econometrics*, Wiley, New York.