RT-MAE-9111

OPTIMAL PREDICTION OF THE FINITE POPULATION REGRESSION COEFFICIENT IN FINITE POPULATIONS

by

H. Bolfarine, S.N. Elian,

J. Rodrigues

and

S. Zacks

Palavras Chaves: Finite populations; optimal predictors; finite population regression coefficient; weighted least squares
(Key words)

Classificação AMS: 62D05
(AMS Classification)

- Junho de 1991 -

OPTIMAL PREDICTION OF THE FINITE POPULATION REGRESSION COEFFICIENT IN FINITE POPULATIONS

H. Bolfarine, S. N. Elian, J. Rodrigues* and S. Zacks**

IME-USP, ICMSC-USP* and SUNY, Binghamton

Departamento de Estatistica-IME

Universidade de São Paulo

Caixa Postal 20570- CEP 01498 (Agencia Iguatemi)

São Paulo- Brasil

Abstract

In this paper we investigate optimal prediction of the finite population regression coefficient β_N under a general linear regression superpopulation model. Optimal predictors are obtained under Gaussian superpopulation models and also under weaker Gauss-Markov type assumptions. We derive the optimum linear predictor of β_N under the general linear model with a nondiagonal covariance matrix and show that it reduces to the usual least squares estimator of the superpopulation regression coefficient. A linear Bayesian approach for predicting β_N is also proposed.

1. Introduction

Let $\mathcal{P} = \{1, \ldots, N\}$ denote a finite population of N units, where N is known. Associated with the k–th unit of $\mathcal{P}$, there are $p+1$ quantities, $y_k, x_{k1}, \ldots, x_{kp}$, where all but y_k are known, $k = 1, \ldots, N$. Let $\mathbf{y} = (y_1, \ldots, y_N)'$ and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)'$, where the $\mathbf{X}_k = (x_{k1}, \ldots, x_{kp})'$, $k = 1, \ldots, N$. Relating the two sets of variables, we consider the linear model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \tag{1}$$

where $\mathbf{e} \sim N(0, \mathbf{V})$. Let $\psi = (\beta, \mathbf{V})$ denote the parameter in model (1). Note that model (1), as stated, is completely characterized by $(\mathbf{X}; \mathbf{V})$. Throughout the paper, model (1) is considered to be known.

Let $\theta(\mathbf{y})$ be a population quantity of interest. Examples of such quantities are: the population total, $T = \mathbf{1}'_N\mathbf{y} = \sum_{i=1}^N y_i$, where $\mathbf{1}_N$ is a vector of ones of dimension N, and the finite population regression coefficient

$$\beta_N = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{2}$$

Optimal prediction of T is considered in Pereira and Rodrigues (1983) and Rodrigues et al. (1985). See also Bolfarine and Zacks, (1991). A sample $\mathbf{s}$ of size n is selected from $\mathcal{P}$ according to some specified sampling plan in order to obtain information on β_N, the quantity we want to predict. Let $\mathbf{r} = \mathcal{P} - \mathbf{s}$ be the unobserved part of $\mathcal{P}$. After the sample $\mathbf{s}$ has been selected, we may reorder the elements of $\mathbf{y}$ so that we have the corresponding partitions of $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{V}$; that is,

$$\begin{pmatrix} \mathbf{y}_s \\ \mathbf{y}_r \end{pmatrix}, \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{pmatrix} \text{ and } \begin{pmatrix} \mathbf{V}_s & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_r \end{pmatrix}.$$

Estimation of β_N has been the subject of several papers in the recent statistical literature. It is worth mentioning the work of Fuller (1975), where asymptotic properties of some predictors of β_N are studied in the diagonal case. Some other important references on the subject are Konijn (1962), Hidiroglou (1974) and Shah et al. (1977). In Section 2 we consider best unbiased prediction of the finite population regression coefficient β_N under the assumption that $\mathbf{e}$ in (1) is normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{V}$, that we denote by $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V})$, with $\mathbf{V}_{rs} = \mathbf{0}$. The prediction variance of the optimal predictor is also derived. In Section 3 the results of Section 2 are extended to the case of nondiagonal covariance matrices, that is, $\mathbf{V}_{sr} \neq \mathbf{0}$. Section 4 is devoted to linear prediction of β_N, where we only make assumptions on the first and second moments of the distribution of $\mathbf{y}$, as it follows from (1). As expected, the results are similar to the ones derived under normality. We also show that the optimal predictor in the nondiagonal case can take a simple and more intuitive form. Indeed, we show that it reduces to the weighted least squares estimator of the superpopulation regression coefficient β. In Section 5 we consider Bayesian linear prediction of β_N, where we only make assumptions on the first and second moments of the joint distribution of β and $\mathbf{y}$.

2. Optimal prediction of β_N under normality. Diagonal covariance matrices

In the case of the superpopulation model (1) with $\mathbf{V}_{rs} = \mathbf{0}$, one can write β_N as

(3)
$$\beta_N = \mathbf{A}_s\hat{\beta}_s + \mathbf{A}_r\beta_r,$$

where $\hat{\beta}_s$ is the weighted least squares estimator of β, namely,

$$\hat{\beta}_s = \mathbf{B}_s\mathbf{y}_s,$$

$$\mathbf{B}_s = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1},$$

$$\mathbf{A}_s = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s$$

and

$$\mathbf{A}_r = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}_r'\mathbf{V}_r^{-1}\mathbf{X}_r.$$

β_r is the "weighted least squares estimator" of β based on $\mathbf{X}_r$, $\mathbf{V}_r$ and $\mathbf{y}_r$, that is, $\beta_r = (\mathbf{X}_r'\mathbf{V}_r^{-1}\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{V}_r^{-1}\mathbf{y}_r$. Notice that $\mathbf{A}_s + \mathbf{A}_r = \mathbf{I}_p$, the identity matrix of dimension p. Since $\mathbf{y}_r$ has not been observed, β_r is treated as an unknown vector valued quantity. A predictor of β_N, can be written in the form

$$\hat{\beta}_N = \mathbf{A}_s\hat{\beta}_s + \mathbf{A}_r\hat{\beta}_r,$$

where $\hat{\beta}_r$ is a predictor of β_r based on $\mathbf{y}_s$.

Definition 2.1. *A predictor $\hat{\beta}_N$ of β_N is unbiased if and only if*

$$E_\psi[\hat{\beta}_N - \beta_N] = 0,$$

for all ψ.

Notice that $\hat{\beta}_N$ is an unbiased predictor of β_N if and only if $\hat{\beta}_r$ is an unbiased predictor of β_r.

Lemma 2.1. *$\hat{\beta}_s$ is an unbiased predictor of β_N.*

Definition 2.2. *The generalized prediction mean squared error (GMSE) of a predictor $\hat{\beta}_N$ of β_N is*

$$GMSE_\psi[\hat{\beta}_N] = E_\psi[\lambda'(\hat{\beta}_N - \beta_N)(\hat{\beta}_N - \beta_N)'\lambda],$$

for any px1 real vector λ.

Definition 2.3. *$\hat{\beta}_{BUN}$ is the best unbiased predictor of β_N if $\hat{\beta}_{BUN}$ is unbiased and*

$$GMSE_\psi[\hat{\beta}_{BUN}] \leq GMSE_\psi[\hat{\beta}_N],$$

for any other unbiased predictor $\hat{\beta}_N$ and for all ψ.

Theorem 2.1. *Under the superpopulation model (1) with $\mathbf{e} \sim N(0, \mathbf{V})$, the best unbiased predictor of β_N is*

$$\hat{\beta}_{BUN} = \hat{\beta}_s. \tag{4}$$

Furthermore,

$$GMSE_\psi[\hat{\beta}_s] = \lambda'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}_r'\mathbf{V}_r^{-1}\mathbf{X}_r(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\lambda. \tag{5}$$

Proof. Let $\hat{\beta}_{*N} = \mathbf{A}_s\hat{\beta}_s + \mathbf{A}_r\hat{\beta}_{*r}$ be any unbiased predictor of β_N. According to Arnold (1981), if $\mathbf{V}$ is known, $\hat{\beta}_s$ is a complete and sufficient statistics. Moreover, since $\mathbf{V}_{rs} = 0$, $\mathbf{y}_s$ is independent of $\mathbf{y}_r$, which implies that $\hat{\beta}_s$ is also totally sufficient, according to the Definition 3 in Rodrigues et al. (1985). We may then write

$$\begin{aligned} E_\psi[\lambda'(\hat{\beta}_{*N} - \beta_N)(\hat{\beta}_{*N} - \beta_N)'\lambda] &= Var_\psi[\lambda'(\hat{\beta}_{*N} - \beta_N)] \\ &\geq Var\{\lambda' E_\psi[(\hat{\beta}_{*N} - \beta_N)|\mathbf{y}_r, \hat{\beta}_s]\} \\ &= \lambda'\mathbf{A}_r Var_\psi\{E_\psi[\hat{\beta}_{*r}|\mathbf{y}_r, \hat{\beta}_s] - \beta_r\}\mathbf{A}_r'\lambda. \end{aligned}$$

Hence, since $\hat{\beta}_s$ is predictive sufficient,

$$E_\psi[\hat{\beta}_{*r}|\mathbf{y}_r, \hat{\beta}_s] = \hat{\beta}_s.$$

This proves (4). Uniqueness follows from the completeness of $\hat{\beta}_s$. To compute the generalized prediction MSE of $\hat{\beta}_s$, notice first that

$$GMSE_\psi[\hat{\beta}_s] = \lambda'\mathbf{A}_r Var[\hat{\beta}_s - \beta_r]\mathbf{A}_r'\lambda.$$

Hence (5) follows from the fact that

$$Var_\psi[\hat{\beta}_s - \beta_r] = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1} + (\mathbf{X}_r'\mathbf{V}_r^{-1}\mathbf{X}_r)^{-1}.$$

3. Optimal prediction under normality. Nondiagonal covariance matrices

In this section, we consider the case where the covariance matrix $\mathbf{V}$ is not diagonal. After some algebraic manipulations, it can be shown that the we can write

$$\beta_N = \mathbf{H}^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{y}_s + \mathbf{H}^{-1}\mathbf{D}\mathbf{E}^{-1}\mathbf{y}_r, \tag{6}$$

where

$$\mathbf{B} = \mathbf{X}_s' - \mathbf{X}_r'\mathbf{V}_r^{-1}\mathbf{V}_{rs}, \quad \mathbf{C} = \mathbf{V}_s - \mathbf{V}_{sr}\mathbf{V}_r^{-1}\mathbf{V}_{rs},$$

$$\mathbf{D} = \mathbf{X}_r' - \mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{V}_{sr}, \quad \mathbf{E} = \mathbf{V}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{V}_{sr},$$

and

$$\mathbf{H} = \mathbf{B}\mathbf{C}^{-1}\mathbf{X}_s + \mathbf{D}\mathbf{E}^{-1}\mathbf{X}_r.$$

It may be noted that if $\mathbf{V}_{rs} = 0$, then β_N in (6) reduces to the simpler expression (3). Moreover, the $GMSE$ of any predictor $\hat{\beta}_N$ of β_N may be written as

$$E_\psi[\lambda'(\hat{\beta}_N - \beta_N)(\hat{\beta}_N - \beta_N)'\lambda] = E_\psi\{[\lambda'(\hat{\beta}_N - \beta_N)]^2\}, \tag{7}$$

a result which makes it possible to derive the optimal predictor of β_N in this more general context. The main result of this section is stated in the next theorem. We emphasize that the totally sufficient approach considered in the last section only applies to the case where $\mathbf{V}_{rs} = 0$. Let $\mathbf{Q}_s = \mathbf{H}^{-1}\mathbf{B}\mathbf{C}^{-1}$ and $\mathbf{Q}_r = \mathbf{H}^{-1}\mathbf{D}\mathbf{E}^{-1}$, so that we may write $\mathbf{Q} = (\mathbf{Q}_s, \mathbf{Q}_r)$.

Theorem 3.1. *Under the superpopulation model (1) with $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V})$, where $\mathbf{V}$ is not necessarilly diagonal, it follows that the BUP of β_N is given by*

$$\hat{\beta}_{BUN} = \mathbf{Q}_s\mathbf{y}_s + \mathbf{Q}_r[\mathbf{X}_r\hat{\beta}_s + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s)]. \tag{8}$$

Furthermore, the GMSE of $\hat{\beta}_{BUN}$ is given by

$$\begin{aligned} GMSE_\psi[\hat{\beta}_{BUN}] &= \lambda'\mathbf{H}^{-1}\mathbf{D}\mathbf{E}^{-1}\mathbf{D}'\mathbf{H}^{-1}\lambda \\ &\quad + \lambda'\mathbf{Q}_r\mathbf{D}'(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{D}\mathbf{Q}_r'\lambda \\ &= \lambda'\mathbf{Q}_r[\mathbf{E} + \mathbf{D}'(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{D}]\mathbf{Q}_r'\lambda. \end{aligned} \tag{9}$$

Proof. We may write

$$\lambda'\beta_N = \lambda'\mathbf{Q}\mathbf{y},$$

where $\mathbf{Q} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$. Let $\hat{\beta}_N$ be any unbiased predictor of β_N. Ordinary, but length algebraic manipulations yield the following result

$$E_\psi\{[\lambda'\mathbf{Qy} - \lambda'\hat{\beta}_N]^2\} = E_\psi\{[\lambda'\mathbf{Qy} - E_\psi[\lambda'\mathbf{Qy}|\mathbf{y}_s]]^2\} + E_\psi\{[\lambda'\hat{\beta}_N - E_\psi[\lambda'\mathbf{Qy}|\mathbf{y}_s]]^2\} \quad (10)$$

From (10), it follows that the minimum mean squared error predictor of $\lambda'\mathbf{Qy}$ is also the minimum mean squared error estimator of

$$E_\psi[\lambda'\mathbf{Qy}|\mathbf{y}_s] = \lambda'\mathbf{Q}_s\mathbf{y}_s + \lambda'\mathbf{Q}_r E_\psi[\mathbf{y}_r|\mathbf{y}_s]. \quad (11)$$

Thus, if β were known, the best predictor of $\mathbf{Qy}$ would be given by $E[\mathbf{Qy}|\mathbf{y}_s]$. But, unfortunately, this is not the case. Under the assumption that the error vector $\mathbf{e}$ is Gaussian, it follows that

$$E_\psi[\mathbf{y}_r|\mathbf{y}_s] = \mathbf{X}_r\beta + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\beta). \quad (12)$$

Note that when no distributional assumptions is made for model (1), the computation of the expectation $E_\psi[\mathbf{y}_s|\mathbf{y}_r]$ will be extremely complicated. Using (11) and (12), it follows that the predictor of $\lambda'\mathbf{Qy}$ which minimizes (10) must be of the form

$$\lambda'\hat{\beta}_{BUN} = \lambda'\mathbf{Q}_s\mathbf{y}_s + \lambda'\mathbf{Q}_r[\mathbf{X}_r\hat{\beta}_s + \mathbf{V}_{sr}\mathbf{V}_s^{-1}(\mathbf{Y}_s - \mathbf{X}_s\hat{\beta}_s)],$$

where, under normality,
$$\hat{\beta}_s = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s$$

is the best unbiased estimator of β. This proves (8). Furthermore, since $\hat{\beta}_{BUN}$ is unbiased,

$$GMSE_\psi[\hat{\beta}_{BUN}] = E_\psi\{[\lambda'(\hat{\beta}_{BUN} - \beta_N)]^2\} = Var_\psi[\lambda'(\hat{\beta}_{BUN} - \beta_N)]$$
$$= \lambda'\mathbf{Q}_r Var_\psi[\mathbf{X}_r\hat{\beta}_s + \mathbf{V}_{sr}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s) - \mathbf{y}_r]\mathbf{Q}_r'\lambda.$$

Hence, (9) follows from the fact that

$$Var_\psi[\mathbf{X}_r\hat{\beta}_s + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s) - \mathbf{y}_r]$$
$$= Var_\psi[(\mathbf{X}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{X}_s)\hat{\beta}_s + \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{y}_s - \mathbf{y}_r]$$
$$= \mathbf{V}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{V}_{sr} + (\mathbf{X}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{X}_s)(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}(\mathbf{X}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{X}_s)'.$$

4. Optimal Linear Prediction

We now consider the more general situation where no assumption, besides that

$$E[\mathbf{e}] = 0 \quad \text{and} \quad Var[\mathbf{e}] = \mathbf{V}, \quad (13)$$

is made about the distribution of the error vector **e**. This are the usual Gauss-Markov type assumptions. As mentioned in the previous section, in this case, the computation of $E_\psi[\mathbf{y}_r|\mathbf{y}_s]$ is extremely complicated. One way of countering this difficulty is to consider predictors of $\mathbf{y}_r$ which are linear in $\mathbf{y}_s$. Let

$$\hat{\beta}_{LN} = \mathbf{M}'\mathbf{y}_s$$

be any linear predictor of β_N, where **M** is a *pxn* matrix of known entries. Further, as before, let $\mathbf{Q}_s = \mathbf{H}^{-1}\mathbf{B}\mathbf{C}^{-1}$, and $\mathbf{Q}_r = \mathbf{H}^{-1}\mathbf{D}\mathbf{E}^{-1}$.

Lemma 4.1. *A linear predictor $\hat{\beta}_{LN} = \mathbf{M}'\mathbf{y}_s$ is unbiased for β_N if and only if*

$$\mathbf{M}'\mathbf{X}_s = \mathbf{I}_p.$$

As a consequence of Lemma 4.1, we have that $\hat{\beta}_{BUN}$ proposed in the previous section is a (ψ-)unbiased predictor of β_N. Note also that the simple predictor $\hat{\beta}_s$, which is a linear predictor, is also unbiased. Let $\mathbf{M}'\mathbf{y}_s$ be any linear predictor of β_N, where **M** is an *nxp* matrix. The proof of the following result is merely based on algebraic manipulations and so, it is omitted.

Lemma 4.2. *Under the above assumptions, for any px1 vector λ, and any nxp matrix* **M**,

(14)
$$E_\psi\{[\lambda'\mathbf{M}'\mathbf{y}_s - \lambda'\mathbf{Q}\mathbf{y}]^2\} = Var_\psi[\lambda'(\mathbf{M} - \mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1})\mathbf{y}_s] + \lambda'\mathbf{Q}_r\mathbf{V}_r\mathbf{Q}_r'\lambda - \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{V}_{sr}\mathbf{Q}_r'\lambda + \{\lambda'(\mathbf{M}'\mathbf{X}_s - \mathbf{I}_p)\beta\}^2.$$

Notice further that if $\lambda'\mathbf{M}'\mathbf{y}_s$ minimizes (14), then $\mathbf{M}'\mathbf{y}_s$ minimizes the GMSE given in the Definition 2.2. The main result of this section is stated next.

Theorem 4.1. *Under the Gauss-Markov assumptions stated above, the unbiased linear predictor with minimum GMSE, $\hat{\beta}_{BLN}$, is as given in (8), that is, $\hat{\beta}_{BLN} = \hat{\beta}_{BUN}$, with GMSE as given in (9).*

Proof. Let $\hat{\beta}_{LN} = \mathbf{M}'\mathbf{y}_s$ be any linear unbiased predictor of β_N. According to Lemma 4.1, since $\hat{\beta}_{LN}$ is unbiased,

$$\lambda'\mathbf{M}'\mathbf{X}_s\beta - \lambda'\mathbf{Q}\mathbf{X}\beta = 0,$$

for all λ and β, so that the last term on the right hand side of the expression (14) is zero and $\mathbf{M}'\mathbf{X}_s = \mathbf{Q}\mathbf{X}$. Note also that $\mathbf{Q}\mathbf{X} = \mathbf{I}_p$. Therefore, it follows from (7) and (14) that to find the linear unbiased predictor for β_N with minimum GMSE, it is equivalent to find a predictor which is unbiased for

(15)
$$E_\psi[\lambda'\mathbf{M}'\mathbf{y}_s - \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{y}_s] = (\lambda'\mathbf{Q}\mathbf{X} - \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{X}_s)\beta$$

and has minimum variance in the class of all linear unbiased predictors. Hence, by the Gauss-Markov Theorem (Rao, 1973), it follows that the best linear unbiased estimator of the expectation (15) is given by

$$\lambda'\mathbf{M}_s'\mathbf{y}_s - \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{y}_s = \lambda'(\mathbf{Q}\mathbf{X} - \mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{X}_s)\hat{\beta}_s, \tag{16}$$

where $\hat{\beta}_s$ is the usual least squares estimator of β given in (3). From (16), it follows that

$$\begin{aligned}\lambda'\mathbf{M}_s'\mathbf{y}_s &= \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{y}_s + \lambda'(\mathbf{Q}\mathbf{X} - \mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{X}_s)\hat{\beta}_s \\ &= \lambda'\mathbf{Q}_s\mathbf{y}_s + \lambda'\mathbf{Q}_r[\mathbf{X}_r\hat{\beta}_s + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s)],\end{aligned}$$

minimizes the meam squared error (14), for all $\lambda \in \mathcal{R}^p$. Notice that $\mathbf{Q}\mathbf{X} = \mathbf{Q}_s\mathbf{X}_s + \mathbf{Q}_r\mathbf{X}_r$. Thus,

$$\hat{\beta}_{BLN} = \mathbf{M}_s'\mathbf{y}_s = \mathbf{Q}_s\mathbf{y}_s + \mathbf{Q}_r[\mathbf{X}_r\hat{\beta}_s + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s)],$$

is the predictor of β_N with minimum GMSE, that is, it minimizes $E_\psi\{[\lambda'(\hat{\beta}_{LN} - \hat{\beta}_N)]^2\}$ in the class of all linear unbiased predictors, as was to be proved. The proof that the GMSE of $\hat{\beta}_{BLN}$ is given by (9) is similar to the one presented in Theorem 3.1.

Example 4.1. In this example, we consider the superpopulation model (1) where

$$\mathbf{X} = \mathbf{1}_N \quad \text{and} \quad \mathbf{V} = (1-\rho)\mathbf{I}_N + \rho\mathbf{1}_N\mathbf{1}_N'.$$

$\mathbf{I}_N$ is the identity matrix of order N and $\mathbf{1}_N$ is a vector of ones of dimension N. It can be checked that

$$\beta_N = \bar{y},$$

the population mean. Moreover, $\hat{\beta}_s = \bar{y}_s$, the sample mean. After mildly tedious algebraic manipulations, it can be shown by using (8) that

$$\begin{aligned}\hat{\beta}_{BLN} &= \frac{(1+(N-1)\rho)}{N}\{\frac{1}{1+(N-n-1)\rho}[1 - \frac{\rho n}{1+(N-1)\rho}n\bar{y}_s] \\ &+ \frac{1}{1+(n-1)\rho}[1 - \frac{\rho(N-n)}{1+\rho(N-1)}]\mathbf{1}_{N-n}'\mathbf{1}_{N-n}\bar{y}_s\} \\ &= \frac{1+(N-1)\rho}{N}\left[\frac{n}{1+(N-1)\rho}\bar{y}_s + \frac{(N-n)}{1+(N-1)\rho}\bar{y}_s\right] = \bar{y}_s,\end{aligned}$$

which coincides with the weighted least squares estimator $\hat{\beta}_s$. Indeed, we show next that this result holds in general.

In the next theorem, we show that the optimal predictor $\hat{\beta}_{BLN}$ (or β_{BUN}) always coincides with the weighted least squares estimator $\hat{\beta}_s$, that is, the optimal predictor always takes a simple form. Bolfarine and Rodrigues (1988) have shown that this is not the case with respect to the optimal predictor of the population total T under the general linear model considered above. Only in some special cases the optimal predictor of T can be written as a simple projection predictor.

Theorem 4.3. *Under the general linear model defined by (1) and (13), we have that*

$$\hat{\beta}_{BLN} = \hat{\beta}_s.$$

Furthermore,

$$GMSE[\hat{\beta}_s] = \lambda'[(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}]\lambda. \tag{17}$$

Proof. Since $\mathbf{QX} = \mathbf{I}_p$, we may write

$$\begin{aligned}\hat{\beta}_{BLN} - \hat{\beta}_s &= \mathbf{Q}_s\mathbf{y}_s + \mathbf{Q}_r[\mathbf{X}_r\hat{\beta}_s + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s)] - \mathbf{QX}\hat{\beta}_s\\ &= (\mathbf{Q}_s + \mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1})(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s)\\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s) = 0,\end{aligned}$$

which proves the first part of the result. On the other hand, (17) follows from the fact that

$$Var_\psi[\hat{\beta}_s] = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1},$$

$$Var_\psi[\beta_N] = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

and

$$Cov_\psi(\hat{\beta}_s;\beta_N) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

The following result is a direct consequence of the above theorem.

Corollary 4.1. *Under the assumptions of Theorem 4.3, we have*

$$\lambda'\mathbf{Q}_r[\mathbf{E} + \mathbf{D}'(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{D}]\mathbf{Q}_r'\lambda = \lambda'[(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}]\lambda.$$

Example 4.2. Two stage sampling. In this example, we assume that the population is divided into H clusters of sizes $N_1, \ldots, N_H$. In the first stage a sample $\mathbf{s}$ of n clusters is selected. In the second stage, a sample $\mathbf{s}_j$ of size m_j is selected from cluster j, for all $j \in \mathbf{s}$. In the notation of model (1), we have that

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_H \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{1}_{N_1} \\ \vdots \\ \mathbf{1}_{N_H} \end{pmatrix}$$

and $\mathbf{V} = \mathrm{diag}\{\mathbf{V}_1, \dots, \mathbf{V}_H\}$, where

$$\mathbf{V}_h = \sigma_h^2 \mathbf{I}_{N_h} + \sigma_v^2 \mathbf{J}_{N_h},$$

and $\mathbf{J}_{N_h} = \mathbf{1}_{N_h}\mathbf{1}'_{N_h}$, $h = 1, \dots, H$. The optimal predictor for the population total T has been derived by Royall (1976), which is a somewhat complex expression. After some mild amount of computations, it follows that, in this case,

$$\beta_N = \frac{\sum_{h=1}^{H} \frac{N_h \bar{y}_h}{\sigma_h^2 + N_h \sigma_v^2}}{\sum_{h=1}^{H} \frac{N_h}{\sigma_h^2 + N_h \sigma_v^2}}.$$

Furthermore, after some extensive algebraic manipulations, it can be shown that the optimal predictor of β_N is given by

$$(18)\quad \hat{\beta}_{BLN} = \Big(\sum_{h=1}^{H} \frac{N_h}{\sigma_h^2 + N_h\sigma_v^2}\Big)^{-1} \Big\{ \sum_{h \in s} \frac{m_h \bar{y}_{sh}}{\sigma_h^2 + N_h \sigma_v^2} + \sum_{h \in s} \frac{(N_h - m_h)(\sigma_v^2 m_h \bar{y}_{sh} + \sigma_h^2 \hat{\beta}_s)}{(\sigma_h^2 + m_h\sigma_v^2)(\sigma_h^2 + N_h \sigma_v^2)} \Big\}$$

$$+ \hat{\beta}_s \sum_{h \notin s} \frac{N_h}{\sigma_h^2 + N_h \sigma_v^2},$$

which, according to Theorem 4.3, is equal to

$$\hat{\beta}_s = \frac{\sum_{h \in s} w_h \bar{y}_{sh}}{\sum_{h \in s} w_h},$$

where

$$\bar{y}_{sh} = \frac{1}{m_h} \sum_{j \in s_h} y_{hj} \quad \text{and} \quad w_h = \frac{m_h \sigma_v^2}{\sigma_h^2 + m_h \sigma_v^2}.$$

Moreover the GMSE of the two predictors, which according to (9) and (17) are given by

$$GMSE_\psi[\hat{\beta}_s] = \lambda^2 \{ (\sum_{i \in s} \frac{m_h}{\sigma_h^2 + m_h \sigma_v^2})^{-1} - (\sum_{h=1}^{H} \frac{N_h}{\sigma_h^2 + N_h \sigma_v^2})^{-1} \},$$

and

$$GMSE_\psi[\hat{\beta}_{BLN}] = \lambda^2 (\sum_{h=1}^{H} \frac{N_h}{\sigma_h^2 + N_h \sigma_v^2})^{-2} \{ S + (\sum_{h \in s} \frac{m_h}{\sigma_h^2 + m_h \sigma_v^2})^{-1} S^2 \},$$

where

$$S = \sum_{h \in s} \frac{N_h}{\sigma_h^2 + N_h \sigma_v^2} + \sum_{h \in s} \frac{\sigma_h^2 (N_h - m_h)}{(\sigma_h^2 + m_h \sigma_v^2)(\sigma_h^2 + N_h \sigma_v^2)},$$

also coincides.

5. Bayes Linear Prediction

The main purpose of the present section is to introduce a Bayesian least squares approach for predicting β_N. The approach is characterized by the fact that although being a Bayesian approach it depends only on the first and second moments of the distributions of $\mathbf{y}$ and β. On the contrary, the ordinary Bayesian approach depends on the full joint distribution of $\mathbf{y}$ and β. Our approach is different from the approach of Smouse (1984), in the sense that the predictors to be considered are restricted to be unbiased with respect to the joint distribution of $\mathbf{y}$ and β. Whereas this approach has been considered (La Motte, 1978) for estimating superparameters, the application here is on prediction of β_N, which is a function of the characteristics of the finite population.

We now rewritte model (1) as the conditional superpopulation model characterized by

(18)
$$E[\mathbf{y}|\beta] = \mathbf{X}\beta$$
$$Var[\mathbf{y}|\beta] = \mathbf{V},$$

where, as before, $\mathbf{V}$ is a known matrix. Moreover, suppose that the investigators' prior knowledge about β provides

(19)
$$E[\beta] = \mu$$
$$Var[\beta] = \Sigma,$$

where Σ is known. Hence, as emphasized above, Bayes linear prediction of β_N requires only the first and second moments of $\mathbf{y}$ and β.

Let

(20)
$$\hat{\beta}_{LN} = \mathbf{a} + \mathbf{M}'\mathbf{y}_s,$$

where the px1 vector $\mathbf{a}$ and the nxp matrix $\mathbf{M}$ are known, be any linear predictor of β_N. The model chacterized by (21) and (22) is denoted by model ψ_B in the sequel. In the following lemma, we provide an expression for the Bayesian GMSE of any linear predictor $\hat{\beta}_{LN}$ given in (20). The proof is based on standard calculations and so it is omitted.

Lemma 5.1. *Under model ψ_B, for any px1 vector* $\mathbf{a}$ *and nxp matrix* $\mathbf{M}$, *we have*

$$E_{\psi_B}[\lambda'(\hat{\beta}_N - \beta_N)]^2 = Var_{\psi_B}[\lambda'\mathbf{a} + \mathbf{F}' - \mathbf{P}'\beta]$$
$$+\lambda'\mathbf{Q}_r\mathbf{V}_r\mathbf{Q}_r'\lambda - \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{V}_{sr}\mathbf{Q}_r'\lambda + [\lambda'\mathbf{a} + \lambda'(\mathbf{M}'\mathbf{X}_s - \mathbf{I}_p)\mu]^2,$$

where

$$\mathbf{F} = \lambda'\mathbf{M}' - \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}$$

and

$$\mathbf{P}' = \lambda' - \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{X}_s.$$

Note that Lemma 5.1 is indeed a generalization of Lemma 4.2. Conditions for unbisedness under model ψ_B are stated in the next lemma. The proof is also ommited.

Lemma 5.2. *Under model ψ_B, $\hat{\beta}_{LN}$ is unbiased if and only if*

$$E_{\psi_B}[\lambda'\mathbf{a} + \mathbf{F}'\mathbf{y}_s - \mathbf{P}'\beta] = \lambda'[\mathbf{a} + (\mathbf{M}'\mathbf{X}_s - \mathbf{I}_p)\mu] = 0.$$

From Lemmas 5.1 and 5.2, it follows that the problem of finding the predictor of β_N which is unbiased and minimizes the GMSE with respect to the model ψ_B is equivalent to find $\mathbf{a} + \mathbf{F}'\mathbf{y}_s$ such that

$$E_{\psi_B}[\mathbf{a} + \mathbf{F}'\mathbf{y}_s - \mathbf{P}'\beta] = 0 \tag{21}$$

and

$$Var_{\psi_B}[\mathbf{a} + \mathbf{F}'\mathbf{y}_s - \mathbf{P}'\beta] \tag{22}$$

is a minimum. The problem of minimizing (22) subject to the condition (21) can be solved directly by using some results in Rao (1971, pg 234). By making use of those results, it follows that the optimum $\mathbf{a}$ and $\mathbf{F}$ are given by

$$\mathbf{a}'_* = \mu'\mathbf{P} - \mu'\mathbf{X}'_s\mathbf{F}_*,$$

and

$$\mathbf{F}_* = \mathbf{V}_s^{-1}\mathbf{X}_s(\Sigma_s^{-1} + \mathbf{X}'_s\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{P}.$$

Further,

$$Var_{\psi_B}[\mathbf{a}_* + \mathbf{F}_*\mathbf{y}_s - \mathbf{P}'\beta] = \mathbf{P}'(\Sigma^{-1} + \mathbf{X}'_s\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{P}. \tag{23}$$

Thus, after some algebraic manipulations, it follows that the Bayes linear predictor of β_N with minimum GMSE with respect to model ψ_B is given by

$$\hat{\beta}_{BBLN} = \mathbf{Q}_s\mathbf{y}_s + \mathbf{Q}_r[\mathbf{X}'_r\hat{\hat{\beta}} + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\hat{\hat{\beta}})], \tag{24}$$

where

$$\hat{\hat{\beta}} = \mathbf{C}\hat{\beta}_s + (\mathbf{I} - \mathbf{C})\mu, \tag{25}$$

and

$$\mathbf{C} = [\Sigma^{-1} + \mathbf{X}'_s\mathbf{V}_s^{-1}\mathbf{X}_s]^{-1}(\mathbf{X}'_s\mathbf{V}_s^{-1}\mathbf{X}_s).$$

Furthermore, by using (23) it follows that the GMSE of predictor $\hat{\beta}_{BBLN}$ with respect to model ψ_B is given by

$$E_{\psi_B}[\lambda'(\hat{\beta}_{BBLN} - \beta_N)]^2 = \mathbf{P}'(\mathbf{X}'_s\mathbf{V}_s^{-1}\mathbf{X}_s + \Sigma^{-1})^{-1}\mathbf{P} + \lambda'\mathbf{Q}'_r\mathbf{V}_r\mathbf{Q}'_r\lambda - \lambda'\mathbf{Q}_r\mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{V}_{sr}\mathbf{Q}'_r\lambda.$$

In the case where the prior mean vector μ is unknown, it can be shown, by using also some results in Rao (1971), pg 234, that the Bayes linear predictor of β_N with minimum GMSE with respect to model ψ_B is given by (8), with prediction GMSE (with respect to model ψ_B) given by (9). Note also that predictor (24) coincides with the optimal predictor $\hat{\beta}_{BLN}$ when Σ is so large thatis not absurd to replace Σ^{-1} by the null matrix. This situation is typically known as the noninformative case. The next simple example shows that with respect to an informative prior distribution predictor (24) and estimator (25) do not coincide.

Example 5.1. Suppose that model (1) is such that $\mathbf{X} = \mathbf{1}_N$ and $\mathbf{V} = \sigma^2 \mathbf{I}_N$. This model is typically known as the simple location model. Consider the prior information $E[\beta] = \beta_o$ and $Var[\beta] = \tau^2$. After some algebraic manipulations, it follows from (25) that

$$\hat{\hat{\beta}} = \frac{\frac{n\bar{y}_s}{\sigma^2} + \frac{\beta_o}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}.$$

Moreover, from (24), we have that

$$\hat{\beta}_{BBLN} = \frac{n}{N}\bar{y}_s + (1 - \frac{n}{N})\hat{\hat{\beta}},$$

which clearly does not coincides with $\hat{\hat{\beta}}$ given above.

References

Arnold, S.F. (1981). *The theory of linear models.* New York: Wiley.

Bolfarine, H. and Zacks, S. (1991). Bayes and minimax prediction in finite populations. *Journal of Statistical Planning and Inference*, (accepted for publication).

Fuller, W.A. (1975). Regression analisys for sample surveys. *Sankhya, Series C*, 37, 117-132.

Hidiroglou, M. (1974). Estimation of regression parameters for finite populations. PhD. thesis, Iowa State University, Ames, Iowa.

Konijin, H.S. (1962). Regression Analysis in Sample Surveys. *Journal of the American Statistical Association*, 68, 880-889.

La Motte, L.R. (1978). Bayes linear estimators. *Technometrics*, 20, 281-290.

Pereira, C.A.B., and Rodrigues, J. (1983). Robust Linear Prediction in Finite Populations. *International Statistical Review*, 51, 293-300.

Rao, C.R. (1971). *Linear Statistical Inference and its Applications*, New York: Wiley.

Rodrigues, J., Bolfarine, H. and Rogakto, A. (1985). A general theory of prediction in finite populations. *International Statististical Review*, 53, 239-254.

Shah, B.V., Holt, M.M. and Folsom, R.E. (1977). Inference about regression models from sample survey data. *Bulletin of the International Statistical Institute*, 47, 3, 43–57.

Smouse, E.P. (1984). A note on Bayesian least squares inference for finite population models. *Journal of the American Statistical Association*, 79, 390-392.

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

1990

9001 - JR., HODGES, J.L. RAMSEY, P.H. & WECHSLER, S.: Improved Significance Probabilities of the Wilcoxon Test, 30p.

9002 - PAULA, G.A.: Bias Correction for Exponential Family Nonlinear Models, 10p.

9003 - SCHONMANN, R.H. & TANAKA, N.I.: One Dimensional Caricature of Phase Transition, 16p.

9004 - ZACKS, S. & BOLFARINE, H.: Maximum Likelihood Prediction in Finite Populations, 15p.

9005 - ZACKS, S. & BOLFARINE, H.: Equivariant Prediction of the Population Variance Under Location-Scale Superpopulation Models, 8p.

9006 - BOLFARINE, H.: Finite Population Prediction Under Error in Variables Superpopulation Models, 14p.

9007 - BOLFARINE, H.: Ratio and Regression Estimators Under Error-In-Variables Superpopulation, 16p.

9008 - BOTTER, D.A. & MOTTA, J.M.: Experiments With Three-Treatment Three-Period Crossover Design: Analysis Through the General Linear Models, 15p.

9009 - SCHONMANN, R.H.: Finite Size Scaling Behavior of a Biased Majority Rule Cellular Automaton, 11p.

9010 - BUENO, V.C.: A Martingale Version of Multistate Monotone Systems, 5p.

9011 - CORDEIRO, G.M. & FERRARI, S.L.P.: A Modified Score Tests Statistic Having Chi-Squared Distribution to Order, 12p.

9012 - NEVES, E.J. & SCHONMANN, R.H.: Behavior of Droplets for a Class a Glauber Dynamics at very Low Temperature, 33pg.

9013 - FERRARI, P.A.: Schock Fluctuations in Asymmetric Simple Exclusion, 26p.

9014 - MAGALHAES, M.N. & TANAKA, N.I.: Reverse Process in M/M/1 Queues, 7p.

9015 - CORDEIRO, G.M. & PAULA, G.A.: Estimation Large Sample Parametric Tests and Diagnostic for Non-Exponential Family Nonlinear Models, 31p.

9016 - CORDEIRO, G.M., FERRARI, S.L.P. & PAULA, G.A.: Improved Score Tests for Generalized Linear Models, 25p.

9017 - ESTON, V.R., GALVES, A. JACOBI, C.M., & LANGENVIN, R.: Dominance Switch Between Two Interacting Species and Large Deviations, 8p.

9018 - IRONY, T.Z., PEREIRA, C.A.B. & BARLOW, R.E.: Bayesian Models for Quality Assurance, 22p.

9019 - BOLFARINE, H. & ZACKS, S., Bayes and Minimax Prediction in Finite Populations, 16p.

9020 - BOLFARINE, H.; RODRIGUES, J. & ZACKS, S.: Some Asymptotic Results in Finite Populations, 13p.

9021 - BOLFARINE, H., CORDANI, L.K.: Estimation of a Structural Linear Regression Model With a Known Reliability Ratio, 9pg.

9022 - PAULINO, C.D.M., PEREIRA, C.A.B.: Bayesian Analysis of Categorical Data Informatively Censored. São Paulo, 18pg

9023 - FERRARI, P.A., MARTINEZ, S. & PICCO, P.: Some Properties of Quasi Stationary Distributions in the Birth and Death Chains, 11pg.

9024 - BOLFARINE, H., SANDOVAL, M.C.: Prediction of the finite population distribution function under Gaussian super popultion models, 13 pg.

9025 - BUENO, V.C. & NORROS, I.: Component importance through compensator transforms, 7p

1991

9101 - RAVISHANKAR, K.: Convergence of a random motion to a point, 11p.

9102 - FERRARI, P.A. & KAVISHANKAR, R.: Shocks in asymmetric exclusion automata, 17p.

9103 - ANDRE, C.D.S. & PERES, C.A.: An Algorithm for the Msae Estimation of the Multistage Dose-Responde model, 14p.

9104 - FERRARI, P.A.; Microscopic shocks in one dimensional driven systems, 21p.

9105 - BOLFARINE, H., LEITE, J.G. & RODRIGUES, J.: On the Estimation of the Size of a Finite and Closed Population, 17p.

9106 - IGLESIAS, P., SANDOVAL, M.C. & PEREIRA, C.A.B.: Predictive Likelihood in Finite Populations, 17p.

9107 - CORDEIRO, G.M., PAULA, G.A. & BOTTER, D.A.: Improved Likelihood Ratio Tests for Dispersion Models, 32p.

9108 - WERKEMA, M.C.C. & TANAKA, N.I.: Alguns Aspectos sobre Filtragens em Filas M/M'J/N com "Feedback", 35p.

9109 - WECHSLER, S.; Bayesian Note on Game Theory, 3p.

9110 - WECHSLER, S.; Rationality in Rubin's Sense, 16p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
IME-USP
Caixa Postal 20.570
01498 - São Paulo, Brasil