Chapter 4
Prediction Theory

4.1 Introduction

In the classical design-based inference theory, the population is fixed and the probability distribution of interest is determined by the random mechanism employed to extract the sample. As there are many ways of extracting a sample, the sampling design plays a relevant role and the finding of good estimators depends on it. On the other hand, the prediction theory treats the values of the target variable in all the units of the population as the realization of a random vector. The probability distribution of the population target vector is introduced by a statistical model and the inference procedures are optimized with respect to that distribution.

The prediction theory for finite populations relies on the so-called superpopulation models in which values of the target variables on population elements are considered as realizations of random variables having joint distributions. The model selection, fit, and diagnostic is the first and important step when applying a statistical methodology based on the prediction theory. Contrary to what happens under the sampling design theory, under the prediction theory there is no true model in applications to real data. There will only be useful models that adequately describe the behavior and relationships between the target variables and the auxiliary variables. The emphasis of this approach is thus on the analysis of data rather than on the design of samples. For more information about the prediction theory for finite populations, see e.g. the books of Cassel et al. (1977) Bolfarine and Zacks (1992) or Valliant et al. (2000).

This chapter gives a description of the prediction theory for finite populations. Section 4.1 introduces the basic notation and gives an illustrative example. Section 4.2 deals with the problem of predicting linear population parameters under a general linear model. Section 4.3 proves the general prediction theorem under a superpopulation linear model. Section 4.4 derives the best linear unbiased predictors of population totals under some linear models. It also shows that some of the

D. Morales et al., *A Course on Small Area Estimation and Mixed Models*, Statistics for Social and Behavioral Sciences, https://doi.org/10.1007/978-3-030-63757-6_4

obtained predictors are widely used estimators in the statistical inference of finite populations. Finally, Sect. 4.5 gives the corresponding R codes.

4.2 The Predictive Approach

Let N be the known number of units in a finite population and let y_j be the numeric value of the target variable measured at the population unit j. Without loss of generality, we write the population in the form $U = \{1, 2, \ldots, N\}$. A sample is a subset of U, i.e. $s \subset U$.

The *general problem* is to select sample $s \subset U$ of size n and to use the numerical values $y_1, \ldots, y_n$, associated to the units of s, for estimating

$$h(y_1, \ldots, y_N),$$

where the functional form of h is known (for example a linear function).

The *predictive approach* treats the numerical values $y_1, \ldots, y_N$ (y-values) as realizations of random variables $Y_1, \ldots, Y_N$. After observing the sample, estimating $h(y_1, \ldots, y_N)$ is in fact predicting the value of a function of the non-observed variables Y_j. The predictive approach models the relationships between the variables through the joint probability distribution of $(Y_1, \ldots, Y_N)$. The predictions are done with respect to that distribution (the model distribution).

We use the term "prediction" in the sense of "guessing," with statistical techniques, the values of the non-observed random variables Y. We do not use this term in the sense of "guessing" future values that might occur (like in time series).

Let $r \subset U$ be the set of non-sampled population units, i.e. $U = s \cup r$. The y-values in s are known, but the ones in r are not. The prediction effort is addressed to the y-values in r, or to a function of them. By using together what is known, for individuals in s, and what is predicted, for individuals in r, we can get predictions for the population U. The following example follows the same steps as the one appearing in Section 1.2 of Valliant et al. (2000).

Example 4.1 For a given region and time period, Table 4.1 shows data from hospitals. We want to estimate $T = \sum_{j=1}^{N} y_j$. We have a sample s of size $n = 10$, excluding the case 11. Note that $\sum_{j \in s} y_j$ is known and that

$$T = \sum_{j \in s} y_j + y_{11}.$$

Therefore, estimating T is equivalent to predicting y_{11}.

Let us assume that a simple regression model M holds, i.e.

$$E_M[Y_j] = \beta x_j, \; j = 1, \ldots, N, \qquad \text{cov}_M(Y_j, Y_k) = \begin{cases} \sigma^2 x_j & \text{if } j = k, \\ 0 & \text{otherwise,} \end{cases}$$

Table 4.1 Data from hospitals

Hospital	N. of beds (x)	N. of discharges (y)
1	56	180
2	59	269
3	75	236
4	83	222
5	114	361
6	117	400
7	119	337
8	121	394
9	151	600
10	209	506
11	251	617
Total for $N = 11$	1355	4122
Totals for $n = 10$	1104	3505

where β is an unknown parameter that we have to estimate. The best linear unbiased estimator (BLUE) can be obtained by minimizing the weighted sum of squared errors

$$SWSE = \sum_{j \in s} \frac{1}{\sigma^2 x_j}(y_j - \beta x_j)^2.$$

By taking derivatives and equating to zero, we have

$$0 = \frac{\partial SWSE}{\partial \beta} = -\sum_{j \in s} \frac{2(y_j - \beta x_j)x_j}{\sigma^2 x_j} \iff \hat{\beta} = \frac{\sum_{j \in s} y_j}{\sum_{j \in s} x_j} = \frac{3505}{1104} = 3.175.$$

In a conventional regression analysis we should calculate a confidence interval for β or we should test the hypothesis $H_0 : \beta = 0$. In this case $\hat{\beta}$ is only an intermediate step for arriving to the final target: the estimation of T.

In hospital 11, we have $x_{11} = 251$, the corresponding prediction is $\hat{y}_{11} = \hat{\beta} x_{11} = 3.175 \cdot 251 = 796.88$ and

$$\hat{T} = \sum_{j \in s} y_j + \hat{y}_{11} = 3505 + 796.88 = 4301.88.$$

In this case the relative error is

$$\frac{|\hat{T} - T|}{T} 100\% = \frac{|4301.88 - 4122|}{4122} 100\% \approx 4.36\%,$$

which is rather moderate.

Example 4.1 can be extended to the case

$$T = \sum_{j\in s} y_j + \sum_{j\in r} y_j.$$

In this context, the estimator of T is

$$\hat{T} = \sum_{j\in s} y_j + \sum_{j\in r} \hat{\beta} x_j = \sum_{j\in s} y_j + \frac{\sum_{j\in s} y_j}{\sum_{j\in s} x_j} \sum_{j\in r} x_j = \left(\frac{\sum_{j\in s} y_j}{\sum_{j\in s} x_j}\right) \sum_{j\in s} x_j + \left(\frac{\sum_{j\in s} y_j}{\sum_{j\in s} x_j}\right) \sum_{j\in r} x_j$$

$$= \left(\frac{\sum_{j\in s} y_j}{\sum_{j\in s} x_j}\right) \sum_{j\in U} x_j = N\overline{y}_s \frac{\overline{x}}{\overline{x}_s} \triangleq \hat{T}_R \text{ (ratio estimator)},$$

where $\overline{y}_s = \frac{1}{n}\sum_{j\in s} y_j$, $\overline{x}_s = \frac{1}{n}\sum_{j\in s} x_j$ and $\overline{x} = \frac{1}{N}\sum_{j=1}^{N} x_j$. Section 4.3 extends these ideas to the prediction of linear parameters.

4.3 Prediction Theory Under the Linear Model

Let us consider a finite population $U = \{1, \ldots, N\}$. Let $\boldsymbol{y} = (y_1, \ldots, y_N)'$ be the vector containing the values of a variable Y in all the population units. The *target* is to estimate a linear combination of $y_1, \ldots, y_N$, $\boldsymbol{\gamma}'\boldsymbol{y}$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)'$ is a vector containing N known constants. For example,

- If $\gamma_j = 1$, $j = 1 \ldots, N$, then $\boldsymbol{\gamma}'\boldsymbol{y} = \sum_{j=1}^{N} y_j$ is the population total,
- If $\gamma_j = \frac{1}{N}$, $j = 1 \ldots, N$, then $\boldsymbol{\gamma}'\boldsymbol{y} = \frac{1}{N}\sum_{j=1}^{N} y_j$ is the population mean.

Let us consider a sample $s \subset U$ of $n \leq N$ units. Let $r = U - s$ be the set of non-sampled units. Without loss of generality, we renumber the population units and we write $\boldsymbol{y} = (\boldsymbol{y}_s', \boldsymbol{y}_r')'$, where

- $\boldsymbol{y}_s$ is the vector of size n containing the values of Y in the observed units,
- $\boldsymbol{y}_r$ is the vector of size $N - n$ containing the values of Y in the non-observed units.

Similarly, we write $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_s', \boldsymbol{\gamma}_r')'$ and $\boldsymbol{\gamma}'\boldsymbol{y} = \boldsymbol{\gamma}_s'\boldsymbol{y}_s + \boldsymbol{\gamma}_r'\boldsymbol{y}_r$. Note that the problem of estimating $\boldsymbol{\gamma}'\boldsymbol{y}$ is equivalent to the problem of predicting the value of the non-observed random variable $\boldsymbol{\gamma}_r'\boldsymbol{y}_r$.

Definition 4.1 A *linear estimator* of $\theta = \boldsymbol{\gamma}'\boldsymbol{y}$ is $\hat{\theta} = \boldsymbol{g}_s'\boldsymbol{y}_s$, where $\boldsymbol{g}_s = (g_1, \ldots, g_n)'$ is a vector of n coefficients.

Definition 4.2 The *estimation error* of the estimator $\hat{\theta} = \boldsymbol{g}_s'\boldsymbol{y}_s$ is $\hat{\theta} - \theta = \boldsymbol{g}_s'\boldsymbol{y}_s - \boldsymbol{\gamma}'\boldsymbol{y}$.

We can write the estimation error as a function of the observed and non-observed measurements, i.e.

$$\hat{\theta} - \theta = \boldsymbol{g}_s' \boldsymbol{y}_s - \boldsymbol{\gamma}' \boldsymbol{y} = (\boldsymbol{g}_s' - \boldsymbol{\gamma}_s') \boldsymbol{y}_s - \boldsymbol{\gamma}_r' \boldsymbol{y}_r = \boldsymbol{a}' \boldsymbol{y}_s - \boldsymbol{\gamma}_r' \boldsymbol{y}_r, \quad \text{with } \boldsymbol{a} = \boldsymbol{g}_s - \boldsymbol{\gamma}_s.$$

Note that

- The first component, $\boldsymbol{a}' \boldsymbol{y}_s$, depends only on the sampled units and its value can be calculated after observing the sample s.
- The second component depends on the non-sampled units and its value should be predicted.
- An "ideal" best estimator has the property $0 = \hat{\theta} - \theta = \boldsymbol{a}' \boldsymbol{y}_s - \boldsymbol{\gamma}_r' \boldsymbol{y}_r$. Therefore, using $\boldsymbol{g}_s' \boldsymbol{y}_s$ for estimating $\boldsymbol{\gamma}' \boldsymbol{y}$ is equivalent to using $\boldsymbol{a}' \boldsymbol{y}_s$ for predicting $\boldsymbol{\gamma}_r' \boldsymbol{y}_r$. This is to say, finding a good "$\boldsymbol{g}_s$" is equivalent to finding a good "$\boldsymbol{a}$."

In this section we study the prediction problem under the general linear model M:

$$E_M[\boldsymbol{y}] = \boldsymbol{X}\boldsymbol{\beta}, \qquad \text{var}_M(\boldsymbol{y}) = \boldsymbol{V}, \tag{4.1}$$

where $\boldsymbol{X}_{N \times p}$ is the matrix of auxiliary variables, $\boldsymbol{\beta}_{p \times 1}$ is the vector of unknown regression parameters and $\boldsymbol{V}_{N \times N}$ is a known positive definite covariance matrix. We assume that the values of the auxiliary variables are known in all the population units, i.e. $\boldsymbol{X}_{N \times p}$ is known.

We sort the population units and we express the matrices X and V in the following block form:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_s \\ \boldsymbol{X}_r \end{pmatrix}, \qquad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_s & \boldsymbol{V}_{sr} \\ \boldsymbol{V}_{rs} & \boldsymbol{V}_r \end{pmatrix},$$

where $\boldsymbol{X}_s$ is $n \times p$, $\boldsymbol{X}_r$ is $(N-n) \times p$, $\boldsymbol{V}_s$ is $n \times n$, $\boldsymbol{V}_r$ is $(N-n) \times (N-n)$, $\boldsymbol{V}_{sr}$ is $n \times (N-n)$, and $\boldsymbol{V}_{rs} = \boldsymbol{V}_{sr}'$. We further assume that $\boldsymbol{V}_s$ is positive definite and $\boldsymbol{X}_s$ has a full rank.

Definition 4.3 The estimator $\hat{\theta}$ is *unbiased* for θ under the model M if and only if $E_M[\hat{\theta} - \theta] = 0$. We can also say *predictively unbiased* or unbiased with respect to the model distribution.

Note that $E_M[\hat{\theta}] = \theta$ is not correct because θ is random.

Definition 4.4 The error variance (prediction variance) of $\hat{\theta}$ under M is $\text{var}_M(\hat{\theta} - \theta)$.

If $\hat{\theta}$ is predictively unbiased, then its error variance is equal to its mean squared error, i.e. $\text{var}_M(\hat{\theta} - \theta) = E_M[(\hat{\theta} - \theta)^2]$.

Example 4.2 (Ratio Estimator) We show that the ratio estimator of the total $T = \sum_{j=1}^{N} y_j$, $\hat{T}_R = N\overline{y}_s \frac{\overline{x}}{\overline{x}_s}$, is the best linear unbiased predictor (BLUP) if we work under the model

$$E_M[Y_j] = \beta x_j, \; j = 1, \ldots, N, \qquad \mathrm{cov}_M(Y_j, Y_k) = \begin{cases} \sigma^2 x_j & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Here, the best estimator means the estimator which minimizes the error variance. For a sample s, we have $T = \sum_{j\in s} y_j + \sum_{j\in r} y_j$. If we should know the value of the parameter β (which is unknown), then we could estimate T with $T^* = \sum_{j\in s} y_j + \beta \sum_{j\in r} x_j$, because $E_M[T^* - T] = 0$. On the other hand, every estimator $\hat{T}$ of T can be written as

$$\hat{T} = \sum_{j\in s} y_j + \left[\frac{\hat{T} - \sum_{j\in s} y_j}{\sum_{j\in r} x_j}\right] \sum_{j\in r} x_j,$$

so that $\hat{T}$ has the same form as T^* and $(\hat{T} - \sum_{j\in s} y_j)/\sum_{j\in r} x_j$ estimates β. We can write

$$\hat{T} = \sum_{j\in s} y_j + \hat{\beta} \sum_{j\in r} x_j \quad \text{and} \quad \hat{T} - T = \hat{\beta} \sum_{j\in r} x_j - \sum_{j\in r} y_j.$$

Therefore, $\hat{T}$ is predictively unbiased if

$$E_M\Big[\hat{\beta} \sum_{j\in r} x_j - \sum_{j\in r} y_j\Big] = \big(E_M[\hat{\beta}] - \beta\big) \sum_{j\in r} x_j = 0.$$

This is to say, $\hat{T}$ is predictively unbiased for T if and only if $\hat{\beta}$ is predictively unbiased for β.

The error variance of $\hat{T} = \sum_{j\in s} y_j + \hat{\beta} \sum_{j\in r} x_j$ is

$$\mathrm{var}_M\big(\hat{T} - T\big) = \mathrm{var}_M\Big(\hat{\beta} \sum_{j\in r} x_j - \sum_{j\in r} y_j\Big) = \Big(\sum_{j\in r} x_j\Big)^2 \mathrm{var}_M\big(\hat{\beta}\big) + \mathrm{var}_M\Big(\sum_{j\in r} y_j\Big).$$

For minimizing the error variance of $\hat{T}$ we have to minimize the variance of $\hat{\beta}$. Assume that we are restricted to linear unbiased estimators of β, i.e.

$$\hat{\beta} = \sum_{j\in s} a_j y_j, \quad E_M[\hat{\beta}] = \beta \sum_{j\in s} a_j x_j = \beta, \quad \sum_{j\in s} a_j x_j = 1.$$

The variance of $\hat{\beta}$ is

$$\mathrm{var}_M(\hat{\beta}) = \sigma^2 \sum_{j\in s} a_j^2 x_j.$$

We find the best linear unbiased estimator (BLUE) of β by applying the Lagrange multiplier method. The Lagrangian function is

$$L = \sigma^2 \sum_{j\in s} a_j^2 x_j + \lambda \left(\sum_{j\in s} a_j x_j - 1 \right).$$

By taking derivatives, we have

$$0 = \frac{\partial L}{\partial a_j} = 2\sigma^2 a_j x_j + \lambda x_j, \quad j \in s, \tag{4.2}$$

$$0 = \frac{\partial L}{\partial \lambda} = \sum_{j\in s} a_j x_j - 1, \tag{4.3}$$

and

$$0 = \sum_{j\in s} \frac{\partial L}{\partial a_j} = 2\sigma^2 \sum_{j\in s} a_j x_j + \lambda n \overline{x}_s = 2\sigma^2 + \lambda n \overline{x}_s,$$

which implies

$$\lambda = -\frac{2\sigma^2}{n\overline{x}_s}.$$

By substituting in (4.2), we get

$$a_j = -\lambda \frac{1}{2\sigma^2} = \frac{2\sigma^2}{n\overline{x}_s} \frac{1}{2\sigma^2} = \frac{1}{n\overline{x}_s}.$$

The BLUE of β is

$$\hat{\beta} = \sum_{j\in s} \frac{1}{n\overline{x}_s} y_j = \frac{\overline{y}_s}{\overline{x}_s}.$$

The BLUP of the total T, obtained from the BLUE of β, is

$$\begin{aligned} \hat{T} &= \sum_{j\in s} y_j + \hat{\beta} \sum_{j\in r} x_j = \sum_{j\in s} y_j + \frac{\overline{y}_s}{\overline{x}_s} \sum_{j\in r} x_j \\ &= \frac{n\overline{y}_s}{\overline{x}_s} \overline{x}_s + \frac{\overline{y}_s}{\overline{x}_s} \sum_{j\in r} x_j = N \frac{\overline{x}}{\overline{x}_s} \overline{y}_s = \hat{T}_R \text{ (ratio estimator).} \end{aligned}$$

The estimation error is

$$\hat{T}_R - T = \frac{N\overline{x}}{n\overline{x}_s}\sum_{j\in s} y_j - \sum_{j\in U} y_j = \left(\frac{N\overline{x}}{n\overline{x}_s} - 1\right)\sum_{j\in s} y_j - \sum_{j\in r} y_j$$
$$= \frac{(N-n)\overline{x}_r}{n\overline{x}_s}\sum_{j\in s} y_j - \sum_{j\in r} y_j,$$

where $\overline{x}_r = \frac{1}{N-n}\sum_{j\in r} x_j$. The error variance is

$$\mathrm{var}_M\left(\hat{T}_R - T\right) = \left(\frac{(N-n)\overline{x}_r}{n\overline{x}_s}\right)^2 n\overline{x}_s\sigma^2 + (N-n)\overline{x}_r\sigma^2$$
$$= \frac{(N-n)^2\overline{x}_r^2}{n\overline{x}_s}\sigma^2 + (N-n)\overline{x}_r\sigma^2 = (N-n)\overline{x}_r\sigma^2\left(\frac{(N-n)\overline{x}_r}{n\overline{x}_s} + 1\right)$$
$$= (N-n)\overline{x}_r\sigma^2\frac{(N-n)\overline{x}_r + n\overline{x}_s}{n\overline{x}_s} = \frac{(N-n)N}{n}\frac{\overline{x}_r\overline{x}}{\overline{x}_s}\sigma^2$$
$$= \frac{N^2}{n}(1-f)\frac{\overline{x}_r\overline{x}}{\overline{x}_s}\sigma^2, \quad \text{where } f = \frac{n}{N}.$$

4.4 The General Prediction Theorem

The following theorem gives the best linear unbiased predictor of a linear parameter under a superpopulation linear model (4.1). It also gives the corresponding prediction variance. For more details, see Chapter 2 of Valliant et al. (2000).

Theorem 4.1 *Among linear predictively unbiased estimators* $\hat{\theta} = \boldsymbol{g}_s'\boldsymbol{y}_s$ *of* $\theta = \boldsymbol{\gamma}'\boldsymbol{y}$, *the error variance is minimized by*

$$\hat{\theta}_{opt} = \boldsymbol{\gamma}_s'\boldsymbol{y}_s + \boldsymbol{\gamma}_r'\left[\boldsymbol{X}_r\hat{\boldsymbol{\beta}} + \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}})\right], \tag{4.4}$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s$. *The error variance of* $\hat{\theta}_{opt}$ *is*

$$\mathrm{var}_M(\hat{\theta}_{opt} - \theta) = \boldsymbol{\gamma}_r'(\boldsymbol{V}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr})\boldsymbol{\gamma}_r$$
$$+ \boldsymbol{\gamma}_r'(\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)(\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)^{-1}(\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)'\boldsymbol{\gamma}_r.$$

Proof The error variance is

$$E_M\left[(\boldsymbol{g}_s'\boldsymbol{y}_s - \boldsymbol{\gamma}'\boldsymbol{y})^2\right] = E_M\left[(\boldsymbol{a}'\boldsymbol{y}_s - \boldsymbol{\gamma}_r'\boldsymbol{y}_r)^2\right]$$
$$= \mathrm{var}_M\left(\boldsymbol{a}'\boldsymbol{y}_s - \boldsymbol{\gamma}_r'\boldsymbol{y}_r\right) + \left(E_M\left[\boldsymbol{a}'\boldsymbol{y}_s - \boldsymbol{\gamma}_r'\boldsymbol{y}_r\right]\right)^2$$
$$= \boldsymbol{a}'\boldsymbol{V}_s\boldsymbol{a} - 2\boldsymbol{a}'\boldsymbol{V}_{sr}\boldsymbol{\gamma}_r + \boldsymbol{\gamma}_r'\boldsymbol{V}_r\boldsymbol{\gamma}_r + \left[(\boldsymbol{a}'\boldsymbol{X}_s - \boldsymbol{\gamma}_r'\boldsymbol{X}_r)\boldsymbol{\beta}\right]^2,$$

where $\boldsymbol{a} = \boldsymbol{g}_s - \boldsymbol{\gamma}_s$. Since we are assuming that $\hat{\theta} = \boldsymbol{g}_s'\boldsymbol{y}_s$ is unbiased, then

$$E_M\left[\boldsymbol{g}_s'\boldsymbol{y}_s - \boldsymbol{\gamma}'\boldsymbol{y}\right] = E_M\left[\boldsymbol{a}'\boldsymbol{y}_s - \boldsymbol{\gamma}_r'\boldsymbol{y}_r\right] = (\boldsymbol{a}'\boldsymbol{X}_s - \boldsymbol{\gamma}_r'\boldsymbol{X}_r)\boldsymbol{\beta} = 0,$$

i.e. the last term in the previous equation vanishes. The Lagrangian function for minimizing the error variance with respect to $\boldsymbol{a}$ is

$$L = L(\boldsymbol{a}, \boldsymbol{\lambda}) = \boldsymbol{a}'\boldsymbol{V}_s\boldsymbol{a} - 2\boldsymbol{a}'\boldsymbol{V}_{sr}\boldsymbol{\gamma}_r + 2(\boldsymbol{a}'\boldsymbol{X}_s - \boldsymbol{\gamma}_r'\boldsymbol{X}_r)\boldsymbol{\lambda}.$$

By taking derivatives with respect to $\boldsymbol{\lambda}$ and equating to zero, we get

$$0 = \frac{\partial L}{\partial \boldsymbol{\lambda}} = 2\boldsymbol{a}'\boldsymbol{X}_s - 2\boldsymbol{\gamma}_r'\boldsymbol{X}_r. \tag{4.5}$$

By taking derivatives with respect to $\boldsymbol{a}$, we get

$$0 = \frac{\partial L}{\partial \boldsymbol{a}} = 2\boldsymbol{V}_s\boldsymbol{a} - 2\boldsymbol{V}_{sr}\boldsymbol{\gamma}_r + 2\boldsymbol{X}_s\boldsymbol{\lambda} \tag{4.6}$$

and

$$\boldsymbol{a} = \boldsymbol{V}_s^{-1}(\boldsymbol{V}_{sr}\boldsymbol{\gamma}_r - \boldsymbol{X}_s\boldsymbol{\lambda}). \tag{4.7}$$

On the other hand, from (4.6) we have

$$\boldsymbol{X}_s\boldsymbol{\lambda} = \boldsymbol{V}_{sr}\boldsymbol{\gamma}_r - \boldsymbol{V}_s\boldsymbol{a}. \tag{4.8}$$

Pre-multiplying (4.8) by $\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}$, and taking into account that $\boldsymbol{X}_s'\boldsymbol{a} = \boldsymbol{X}_r'\boldsymbol{\gamma}_r$, we get

$$\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{\lambda} = \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr}\boldsymbol{\gamma}_r - \boldsymbol{X}_s'\boldsymbol{a} = \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr}\boldsymbol{\gamma}_r - \boldsymbol{X}_r'\boldsymbol{\gamma}_r = (\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr} - \boldsymbol{X}_r')\boldsymbol{\gamma}_r$$

and

$$\boldsymbol{\lambda} = \boldsymbol{A}_s^{-1}(\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr} - \boldsymbol{X}_r')\boldsymbol{\gamma}_r, \qquad \text{where} \quad \boldsymbol{A}_s = \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s.$$

By substituting $\boldsymbol{\lambda}$ in (4.7) we obtain the optimal value of $\boldsymbol{a}$, i.e.

$$\boldsymbol{a}_{opt} = \boldsymbol{V}_s^{-1}\left[\boldsymbol{V}_{sr} + \boldsymbol{X}_s\boldsymbol{A}_s^{-1}(\boldsymbol{X}_r' - \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr})\right]\boldsymbol{\gamma}_r.$$

The best linear unbiased estimator of $\boldsymbol{\gamma}_r'\boldsymbol{y}_r$ is

$$\begin{aligned}\boldsymbol{a}_{opt}'\boldsymbol{y}_s &= \boldsymbol{\gamma}_r'\left[\boldsymbol{V}_{sr}' + (\boldsymbol{X}_r - \boldsymbol{V}_{sr}'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\right]\boldsymbol{V}_s^{-1}\boldsymbol{y}_s\\ &= \boldsymbol{\gamma}_r'\left[\boldsymbol{V}_{sr}'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s + \boldsymbol{X}_r\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s - \boldsymbol{V}_{sr}'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s\right]\\ &= \boldsymbol{\gamma}_r'\left[\boldsymbol{X}_r\hat{\boldsymbol{\beta}} + \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}})\right],\end{aligned}$$

where

$$\hat{\boldsymbol{\beta}} = \boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s = (\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s.$$

Therefore, we obtain

$$\hat{\theta}_{opt} = \boldsymbol{\gamma}_s'\boldsymbol{y}_s + \boldsymbol{a}_{opt}'\boldsymbol{y}_s = \boldsymbol{\gamma}_s'\boldsymbol{y}_s + \boldsymbol{\gamma}_r'\left[\boldsymbol{X}_r\hat{\boldsymbol{\beta}} + \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}})\right].$$

Finally, the error variance is

$$\begin{aligned}V_M &= \mathrm{var}_M(\hat{\theta}_{opt} - \theta) = \boldsymbol{a}_{opt}'\boldsymbol{V}_s\boldsymbol{a}_{opt} - 2\boldsymbol{a}_{opt}'\boldsymbol{V}_{sr}\boldsymbol{\gamma}_r + \boldsymbol{\gamma}_r'\boldsymbol{V}_r\boldsymbol{\gamma}_r\\ &= \boldsymbol{\gamma}_r'\left[\boldsymbol{V}_{rs} + (\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\right]\boldsymbol{V}_s^{-1}\left[\boldsymbol{V}_{sr} + \boldsymbol{X}_s\boldsymbol{A}_s^{-1}(\boldsymbol{X}_r' - \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr})\right]\boldsymbol{\gamma}_r\\ &- 2\boldsymbol{\gamma}_r'\left[\boldsymbol{V}_{rs} + (\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\right]\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr}\boldsymbol{\gamma}_r + \boldsymbol{\gamma}_r'\boldsymbol{V}_r\boldsymbol{\gamma}_r\\ &= \boldsymbol{\gamma}_r'\left(\boldsymbol{V}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr}\right)\boldsymbol{\gamma}_r + \boldsymbol{\gamma}_r'\left[\boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{A}_s^{-1}(\boldsymbol{X}_r' - \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr})\right]\boldsymbol{\gamma}_r\\ &+ \boldsymbol{\gamma}_r'\left[(\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{A}_s^{-1}(\boldsymbol{X}_r' - \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr})\right]\boldsymbol{\gamma}_r\\ &- \boldsymbol{\gamma}_r'\left[(\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr}\right]\boldsymbol{\gamma}_r\\ &= \boldsymbol{\gamma}_r'\left(\boldsymbol{V}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{V}_{sr}\right)\boldsymbol{\gamma}_r + \boldsymbol{\gamma}_r'(\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)\boldsymbol{A}_s^{-1}(\boldsymbol{X}_r - \boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)'\boldsymbol{\gamma}_r.\end{aligned}$$

□

Corollary 4.1 *The estimator* $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s$ *minimizes the weighted sum of squared residuals* $SSE = (\boldsymbol{y}_s - \boldsymbol{X}_s\boldsymbol{\beta})'\boldsymbol{V}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s\boldsymbol{\beta})$.

Proof It holds that

$$SSE = \boldsymbol{y}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s + \boldsymbol{\beta}'\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s.$$

Therefore, we have

$$0 = \frac{\partial SSE}{\partial \boldsymbol{\beta}} = 2\boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{X}_s \boldsymbol{\beta} - 2\boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{y}_s \quad \text{and} \quad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{X}_s)^{-1} \boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{y}_s. \tag{4.9}$$

Since

$$\frac{\partial^2 SSE}{\partial \boldsymbol{\beta}^2} = 2\boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{X}_s$$

and we assume that $\boldsymbol{V}_s$ is symmetric and positive definite matrix, it follows that $\boldsymbol{V}_s^{-1}$ as well as $\boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{X}_s$ are positive definite matrices and thus $\hat{\boldsymbol{\beta}}$ is the point of minima of the function SSE. □

The equations $\boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{X}_s \hat{\boldsymbol{\beta}} = \boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{y}_s$, appearing in (4.9), are called *normal equations*. They are p equations with p unknowns $\beta_1, \ldots, \beta_p$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$. Further, the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is called *least squares estimator.*

Corollary 4.2 *If* $\boldsymbol{V}_{rs} = 0$, *then the BLUP and the error variance are*

$$\hat{\theta}_{opt} = \boldsymbol{\gamma}_s' \boldsymbol{y}_s + \boldsymbol{\gamma}_r' \boldsymbol{X}_r \hat{\boldsymbol{\beta}}, \quad \text{var}_M(\hat{\theta}_{opt} - \theta) = \boldsymbol{\gamma}_r' \left(\boldsymbol{V}_r + \boldsymbol{X}_r \boldsymbol{A}_s^{-1} \boldsymbol{X}_r' \right) \boldsymbol{\gamma}_r.$$

Therefore, it holds that $\hat{\theta}_{opt} = \boldsymbol{\gamma}_s' \boldsymbol{y}_s + \boldsymbol{\gamma}_r' \hat{\boldsymbol{y}}_r$, *where* $\hat{\boldsymbol{y}}_r = \boldsymbol{X}_r \hat{\boldsymbol{\beta}}$.

Proposition 4.1 *Among the linear predictively unbiased estimators* $\hat{\theta}$ *of* θ, *the variance is minimized by* $\hat{\theta}^* = \boldsymbol{\gamma}' \boldsymbol{X} \hat{\boldsymbol{\beta}}$, *where* $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{X}_s)^{-1} \boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{y}_s$. *The estimator variance is*

$$\text{var}_M(\hat{\theta}^*) = \boldsymbol{\gamma}' \boldsymbol{X} \boldsymbol{A}_s^{-1} \boldsymbol{X}' \boldsymbol{\gamma}, \quad \textit{with} \quad \boldsymbol{A}_s = \boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{X}_s.$$

Proof Let $\hat{\theta} = \boldsymbol{g}_s' \boldsymbol{y}_s$ be a linear estimator of $\theta = \boldsymbol{\gamma}' \boldsymbol{y}$. The variance of $\hat{\theta}$ is $\text{var}_M(\hat{\theta}) = \boldsymbol{g}_s' \boldsymbol{V}_s \boldsymbol{g}_s$. As $\hat{\theta}$ is predictively unbiased, it holds that

$$0 = E_M\left[\hat{\theta} - \theta\right] = \boldsymbol{g}_s' \boldsymbol{X}_s \boldsymbol{\beta} - \boldsymbol{\gamma}' \boldsymbol{X} \boldsymbol{\beta} = (\boldsymbol{g}_s' \boldsymbol{X}_s - \boldsymbol{\gamma}' \boldsymbol{X}) \boldsymbol{\beta}.$$

The Lagrangian function is

$$L = \boldsymbol{g}_s' \boldsymbol{V}_s \boldsymbol{g}_s + 2(\boldsymbol{g}_s' \boldsymbol{X}_s - \boldsymbol{\gamma}' \boldsymbol{X}) \boldsymbol{\lambda}.$$

It holds that

$$0 = \frac{\partial L}{\partial \boldsymbol{g}_s} = 2\boldsymbol{V}_s \boldsymbol{g}_s + 2\boldsymbol{X}_s \boldsymbol{\lambda}. \tag{4.10}$$

From (4.10) we get $\boldsymbol{g}_s = -\boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{\lambda}$. Pre-multiplying (4.10) by $\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}$, and taking into account that $\boldsymbol{X}_s'\boldsymbol{g}_s = \boldsymbol{X}'\boldsymbol{\gamma}$, we obtain

$$0 = \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_s\boldsymbol{g}_s + \boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{\lambda} = \boldsymbol{X}'\boldsymbol{\gamma} + \boldsymbol{A}_s\boldsymbol{\lambda}.$$

Therefore, we have

$$\boldsymbol{\lambda} = -\boldsymbol{A}_s^{-1}\boldsymbol{X}'\boldsymbol{\gamma} \qquad \text{and} \qquad \boldsymbol{g}_s^* = \boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{A}_s^{-1}\boldsymbol{X}'\boldsymbol{\gamma}.$$

By substituting the expressions of $\hat{\theta}$ and $\mathrm{var}_M(\hat{\theta})$, we get

$$\hat{\theta}^* = \boldsymbol{g}_s^{*\prime}\boldsymbol{y}_s = \boldsymbol{\gamma}'\boldsymbol{X}\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s = \boldsymbol{\gamma}'\boldsymbol{X}\hat{\boldsymbol{\beta}},$$

$$\mathrm{var}_M(\hat{\theta}^*) = \boldsymbol{g}_s^{*\prime}\boldsymbol{V}_s\boldsymbol{g}_s^* = \boldsymbol{\gamma}'\boldsymbol{X}\boldsymbol{A}_s^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{V}_s\boldsymbol{V}_s^{-1}\boldsymbol{X}_s\boldsymbol{A}_s^{-1}\boldsymbol{X}'\boldsymbol{\gamma} = \boldsymbol{\gamma}'\boldsymbol{X}\boldsymbol{A}_s^{-1}\boldsymbol{X}'\boldsymbol{\gamma},$$

which completes the proof. □

The following remarks give some comments of interest about the best linear unbiased predictors.

Remark 4.1

1. The estimator $\hat{\theta}_{opt}$ can be expressed in the form

   $$\hat{\theta}_{opt} = \boldsymbol{\gamma}_s'\boldsymbol{y}_s + \boldsymbol{\gamma}_r'\boldsymbol{X}_r\hat{\boldsymbol{\beta}} + \boldsymbol{\gamma}_r'\boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}}) = \boldsymbol{\gamma}_s'\boldsymbol{y}_s + \boldsymbol{\gamma}_r'\hat{\boldsymbol{y}}_r + \boldsymbol{\gamma}_r'\boldsymbol{V}_{rs}\boldsymbol{V}_s^{-1}\boldsymbol{e}_s.$$

   Therefore, $\hat{\theta}_{opt}$ uses the sampling units to reconstruct the "sample part" of the parameter $\theta = \boldsymbol{\gamma}'\boldsymbol{y}$. The term $\hat{\boldsymbol{y}}_r$ predicts the values of y in the non-sampled units and uses these predictions for reconstructing the non-sample part of the parameter. Finally, it uses the sample residuals $\boldsymbol{e}_s = \boldsymbol{y}_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}}$ for correcting the bias and obtaining the predictive unbiasedness.
2. The estimator $\hat{\theta}^* = \boldsymbol{\gamma}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{\gamma}'\hat{\boldsymbol{y}}$ does not use explicitly the observed sample values and it reconstructs the parameter by using only the predictions of the y-values.

Remark 4.2 Let us assume that $\boldsymbol{V}_{rs} = \boldsymbol{0}$. If the target parameter is $y_j = \boldsymbol{\eta}'\boldsymbol{y}$, with $\boldsymbol{\eta} = (0, \ldots, 0, 1^{(j)}, 0, \ldots, 0)$, then the BLUP is

$$\hat{y}_j = \begin{cases} y_j & \text{if } j \in s, \\ \boldsymbol{x}_j\hat{\boldsymbol{\beta}} = \tilde{y}_j & \text{if } j \in r, \end{cases}$$

where $\boldsymbol{x}_j$ is the row j of matrix $\boldsymbol{X}$. For any other parameter $\theta = \boldsymbol{\gamma}'\boldsymbol{y}$, we have

$$\hat{\theta}_{opt} = \boldsymbol{\gamma}_s'\boldsymbol{y}_s + \boldsymbol{\gamma}_r'\boldsymbol{X}_r\hat{\boldsymbol{\beta}} = \sum_{j\in s}\gamma_j y_j + \sum_{j\in r}\gamma_j\boldsymbol{x}_j\hat{\boldsymbol{\beta}} = \sum_{j\in s}\gamma_j y_j + \sum_{j\in r}\gamma_j\tilde{y}_j,$$

$$\hat{\theta}^* = \boldsymbol{\gamma}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \sum_{j\in U}\gamma_j\boldsymbol{x}_j\hat{\boldsymbol{\beta}} = \sum_{j\in U}\gamma_j\tilde{y}_j.$$

The estimator $\hat{\theta}_{opt}$ is called *predictive* and the estimator $\hat{\theta}^*$ is called *projective*.

4.5 BLUPs for Some Simple Models

In some cases the BLUPs are classical estimators appearing in survey sampling methods for finite populations. In the following examples, we assume that

1. The *prediction target* is $T = \sum_{j=1}^{N} y_j$; this is to say, $T = \theta = \boldsymbol{\gamma}'\boldsymbol{y}$ with $\boldsymbol{\gamma} = (1, \ldots, 1)'_{N\times 1}$.
2. The *notation* $e_j \sim (a, b)$ is used for indicating that e_j is a random error with $E[e_j] = a$ and $\text{var}(e_j) = b$. For vector and matrices, we use the notation

$$\mathbf{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1}, \quad \boldsymbol{I}_n = \begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1 \end{pmatrix}_{n\times n}, \quad \mathbf{1}_{n\times n} = \begin{pmatrix} 1 & \ldots & 1 \\ \vdots & \ddots & \vdots \\ \mathbf{1} & \ldots & 1 \end{pmatrix}_{n\times n}.$$

Example 4.3 (Expansive Estimator) Let us consider the model $y_j = \mu + e_j,\ j = 1, \ldots, N$, with uncorrelated random errors $e_j \sim (0, \sigma^2)$. In the framework of the general linear model, $E_M[\boldsymbol{y}] = \boldsymbol{X}\boldsymbol{\beta},\ \text{var}_M(\boldsymbol{y}) = \boldsymbol{V}$, we have that $\boldsymbol{\beta} = \mu,\ \boldsymbol{X} = \mathbf{1}_N,\ \boldsymbol{V} = \sigma^2\boldsymbol{I}_N$. Therefore, we have

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{y}_s = \left(\sigma^{-2}\mathbf{1}_n'\boldsymbol{I}_n\mathbf{1}_n\right)^{-1}\mathbf{1}_n'\sigma^{-2}\boldsymbol{I}_n\boldsymbol{y}_s = \frac{1}{n}\sum_{j=1}^{n} y_j = \overline{y}_s.$$

The BLUP is

$$\hat{T} = \boldsymbol{\gamma}_s'\boldsymbol{y}_s + \boldsymbol{\gamma}_r'\boldsymbol{X}_r\hat{\boldsymbol{\beta}} = n\overline{y}_s + (N-n)\overline{y}_s = N\overline{y}_s.$$

The error variance is

$$\begin{aligned}
\text{var}_M(\hat{T} - T) &= \boldsymbol{\gamma}_r'(\boldsymbol{V}_r + \boldsymbol{X}_r(\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_r')\boldsymbol{\gamma}_r \\
&= \mathbf{1}_{N-n}'\left(\sigma^2\boldsymbol{I}_{N-n} + \mathbf{1}_{N-n}\frac{\sigma^2}{n}\mathbf{1}_{N-n}'\right)\mathbf{1}_{N-n} \\
&= \mathbf{1}_{N-n}'\sigma^2\left(\boldsymbol{I}_{N-n} + \frac{1}{n}\mathbf{1}_{(N-n)\times(N-n)}\right)\mathbf{1}_{N-n} = \sigma^2\left[(N-n) + \frac{1}{n}(N-n)^2\right] \\
&= \frac{\sigma^2(N-n)N}{n} = \frac{N^2(1-f)\sigma^2}{n}, \quad \text{where} \quad f = \frac{n}{N}.
\end{aligned}$$

Example 4.4 (Linear Regression Estimator) Let us consider the model $y_j = a + bx_j + e_j,\ j = 1, \ldots, N$, with uncorrelated random errors $e_j \sim (0, \sigma^2)$. Under the general linear model, $E_M[\boldsymbol{y}] = \boldsymbol{X\beta}$, $\text{var}_M(\boldsymbol{y}) = \boldsymbol{V}$, we have

$$\boldsymbol{\beta} = \begin{pmatrix} a \\ b \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}, \quad \boldsymbol{V} = \sigma^2 \boldsymbol{I}_N.$$

Therefore, we have

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = (\boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{X}_s)^{-1} \boldsymbol{X}_s' \boldsymbol{V}_s^{-1} \boldsymbol{y}_s = \begin{pmatrix} n & \sum_{j=1}^n x_j \\ \sum_{j=1}^n x_j & \sum_{j=1}^n x_j^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^n y_j \\ \sum_{j=1}^n x_j y_j \end{pmatrix}$$

$$= \frac{\begin{pmatrix} \sum_{j=1}^n x_j^2 & -\sum_{j=1}^n x_j \\ -\sum_{j=1}^n x_j & n \end{pmatrix} \begin{pmatrix} \sum_{j=1}^n y_j \\ \sum_{j=1}^n x_j y_j \end{pmatrix}}{n\left(\sum_{j=1}^n x_j^2\right) - \left(\sum_{j=1}^n x_j\right)^2} = \begin{pmatrix} \frac{\left(\sum_{j=1}^n y_j\right)\left(\sum_{j=1}^n x_j^2\right) - \left(\sum_{j=1}^n x_j\right)\left(\sum_{j=1}^n x_j y_j\right)}{n\left(\sum_{j=1}^n x_j^2\right) - \left(\sum_{j=1}^n x_j\right)^2} \\ \frac{n\left(\sum_{j=1}^n x_j y_j\right) - \left(\sum_{j=1}^n x_j\right)\left(\sum_{j=1}^n y_j\right)}{n\left(\sum_{j=1}^n x_j^2\right) - \left(\sum_{j=1}^n x_j\right)^2} \end{pmatrix}.$$

The estimators of b and a are

$$\hat{b} = \frac{\left(\frac{1}{n}\sum_{j=1}^n x_j y_j\right) - \left(\frac{1}{n}\sum_{j=1}^n x_j\right)\left(\frac{1}{n}\sum_{j=1}^n y_j\right)}{\left(\frac{1}{n}\sum_{j=1}^n x_j^2\right) - \left(\frac{1}{n}\sum_{j=1}^n x_j\right)^2} = \frac{\frac{1}{n}\sum_{j=1}^n (x_j - \overline{x}_s)(y_j - \overline{y}_s)}{\frac{1}{n}\sum_{j=1}^n (x_j - \overline{x}_s)^2},$$

$$\hat{a} = \frac{\left[\overline{y}_s\left(\frac{1}{n}\sum_{j=1}^n x_j^2\right) - \overline{y}_s \overline{x}_s^2\right] - \left[\overline{x}_s\left(\frac{1}{n}\sum_{j=1}^n x_j y_j\right) - \overline{y}_s \overline{x}_s^2\right]}{\left(\frac{1}{n}\sum_{j=1}^n x_j^2\right) - \overline{x}_s^2}$$

$$= \overline{y}_s - \overline{x}_s \frac{\left(\frac{1}{n}\sum_{j=1}^n x_j y_j\right) - \left(\frac{1}{n}\sum_{j=1}^n x_j\right)\left(\frac{1}{n}\sum_{j=1}^n y_j\right)}{\left(\frac{1}{n}\sum_{j=1}^n x_j^2\right) - \overline{x}_s^2} = \overline{y}_s - \overline{x}_s \hat{b}.$$

The BLUP is

$$\hat{T} = \boldsymbol{\gamma}_s' \boldsymbol{y}_s + \boldsymbol{\gamma}_r' \boldsymbol{X}_r \hat{\boldsymbol{\beta}} = n\overline{y}_s + \mathbf{1}_{N-n}' \begin{pmatrix} 1 & x_{n+1} \\ \vdots & \vdots \\ \mathbf{1} & x_N \end{pmatrix} \begin{pmatrix} \overline{y}_s - \overline{x}_s \hat{b} \\ \hat{b} \end{pmatrix}$$

$$= n\overline{y}_s + \left((N-n), \sum_{j=n+1}^N x_j \right) \begin{pmatrix} \overline{y}_s - \overline{x}_s \hat{b} \\ \hat{b} \end{pmatrix} = n\overline{y}_s + (N-n)\overline{y}_s - (N-n)\overline{x}_s \hat{b}$$

$$+ (N\overline{x} - n\overline{x}_s)\hat{b} = N\overline{y}_s + N(\overline{x} - \overline{x}_s)\hat{b} = N\left[\overline{y}_s + (\overline{x} - \overline{x}_s)\hat{b}\right].$$

The error variance is

$$
\begin{aligned}
V_M =\ & \mathrm{var}_M(\hat{T}-T) = \boldsymbol{\gamma}_r'(\boldsymbol{V}_r + \boldsymbol{X}_r(\boldsymbol{X}_s'\boldsymbol{V}_s^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_r')\boldsymbol{\gamma}_r = \sigma^2 \mathbf{1}_{N-n}' \\
& \cdot \left(\boldsymbol{I}_{N-n} + \begin{pmatrix} 1 & x_{n+1} \\ \vdots & \vdots \\ \mathbf{1} & x_N \end{pmatrix} \frac{\begin{pmatrix} \sum_{j=1}^n x_j^2 & -\sum_{j=1}^n x_j \\ -\sum_{j=1}^n x_j & n \end{pmatrix}}{n\left(\sum_{j=1}^n x_j^2\right) - \left(\sum_{j=1}^n x_j\right)^2} \begin{pmatrix} 1 & \dots & 1 \\ x_{n+1} & \dots & x_N \end{pmatrix} \right) \mathbf{1}_{N-n} \\
& = \sigma^2(N-n) + \sigma^2 \left(N-n,\ \sum_{j=n+1}^N x_j\right) \frac{\begin{pmatrix} \sum_{j=1}^n x_j^2 & -\sum_{j=1}^n x_j \\ -\sum_{j=1}^n x_j & n \end{pmatrix}}{n\sum_{j=1}^n (x_j - \overline{x}_s)^2} \begin{pmatrix} N-n \\ \sum_{j=n+1}^N x_j \end{pmatrix} \\
& = \sigma^2(N-n)\left\{1 + \frac{(N-n)\sum_{j=1}^n x_j^2 - 2\sum_{j=1}^n x_j \sum_{j=n+1}^N x_j + \frac{n}{N-n}\left(\sum_{j=n+1}^N x_j\right)^2}{n\sum_{j=1}^n (x_j - \overline{x}_s)^2}\right\} \\
& = \sigma^2(N-n)\left\{1 + \frac{A}{B}\right\}.
\end{aligned}
$$

By taking into account that $n\sum_{j=1}^n x_j^2 - \left(\sum_{j=1}^n x_j\right)^2 = n\sum_{j=1}^n (x_j - \overline{x}_s)^2$ and that $f = n/N$, we get

$$
\begin{aligned}
A =\ & (N-n)\sum_{j=1}^n (x_j - \overline{x}_s)^2 + \frac{N-n}{n}\left(\sum_{j=1}^n x_j\right)^2 - 2\sum_{j=1}^n x_j \left(\sum_{j=1}^N x_j - \sum_{j=1}^n x_j\right) \\
& + \frac{n}{N-n}\left(\sum_{j=1}^N x_j - \sum_{j=1}^n x_j\right)^2 = \frac{1}{n(N-n)}\left\{ n(N-n)^2 \sum_{j=1}^n (x_j - \overline{x}_s)^2 + (N-n)^2 \left(\sum_{j=1}^n x_j\right)^2 \right. \\
& - \left. 2n(N-n)\left[\sum_{j=1}^N x_j \sum_{j=1}^n x_j - \left(\sum_{j=1}^n x_j\right)^2\right] + n^2\left[\left(\sum_{j=1}^N x_j\right)^2 - 2\sum_{j=1}^N x_j \sum_{j=1}^n x_j + \left(\sum_{j=1}^n x_j\right)^2\right]\right\} \\
& = \frac{1}{n(N-n)}\left\{ n(N-n)^2 \sum_{j=1}^n (x_j - \overline{x}_s)^2 + N^2\left(\sum_{j=1}^n x_j\right)^2 - 2nN\sum_{j=1}^N x_j \sum_{j=1}^n x_j + n^2\left(\sum_{j=1}^N x_j\right)^2 \right\}.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
V_M &= \sigma^2(N-n)\left\{ \frac{N}{n} + \frac{N^2\left(\sum_{j=1}^n x_j\right)^2 - 2nN\sum_{j=1}^N x_j \sum_{j=1}^n x_j + n^2\left(\sum_{j=1}^N x_j\right)^2}{n^2(N-n)\sum_{j=1}^n (x_j - \overline{x}_s)^2} \right\} \\
&= \frac{\sigma^2(N-n)N}{n}\left\{ 1 + \frac{n^2 N \overline{x}_s^2 - 2n^2 N \overline{x}_s \overline{x} + n^2 N \overline{x}^2}{n(N-n)\sum_{j=1}^n (x_j - \overline{x}_s)^2} \right\}
\end{aligned}
$$

$$= \frac{N^2}{n}(1-f)\sigma^2\left[1+\frac{(\overline{x}_s-\overline{x})^2}{(1-f)\frac{1}{n}\sum_{j=1}^{n}(x_j-\overline{x}_s)^2}\right].$$

4.6 R Codes for BLUPs

This section gives R codes for calculating the expansive estimator and the linear regression estimator described in Examples 4.3 and 4.4, respectively. The target is estimating the population average of variable INCOME from the survey data file `LFS20.txt`. Let us note that there is a small difference with respect to Examples 4.3 and 4.4 since in the application we estimate population means instead of population totals. That means that the derived formulas for BLUP and error variance must be divided by N and N^2, respectively.

The following code reads the data file:

```
# Survey data
dat <- read.table("LFS20.txt", header=TRUE, sep = "\t", dec = ".")
# Sample size
n <- nrow(dat); n
# Rename variables
y <- dat$INCOME; x <- dat$REGISTERED
# Auxiliary data
dataux <- read.table("Nds20.txt", header=TRUE, sep = "\t", dec = ".")
```

Continuation of Example 4.3 (Expansive Estimator of the Average Income)

```
mod1 <- lm(y~1)                                   # Assumed model
sigma12 <- as.numeric(anova(mod1)[3])              # Model error variance
beta1 <- as.numeric(mod1$coefficients)             # Regression parameter
Npop <- sum(dataux$N)                              # Population size
f <- n/Npop; f                                     # Sampling fraction
Mincome1 <- beta1                                  # Expansive estimator
Mincome1; mean(y)                                  # Checking
varMincome1 <- (1-f)*sigma12/n                     # Estimator error variance
```

Continuation of Example 4.4 (Linear Regression Estimator of the Average Income)

```
mod2 <- lm(y~x)                                    # Assumed model
sigma22 <- anova(mod2)[2,3]                        # Model error variance
beta2 <- mod2$coefficients                         # Regression parameters
Npop <- sum(dataux$N)                              # Population size
f <- n/Npop; f                                     # Sampling fraction
ymean <- mean(y); xmean <- mean(x)                 # Sample means of y and x
Xmean <- sum(dataux$reg)/sum(dataux$N)             # Population mean of x
Mincome2 <- as.numeric(ymean+(Xmean-xmean)*        # Linear regression estimator
                       beta2[2])
xvar <- (n-1)*var(x)/n; xvar                       # Sample variance of x
varMincome2 <- ((1-f)*sigma22/n)*
           (1+((xmean-Xmean)^2/((1-f)*xvar)))      # Estimator error variance
```

The R code to save the results is

```
model1 <- c(beta1, NA, sigma12, Mincome1, varMincome1)
model2 <- c(beta2, sigma22, Mincome2, varMincome2)
labels <- c("intercept", "beta1", "sigma2", "Mincome", "Mincome variance")
output <- data.frame(labels, model1, model2)
```

For model 1 (introduced in Example 4.3) and model 2 (introduced in Example 4.4), Table 4.2 presents the estimated intercept (intercept), covariate regression coeffi-

Table 4.2 Results of expansive estimator (model 1) and linear regression estimator (model 2) of the average income

	model 1	model 2
intercept	46,925.32	47,709.49
beta1		−9686.78
sigma2	157,023,019.44	150,178,380.43
Mincome	46,925.32	46,881.85
Mincome variance	148,684.02	142,241.59

cient (beta1), error variance (sigma2), average income (Mincome), and variance of average income estimator (Mincome variance).

For estimating the population average income, the linear regression estimator (derived under model 2) has lower estimated variance than the expansive estimator (derived under model 1).

References

Bolfarine, H., Zacks, S.: Prediction Theory for Finite Populations. Springer, Berlin (1992)

Cassel, C., Särndal, C.E., Wretman, J.: Foundations of Inference in Survey Sampling. Wiley, New York (1977)

Valliant, R., Dorfman, A.H., Royall, R.M.: Finite Population Sampling and Inference. A Prediction Approach. John Wiley, New York (2000)