Bayesian Methods in Finite Population Sampling

Mark Sebastian Hamner
Department of Mathematics and Computer Science, Texas Woman's University, Denton, Texas

John W. Seaman, Jr. and Dean M. Young
Institute of Statistics, Baylor University, Waco, Texas

Key Words: Finite population, Population total, General regression estimator, Bayesian prediction

Introduction

The focus of this paper is on the prediction of a finite population total T by taking a sample of size n from a population of size N units. For example, suppose we wish to estimate the average total amount a university student can expect to borrow before graduation in a certain region of the country. We can use sample survey information from graduating students to predict the total loans for all students in that region and from that we can estimate the average total amount a student will borrow.

Classical theory models the data collection procedure with a sampling design, a probability function defined on the sample space, S, of all possible samples of size n. The sampling design along with unbiasedness requirements yields a frequentist approach to relating observed with unobserved population units. In contrast, a superpopulation model provides the stochastic structure for Bayesian inferential purposes. In this paper, under an ignorable design, we develop a fully Bayesian approach to the prediction of the population total T that is a function of the mean of the posterior predictive distribution. We show that the Bayesian estimator has the general form of the familiar general regression estimator found in model assisted survey sampling.

1. Matrix Empirical Bayes π-estimator

Gosh and Meeden (1997, p.164) present an empirical Bayes estimator of the finite population mean for a quantity of interest $y_1, y_2, \ldots, y_N$. In this section we will present the results of Gosh and Meeden in a more convenient form using matrices. Let $\mathbf{s}$ be a sample of n elements from $\{1,2 \ldots, N\}$ and let $\mathbf{r}$ represent the ($N - n$) non-sampled elements.

We can rewrite the results of Gosh and Meeden's population average estimate to an empirical population total estimate in matrix form by assuming the superpopulation model

$$\mathbf{Y} = \mathbf{A1\theta} + \mathbf{e}$$

where $\mathbf{Y}$ is $N \times 1$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y_s} \\ \hdashline \mathbf{y_r} \end{bmatrix}$$

such that $\mathbf{y_s}$ is an $n \times 1$ vector of sampled units, and $\mathrm{y}_\mathbf{r}$ is an $(N-n) \times 1$ vector of the non-sampled units. Define $\mathbf{A}$ to be the known $N \times N$ diagonal matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{A_r} \end{bmatrix},$$

where $\mathbf{A_s}$ is an $n \times n$ diagonal matrix with diagonal elements $a_1, a_2, \ldots, a_n$, and $\mathbf{A_r}$ is $(N-n) \times (N-n)$ diagonal matrix with diagonal elements $a_{n+1}, a_{n+2}, \ldots, a_N$. Let $\mathbf{1}$ be an $N \times 1$ vector of ones

$$\mathbf{1} = \begin{bmatrix} \mathbf{1_s} \\ \hdashline \mathbf{1_r} \end{bmatrix},$$

where $\mathbf{1_s}$ is an $n \times 1$ vector of ones, $\mathbf{1_r}$ is a $(N-n) \times 1$ vector of ones, θ is a 1×1 unknown parameter, and $\mathbf{e} \sim N(\mathbf{0},\mathbf{V})$, where $N(\mathbf{u},\mathbf{W})$ denotes a multivariate normal distribution with mean vector $\mathbf{u}$ and covariance matrix $\mathbf{W}$. Assume $\mathbf{V}$ is a known positive definite diagonal matrix that may be partitioned as

$$\mathbf{V} = \begin{bmatrix} \mathbf{V_{ss}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V_{rr}} \end{bmatrix},$$

where $\mathbf{V_{ss}}$ is an $n \times n$ diagonal matrix with diagonal elements σ_i^2 for $i = 1, 2, \ldots$, n, and $\mathbf{V_{rr}}$ is an $(N-n) \times (N-n)$ diagonal matrix with diagonal elements σ_j^2 for $j = n+1, n+2, \ldots, N$. To estimate the population average first note that the population total is

$$T = \mathbf{1}'\mathbf{Y} = \begin{bmatrix} \mathbf{1}'_\mathbf{s} & \mathbf{1}'_\mathbf{r} \end{bmatrix} \begin{bmatrix} \mathbf{y_s} \\ \hdashline \mathbf{y_r} \end{bmatrix} = \mathbf{1}'_\mathbf{s}\mathbf{y_s} + \mathbf{1}'_\mathbf{r}\mathbf{y_r}$$

Thus, our estimate of the population total involves estimating the non-sampled units $\mathbf{y_r}$ of the population total T. The estimate of the population total for a fully Bayesian method under squared error loss requires the posterior predictive expectation $E(\mathbf{y_r}|\mathbf{y_s})$ for estimation of $\mathbf{y_r}$. This posterior expectation can be obtained by the nested expectations

$$E(\mathbf{y_r}|\mathbf{y_s}) = E_\theta\left[E_{\mathbf{y_r}}(\mathbf{y_r} \mid \mathbf{y_s}, \boldsymbol{\theta},) \middle| \mathbf{y_s}\right],$$

where $E_{\mathbf{y_r}}(\cdot)$ is the expectation with respect to $\mathbf{y_r}$ and $E_\theta(\cdot)$ is the expectation with respect to the posterior of $\boldsymbol{\theta}$. Under model $\mathbf{Y} = \mathbf{A1\theta} + \mathbf{e}$, $E_{\mathbf{y_r}}(\mathbf{y_r} \mid \mathbf{y_s}, \theta,) = \mathbf{A_r 1_r \theta}$. Now, instead of taking the expectation

$E_{\boldsymbol{\theta}}(\mathbf{A_r 1_r \theta})$, an empirical Bayes result ensues by estimating the unknown parameter $\boldsymbol{\theta}$ using a weighted least squares estimate. Let $\mathbf{X_s} = \mathbf{A_s 1_s}$, then the least squares estimate of $\boldsymbol{\theta}$ is

$$\hat{\theta}_{LE} = \left(\mathbf{X}_s' \mathbf{V}_{ss}^{-1} \mathbf{X}_s\right)^{-1} \mathbf{X}_s' \mathbf{V}_{ss}^{-1} \mathbf{y}_s$$

$$= \left(\mathbf{1}_s' \mathbf{A}_s \mathbf{V}_{ss}^{-1} \mathbf{A}_s \mathbf{1}_s\right)^{-1} \mathbf{1}_s' \mathbf{A}_s \mathbf{V}_{ss}^{-1} \mathbf{y}_s \qquad (1.1)$$

Thus, an empirical Bayes estimate of the population total is

$$\hat{T}_{EB} = \mathbf{1}_s' \mathbf{y}_s + \mathbf{1}_r' E_{\mathbf{y}_r}\left(\mathbf{y}_r \mid \mathbf{y}_s, \hat{\theta}_{LE},\right)$$

$$= \mathbf{1}_s' \mathbf{y}_s + \mathbf{1}_r' \mathbf{A}_r \mathbf{1}_r \hat{\theta}_{LE}$$

$$= \mathbf{1}_s' \mathbf{y}_s$$

$$+ \mathbf{1}_r' \mathbf{A}_r \mathbf{1}_r \left(\mathbf{1}_s' \mathbf{A}_s \mathbf{V}_{ss}^{-1} \mathbf{A}_s \mathbf{1}_s\right)^{-1} \mathbf{1}_s' \mathbf{A}_s \mathbf{V}_{ss}^{-1} \mathbf{y}_s . \qquad (1.2)$$

Next we use the matrix form of the empirical Bayes estimate and derive the Horvitz-Thompson estimator using the assumptions noted by Gosh and Meeden. Let π_i represent the inclusion probability for an individual k from a finite population of size N where $i = 1,2,\ldots, N$. Define $\boldsymbol{\pi}$ to be an $N \times N$ diagonal matrix whose diagonal elements, $\pi_{ii} \equiv \pi_i$, $i = 1,2,\ldots N$, are the inclusion probabilities. We can write $\boldsymbol{\pi}$ as

$$\boldsymbol{\pi} = \left[\begin{array}{c|c} \boldsymbol{\pi}_s & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\pi}_r \end{array}\right]$$

where $\boldsymbol{\pi}_s$ is an $n \times n$ diagonal matrix and $\boldsymbol{\pi}_r$ is an $(N - n) \times (N - n)$ diagonal matrix. One of the conditions of the inclusion probabilities is that $\sum_{i=1}^{N} \pi_i = n$. In matrix form this condition for the sum of the diagonal elements of $\boldsymbol{\pi}$ can be written as

$$n = \mathbf{1}' \boldsymbol{\Pi} \mathbf{1} = \left[\mathbf{1}_s' \mid \mathbf{1}_r'\right] \left[\begin{array}{c|c} \boldsymbol{\pi}_s & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\pi}_r \end{array}\right] \left[\begin{array}{c} \mathbf{1}_s \\ \hline \mathbf{1}_r \end{array}\right]$$

$$= \mathbf{1}_s' \boldsymbol{\pi}_s \mathbf{1}_s + \mathbf{1}_r' \boldsymbol{\pi}_r \mathbf{1}_r ,$$

which implies

$$\mathbf{1}_r' \boldsymbol{\pi}_r \mathbf{1}_r = n - \mathbf{1}_s' \boldsymbol{\pi}_s \mathbf{1}_s \ . \qquad (1.3)$$

Let $\mathbf{A} = \boldsymbol{\pi}$, $\mathbf{V} = \boldsymbol{\pi}(\mathbf{I} - \boldsymbol{\pi})^{-1}\boldsymbol{\pi}$ where $\mathbf{I}$ is the $N \times N$ identity matrix

$$\mathbf{I} = \left[\begin{array}{c|c} \mathbf{I}_s & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_r \end{array}\right]$$

such that $\mathbf{I}_s$ is an $n \times n$ identity matrix and $\mathbf{I}_r$ is an $(N - n) \times (N - n)$ identity matrix. Using equations (1.3) and the assumptions made for $\mathbf{A}$ and $\mathbf{V}$, the empirical Bayes estimator (1.2) becomes

$$\hat{T}_{EB} = \left[\mathbf{1}_s' + \mathbf{1}_s' \left(\mathbf{I}_s - \boldsymbol{\pi}_s\right) \boldsymbol{\pi}_s^{-1}\right] \mathbf{y}_s$$

$$= \mathbf{1}_s' \boldsymbol{\pi}_s^{-1} \mathbf{y}_s \qquad (1.4)$$

Equation (1.4) is the matrix representation of the Horvitz-Thompson estimator, sometimes referred to as the *π-estimator*, for the population total.

2. *General Regression Estimator*

In this section we introduce the general regression estimator and present it in a more convenient form using matrices. This facilitates subsequent derivations.

As a means to possibly improve the basic $\boldsymbol{\pi}$-estimator using auxiliary information, Sarndal et al. (1992) employ classical sampling design theory, using inclusion probabilities, and the regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ where $\mathbf{e} \sim N(\mathbf{0},\mathbf{V})$. The latter is used, however, only as a means to obtain an estimate of $\boldsymbol{\beta}$. Hence, unbiasedness and variance expressions are derived under the sampling design. In short, Sarndal et al. (1992) do not assume that the regression model generated the sample. Thus, the *general regression estimator* (GRE) derived in Sarndal et al. (1992) is *model assisted* but not model dependent. Sarndal et al. (p.225, 1992) define the GRE as

$$\hat{T}_{\mathrm{GRE}} \equiv \sum_{k=1}^{n} \frac{y_k}{\pi_k} + \sum_{j=1}^{p} \hat{\beta}_j \left(\sum_{k=1}^{N} X_{jk} - \sum_{k=1}^{n} \frac{X_{jk}}{\pi_k} \right),$$

where y_k is a variable of interest, like loan amount, for $k = 1, 2,\ldots, N$, π_k is the inclusion probability, $\hat{\beta}_j$ is an unknown regression coefficient for $j = 1, 2,\ldots, p$, and x_{jk} is a known auxiliary variable. Notice that the GRE is equal to the π-estimator plus an adjustment term. Using a regression model to assist in the estimate for $\boldsymbol{\beta} \equiv (\beta_1,\ldots,\beta_p)'$, Sardnal et al. (1992, p. 228) suggest the estimate

$$\hat{\boldsymbol{\beta}} \equiv \left(\sum_{k=1}^{n} \frac{\mathbf{X}_k \mathbf{X}_k'}{\sigma_k^2 \pi_k} \right)^{-1} \sum_{k=1}^{n} \frac{\mathbf{X}_k y_k}{\sigma_k^2 \pi_k} .$$

Still, under a simple random sampling design in which $\boldsymbol{\pi} = \left(\frac{n}{N}\right)^{-1} \mathbf{I}$, $\hat{\boldsymbol{\beta}}$ is the least squares estimator $\hat{\boldsymbol{\beta}}_s \equiv \left(\mathbf{X}_s' \boldsymbol{\Sigma}^{-1} \mathbf{X}_s\right)^{-1} \mathbf{X}_s' \boldsymbol{\Sigma}^{-1} \mathbf{y}_s$, where $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite covariance matrix. Now, using the notation established in Section 1, the GRE can be rewritten as

$$\hat{T}_{\mathrm{GRE}} = \mathbf{1}' \mathbf{X} \hat{\boldsymbol{\beta}}_s + \mathbf{1}_s' \boldsymbol{\pi}_s^{-1} \left(\mathbf{y}_s - \mathbf{X}_s' \hat{\boldsymbol{\beta}}_s \right) . \qquad (2.1)$$

Assuming $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$ and a simple random sampling design we can express (2.1) as

$$\hat{T}_{\text{GRE}} = \mathbf{1}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbf{s}} + \frac{N}{n}\mathbf{1}'_{\mathbf{s}}[\mathbf{I}_{\mathbf{s}} - \mathbf{P}_{\mathbf{s}}]\mathbf{y}_{\mathbf{s}}$$

where, again, $\hat{\boldsymbol{\beta}}_{\mathbf{s}} = (\mathbf{X}'_{\mathbf{s}}\mathbf{X}_{\mathbf{s}})^{-1}\mathbf{X}_{\mathbf{s}}\mathbf{y}_{\mathbf{s}}$ is the least square estimator of $\boldsymbol{\beta}$ and $\mathbf{P}_{\mathbf{s}} = \mathbf{X}_{\mathbf{s}}(\mathbf{X}'_{\mathbf{s}}\mathbf{X}_{\mathbf{s}})^{-1}\mathbf{X}'_{\mathbf{s}}$ is the projection matrix onto the column space of $\mathbf{X}_{\mathbf{s}}$.

3. An Empirical Bayes General Regression Estimator

In this section we introduce a superpopulation model and obtain an empirical Bayes estimator of the population total. Our empirical Bayes estimator of the population total $T = \mathbf{1}'_{\mathbf{s}}\mathbf{y}_{\mathbf{s}} + \mathbf{1}'_{\mathbf{r}}\mathbf{y}_{\mathbf{r}}$ requires derivation of the mean of the posterior predictive distribution $E(\mathbf{y}_{\mathbf{r}}|\mathbf{y}_{\mathbf{s}})$. Finally, we show that our empirical Bayes estimator is equal to the general regression estimator.

Royal and Pfeffermann (1982) focus attention on necessary assumptions needed for robustness of their statistical procedures for predicting the population total T given $\mathbf{y}_{\mathbf{s}}$. They consider their procedure robust if the posterior probability distribution of T using model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ is not greatly affected by instead using model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \mathbf{e}$ where $\mathbf{U}$ are additional regressors with a fixed coefficient vector $\boldsymbol{\gamma}$. Niewenbroek and Renssen (1997) consider estimating the population total T by using two or more surveys to obtain *common variables*, $\mathbf{U}$, as additional regressors observed in both surveys for which the corresponding population totals are unknown and combining them with auxiliary *variables*, $\mathbf{X}$, which have known population totals. They then use these common variables as a tool to improve the estimate of the population total by using what they call an adjusted general regression estimator.

Consider the superpopulation model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \mathbf{e}. \tag{3.1}$$

Here $\mathbf{X}$ is $N \times p$ with full column rank and

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{\mathbf{s}} \\ \hline \mathbf{X}_{\mathbf{r}} \end{bmatrix}$$

such that $\mathbf{X}_{\mathbf{s}}$ is $n \times p$ and $\mathbf{X}_{\mathbf{r}}$ is $(N-n) \times p$. Furthermore, $\mathbf{U}$ is $N \times q$ with full column rank and

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_{\mathbf{s}} \\ \hline \mathbf{U}_{\mathbf{r}} \end{bmatrix}$$

such that $\mathbf{U}_{\mathbf{s}}$ is $n \times q$ and $\mathbf{U}_{\mathbf{r}}$ is $(N-n) \times q$. Finally, γ is $q \times 1$ and $\mathbf{e} \sim N(\mathbf{0},\mathbf{V})$, with $\mathbf{V}$ known such that

$$\mathbf{V} = \left[\begin{array}{c|c} \mathbf{V}_{\mathbf{ss}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{V}_{\mathbf{rr}} \end{array}\right].$$

To estimate T, we shall utilize $E(\mathbf{y}_{\mathbf{r}}|\mathbf{y}_{\mathbf{s}})$. To obtain the latter we first estimate $\boldsymbol{\beta}$ by regressing $\mathbf{y}_{\mathbf{s}}$ on $\mathbf{X}_{\mathbf{s}}$ resulting in an empirical Bayesian procedure. We have

$$\boldsymbol{\beta}|\mathbf{y}_{\mathbf{s}} \sim \mathrm{N}\left(\left(\mathbf{X}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{X}_{\mathbf{s}}\right)^{-1}\mathbf{X}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{y}_{\mathbf{s}},\ \left(\mathbf{X}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{X}_{\mathbf{s}}\right)^{-1}\right).$$

Substituting into (3.1) we have

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathbf{s}} + \mathbf{U}\boldsymbol{\gamma} + \mathbf{e}. \tag{3.2}$$

Using model (3.2) we want an estimator of the population total $T = \mathbf{1}'_{\mathbf{s}}\mathbf{y}_{\mathbf{s}} + \mathbf{1}'_{\mathbf{r}}\mathbf{y}_{\mathbf{r}}$. We must obtain the mean of the posterior predictive distribution $\pi(\mathbf{y}_{\mathbf{r}}|\mathbf{y}_{\mathbf{s}})$.

Using the improper prior $\pi(\gamma) \propto$ constant, the posterior predictive distribution has the form

$$\pi(\mathbf{y}_{\mathbf{r}}|\mathbf{y}_{\mathbf{s}}) = \int_{\mathbb{R}^q} f\left(\mathbf{y}_{\mathbf{r}} \middle| \gamma, \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{U}_{\mathbf{r}}, \mathbf{V}_{\mathbf{rr}}\right) \times \pi\left(\gamma \middle| \mathbf{y}_{\mathbf{s}}, \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{U}_{\mathbf{s}}, \mathbf{V}_{\mathbf{ss}}\right) d\gamma, \tag{3.3}$$

where the likelihood of $\mathbf{y}_{\mathbf{r}} | \gamma, \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{X}_{\mathbf{r}}, \mathbf{U}_{\mathbf{r}}, \mathbf{V}_{\mathbf{rr}}$ is

$$f\left(\mathbf{y}_{\mathbf{r}} \middle| \gamma, \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{X}_{\mathbf{r}}, \mathbf{U}_{\mathbf{r}}, \mathbf{V}_{\mathbf{rr}}\right) = N\left(\mathbf{y}_{\mathbf{r}} \middle| \mathbf{X}_{\mathbf{r}}\hat{\boldsymbol{\beta}}_{\mathbf{s}} + \mathbf{U}_{\mathbf{r}}\gamma, \mathbf{V}_{\mathbf{rr}}\right). \tag{3.4}$$

The posterior of $\gamma | \mathbf{y}_{\mathbf{s}}, \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{U}_{\mathbf{s}}, \mathbf{V}_{\mathbf{ss}}$ can be shown to be

$$\boldsymbol{\pi}(\gamma \mid \mathbf{y}_{\mathbf{s}}, \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{U}_{\mathbf{s}}, \mathbf{V}_{\mathbf{ss}}) = N\left(\gamma \middle| (\mathbf{U}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{U}_{\mathbf{s}})^{-1}\mathbf{U}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}(\mathbf{y}_{\mathbf{s}} - \mathbf{X}_{\mathbf{s}}\hat{\boldsymbol{\beta}}_{\mathbf{s}}), \left(\mathbf{U}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{U}_{\mathbf{s}}\right)^{-1}\right). \tag{3.5}$$

Thus, using (3.4) and (3.5), (3.3) becomes

$$\begin{aligned}\pi(\mathbf{y}_{\mathbf{r}}|\mathbf{y}_{\mathbf{s}}) &= \int_{\mathbb{R}^q} f\left(\mathbf{y}_{\mathbf{r}} \middle| \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{U}_{\mathbf{r}}, \gamma, \sigma^2\right) \pi\left(\gamma \middle| \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{y}_{\mathbf{s}}, \mathbf{U}_{\mathbf{s}}, \sigma^2\right) d\gamma \\ &= \int_{\mathbb{R}^q} N\left(\mathbf{y}_{\mathbf{r}} \middle| \mathbf{X}_{\mathbf{r}}\hat{\boldsymbol{\beta}}_{\mathbf{s}} + \mathbf{U}_{\mathbf{r}}\gamma, \mathbf{V}_{\mathbf{rr}}\right) \times \qquad (3.6) \\ &\quad N\left(\gamma \middle| (\mathbf{U}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{U}_{\mathbf{s}})^{-1}\mathbf{U}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}(\mathbf{y}_{\mathbf{s}} - \mathbf{X}_{\mathbf{s}}\hat{\boldsymbol{\beta}}_{\mathbf{s}}), \left(\mathbf{U}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{U}_{\mathbf{s}}\right)^{-1}\right) d\gamma.\end{aligned}$$

Using equation (3.6), the mean of the posterior predictive distribution $\pi(\mathbf{y}_{\mathbf{r}}|\mathbf{y}_{\mathbf{s}})$ is

$$\begin{aligned}E(\mathbf{y}_{\mathbf{r}}|\mathbf{y}_{\mathbf{s}}) &= \int_{\mathbb{R}^q}\int_{\mathbb{R}^{N-n}} \mathbf{y}_{\mathbf{r}} N\left(\mathbf{y}_{\mathbf{r}} \middle| \mathbf{X}_{\mathbf{r}}\hat{\boldsymbol{\beta}}_{\mathbf{s}} + \mathbf{U}_{\mathbf{r}}\gamma, \mathbf{V}_{\mathbf{rr}}\right) \\ &= E_{\gamma}\left[E_{\mathbf{y}_{\mathbf{r}}}\left(\mathbf{y}_{\mathbf{r}} \middle| \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{U}_{\mathbf{r}}, \gamma, \mathbf{V}_{\mathbf{rr}}\right) \middle| \hat{\boldsymbol{\beta}}_{\mathbf{s}}, \mathbf{y}_{\mathbf{s}}, \mathbf{U}_{\mathbf{s}}, \mathbf{V}_{\mathbf{ss}}\right] \\ &= \mathbf{X}_{\mathbf{r}}\hat{\boldsymbol{\beta}}_{\mathbf{s}} + \mathbf{U}_{\mathbf{r}}(\mathbf{U}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{U}_{\mathbf{s}})^{-1}\mathbf{U}'_{\mathbf{s}}\mathbf{V}_{\mathbf{ss}}^{-1}\left(\mathbf{y}_{\mathbf{s}} - \mathbf{X}_{\mathbf{s}}\hat{\boldsymbol{\beta}}_{\mathbf{s}}\right)\end{aligned}$$

where $E_{\mathbf{y}_{\mathbf{r}}}(\cdot)$ and $E_{\gamma}(\cdot)$ denote expectations with respect to the distributions of $\mathbf{y}_{\mathbf{r}}$ and $\boldsymbol{\gamma}$, respectively.

Thus, our empirical Bayesian estimate of the population total is

$$\hat{T}_{\text{EB}} = \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r' E(\mathbf{y}_r|\mathbf{y}_s)$$
$$= \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r'\Big[\mathbf{X}_r\hat{\boldsymbol\beta}_s$$
$$+ \mathbf{U}_r(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}_s\right)\Big].$$

A special case to our Empirical Bayes estimator leads to an Empirical Bayesian justification for the general regression estimator. First, we use the notation in section 1 and let $\mathbf{V} = \boldsymbol\pi(\mathbf{I} - \boldsymbol\pi)^{-1}\boldsymbol\pi$ and $\mathbf{U} = \boldsymbol\pi\mathbf{1}'$. Then, using equation (1.3) the mean of $\pi(\boldsymbol\gamma \mid \mathbf{y}_s, \hat{\boldsymbol\beta}_s, \mathbf{U}_s, \mathbf{V}_{ss})$ becomes

$$(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}_s\right) =$$
$$[\mathbf{1}_r'\boldsymbol\pi_r\mathbf{1}_r]^{-1}\mathbf{1}_s'(\mathbf{I}_s - \boldsymbol\pi_s)\boldsymbol\pi_s^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}_s\right). \qquad (3.8)$$

Using equation (3.8) then the right hand side of (3.7) becomes

$$\hat{T}_{\text{EB}} = \mathbf{1}'\mathbf{X}\hat{\boldsymbol\beta}_s + \mathbf{1}_s'\boldsymbol\pi_s^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}_s\right).$$

Therefore, with $\mathbf{V} = \boldsymbol\pi(\mathbf{I} - \boldsymbol\pi)^{-1}\boldsymbol\pi$, $\mathbf{U} = \boldsymbol\pi\mathbf{1}'$ and using the mean of the posterior predictive distribution $\pi(\mathbf{y}_r|\mathbf{y}_s)$ as our estimator for the unknown quantity $\mathbf{y}_r$ in T, we have shown that our empirical Bayesian estimator for the population total is the general regression estimator (2.1).

4. *The Superpopulation Model*

Consider the superpopulation model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}\gamma + \mathbf{e}, \qquad (4.1)$$

where $\mathbf{X}$ is $N \times p$ with full column rank, $\mathbf{U}$ is $N \times q$ with full column rank, $\boldsymbol\beta$ is $p \times 1$, $\boldsymbol\gamma$ is $q \times 1$, $\mathbf{e} \sim N(\mathbf{0},\mathbf{V})$, where $N(\mathbf{u},\mathbf{W})$ denotes a multivariate normal distribution with mean vector $\mathbf{u}$ and covariance matrix $\mathbf{W}$. Assume $\mathbf{V}$ is a known positive definite matrix that may be partitioned as

$$\mathbf{V} = \left[\begin{array}{c|c} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \hline \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{array}\right],$$

where $\mathbf{V}_{ss}$ is $n \times n$, $\mathbf{V}_{sr}$ is $n \times (N-n)$, $\mathbf{V}_{rs}$ is $(N-n) \times n$, and $\mathbf{V}_{rr}$ is $(N-n) \times (N-n)$. Assume $\boldsymbol\beta$ and $\boldsymbol\gamma$ are conditionally independent, given $\mathbf{y}_s$, each with improper uniform prior distributions. Using superpopulation model (4.1), we shall obtain the posterior predictive distribution of $\mathbf{y}_r|\mathbf{y}_s$. Then, under squared error loss, we use the mean of the posterior predictive distribution, $E(\mathbf{y}_r|\mathbf{y}_s)$, to estimate $\mathbf{1}_r'\mathbf{y}_r$ in the population total $T = \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r'\mathbf{y}_r$.

The predictive distribution is (Geisser, 1993, p.49)

$$\pi(\mathbf{y}_r|\mathbf{y}_s) = \int_{\mathbb{R}^p}\int_{\mathbb{R}^q} f\left(\mathbf{y}_r|\mathbf{y}_s,\gamma,\beta\right)\pi\left(\gamma|\beta,\mathbf{y}_s\right)\pi\left(\beta|\mathbf{y}_s\right)d\gamma d\beta, \qquad (4.2)$$

where f denotes the appropriate joint or conditional density and $\mathbb{R}^v$ denotes v-dimensional real space. Notice that, if $\mathbf{y}_r$ is independent of $\mathbf{y}_s$, then

$$\pi(\mathbf{y}_r|\mathbf{y}_s) = \int_{\mathbb{R}^p}\int_{\mathbb{R}^q} f\left(\mathbf{y}_r|\gamma,\beta\right)\pi\left(\gamma|\beta,\mathbf{y}_s\right)\pi\left(\beta|\mathbf{y}_s\right)d\gamma d\beta.$$

In the following sections we derive the distributions in the integrand of (4.2).

5. *The Marginal Density of* $\mathbf{y}_r$

In general, independence between $\mathbf{y}_r$ and $\mathbf{y}_s$ may not obtain. Thus, to derive the conditional distribution $f(\mathbf{y}_r|\mathbf{y}_s,\beta,\gamma)$ we use a well-known method of nonsingular linear transformation from multivariate analysis. The distribution of $\mathbf{y}_r$ is

$$f(\mathbf{y}_r|\mathbf{y}_s,\gamma,\beta) = N\Big(\mathbf{y}_r \Big| (\mathbf{X}_r\beta + \mathbf{U}_r\gamma) + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - (\mathbf{X}_s\beta + \mathbf{U}_s\gamma)\right),$$
$$\mathbf{V}_{rr} - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{V}_{sr}\Big) \qquad (5.1)$$

and the distribution of $\mathbf{y}_s$

$$f(\mathbf{y}_s \mid \gamma,\beta) = \mathrm{N}(\mathbf{y}_s \mid \mathbf{X}_s\beta + \mathbf{U}_s\gamma, \mathbf{V}_{ss}).$$

6. *Posterior Distributions for β and γ*

In this section we obtain the conditional and marginal components of the posterior $\pi(\gamma,\beta|\mathbf{y}_s) = \pi(\gamma|\beta,\mathbf{y}_s)\,\pi(\beta|\mathbf{y}_s)$. Since $\pi(\gamma,\beta) \propto$ constant, we have

$$\pi(\gamma,\beta|\mathbf{y}_s) = \pi(\gamma|\beta,\mathbf{y}_s)\,\pi(\beta|\mathbf{y}_s)$$
$$\propto f(\mathbf{y}_s|\gamma,\beta) = N(\mathbf{y}_s|\mathbf{X}_s\beta + \mathbf{U}_s\gamma, \mathbf{V}_{ss}).$$

Thus, the quadratics in the exponential terms in $\pi(\gamma|\beta,\mathbf{y}_s)$ and $\pi(\beta|\mathbf{y}_s)$ are derived from the normal density $N(\mathbf{y}_s|\mathbf{X}_s\beta + \mathbf{U}_s\gamma,\mathbf{V}_{ss})$:

$$\left[\mathbf{y}_s - (\mathbf{X}_s\beta + \mathbf{U}_s\gamma)\right]'\mathbf{V}_{ss}^{-1}\left[\mathbf{y}_s - (\mathbf{X}_s\beta + \mathbf{U}_s\gamma)\right]$$
$$= \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s - \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta - \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma$$
$$- \beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s - \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s + \beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta$$
$$+ \beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma + \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta$$
$$+ \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma. \qquad (6.1)$$

The quadratic form in $\pi(\gamma|\beta,\mathbf{y_s})$ will be derived from the components of equation (6.1) containing gamma coefficients:

$$\begin{aligned}&\gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma - \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma - \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s\\ &+ \beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma + \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta\\ &= \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma - 2\left(\mathbf{y}_s' - \beta'\mathbf{X}_s'\right)\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma. \qquad (6.2)\end{aligned}$$

Since $\mathbf{V}_{ss}^{-1}$ is positive definite and symmetric, then it has full-rank factorization

$$\mathbf{V}_{ss}^{-1} = \mathbf{K}_1'\mathbf{K}_1 \qquad (6.3)$$

where $\mathbf{K}_1$ is a nonsingular $n \times n$ matrix. Let $\tilde{\mathbf{U}}_s$ = $\mathbf{K}_1\mathbf{U}_s$ and define the projection matrix

$$\mathbf{P}_{\tilde{\mathbf{u}}_s} = \tilde{\mathbf{U}}_s(\tilde{\mathbf{U}}_s'\tilde{\mathbf{U}}_s)^{-1}\tilde{\mathbf{U}}_s'.$$

Using (6.3), we can rewrite (6.2) as

$$\begin{aligned}&\gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma - 2\left(\mathbf{y}_s' - \beta'\mathbf{X}_s'\right)\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma\\ &= \gamma'\mathbf{U}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{U}_s\gamma - 2\left(\mathbf{y}_s' - \beta'\mathbf{X}_s'\right)\mathbf{K}_1'\mathbf{K}_1\mathbf{U}_s\gamma\\ &= \left(\tilde{\mathbf{U}}_s\gamma\right)'\left(\tilde{\mathbf{U}}_s\gamma\right) - 2\left[\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\left(\mathbf{y}_s - \mathbf{X}_s\beta\right)\right]'\tilde{\mathbf{U}}_s\gamma. \qquad (6.4)\end{aligned}$$

Define

$$\mathbf{C}_1 = \mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\left(\mathbf{y}_s - \mathbf{X}_s\beta\right). \qquad (6.5)$$

Note that $\mathbf{C}_1$ is a constant with respect to γ. The right hand side of equation (6.4) becomes

$$\begin{aligned}&\left(\tilde{\mathbf{U}}_s\gamma\right)'\left(\tilde{\mathbf{U}}_s\gamma\right) - 2\left[\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\left(\mathbf{y}_s - \mathbf{X}_s\beta\right)\right]'\tilde{\mathbf{U}}_s\gamma\\ &= \left(\tilde{\mathbf{U}}_s\gamma\right)'\left(\tilde{\mathbf{U}}_s\gamma\right) - 2\mathbf{C}_1'\tilde{\mathbf{U}}_s\gamma + \mathbf{C}_1'\mathbf{C}_1 - \mathbf{C}_1'\mathbf{C}_1.\end{aligned}$$

Thus, using (6.3) and (6.5) the quadratic in the exponential term of $\pi(\gamma\,|\,\beta, \mathbf{y}_s)$ is

$$\begin{aligned}\mathbf{Q}_s &\equiv \left(\tilde{\mathbf{U}}_s\gamma\right)'\left(\tilde{\mathbf{U}}_s\gamma\right) - 2\mathbf{C}_1'\tilde{\mathbf{U}}_s\gamma + \mathbf{C}_1'\mathbf{C}_1\\ &= \left(\tilde{\mathbf{U}}_s\gamma - \mathbf{C}_1\right)'\left(\tilde{\mathbf{U}}_s\gamma - \mathbf{C}_1\right)\\ &= \left[\gamma'\tilde{\mathbf{U}}_s' - \left(\mathbf{y}_s - \mathbf{X}_s\beta\right)'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\right]\\ &\quad\times\left[\tilde{\mathbf{U}}_s\gamma - \mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\left(\mathbf{y}_s - \mathbf{X}_s\beta\right)\right]\\ &= \left[\gamma - (\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\beta\right)\right]'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\\ &\quad\times\left[\gamma - (\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\beta\right)\right].\end{aligned}$$

Thus,

$$\begin{aligned}&\pi(\gamma|\beta,\mathbf{y}_s) =\\ &N\left[\gamma\,\middle|\,(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\beta\right), \left(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)^{-1}\right]\\ &\equiv N\left[\gamma\,\middle|\,\mu_\gamma, \mathbf{V}_\gamma\right]. \qquad (6.6)\end{aligned}$$

Now we find the exponential term of the marginal distribution $\pi(\beta|\mathbf{y}_s)$. From (6.5) we have

$$\begin{aligned}\mathbf{C}_1'\mathbf{C}_1 &= \left[\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\left(\mathbf{y}_s - \mathbf{X}_s\beta\right)\right]'\left[\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\left(\mathbf{y}_s - \mathbf{X}_s\beta\right)\right]\\ &= \beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\beta - 2\,\mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\beta\\ &\quad + \mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}\mathbf{y}_s. \qquad (6.7)\end{aligned}$$

Using (6.3), the remaining term of equation (6.1) containing β that was not used to obtain the exponential term in $\pi(\gamma|\beta,\mathbf{y}_s)$ is

$$\begin{aligned}&\beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta - 2\,\mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta\\ &= \beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta - 2\,\mathbf{y}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta. \qquad (6.8)\end{aligned}$$

Thus, the exponential term in $\pi(\beta|\mathbf{y}_s)$ will be derived from the combination of equation (6.7) and equation (6.8):

$$\begin{aligned}&\beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta - \beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\beta\\ &- 2\,\mathbf{y}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta + 2\,\mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\beta\\ &= \beta'\mathbf{X}_s'\mathbf{K}_1'\left(\mathbf{I}_s - \mathbf{P}_{\tilde{\mathbf{u}}_s}\right)\mathbf{K}_1\mathbf{X}_s\beta\\ &- \mathbf{y}_s'\mathbf{K}_1'\left(\mathbf{I}_s - \mathbf{P}_{\tilde{\mathbf{u}}_s}\right)\mathbf{K}_1\mathbf{X}_s\beta. \qquad (6.9)\end{aligned}$$

Let $\mathbf{X}_s$ and $\mathbf{U}_s$ be partitioned as follows

$$\mathbf{X}_s = \left[\mathbf{X}_1 \,\vdots\, \mathbf{X}_2 \,\vdots\, \cdots \,\vdots\, \mathbf{X}_p\right]$$

and

$$\mathbf{U}_s = \left[\mathbf{U}_1 \,\vdots\, \mathbf{U}_2 \,\vdots\, \cdots \,\vdots\, \mathbf{U}_q\right],$$

where $\mathbf{X}_i$ and $\mathbf{U}_j$ are $n \times 1$ column vectors of $\mathbf{X}_s$ and $\mathbf{U}_s$ respectively for i = 1, 2, …, p and j = 1, 2, …, q. Since $\mathbf{V}_{ss}^{-1}$ is symmetric and positive definite then the bilinear form $\mathbf{X}_i'\mathbf{V}_{ss}^{-1}\mathbf{U}_j$ qualifies as an inner product for $\circ^n$. Assume the column space of $\mathbf{X}_s$, $\mathcal{C}(\mathbf{X}_s)$, is orthogonal to the column space of $\mathbf{U}_s$, $\mathcal{C}(\mathbf{U}_s)$, with respect to $\mathbf{V}_{ss}^{-1}$, denoted by $\mathcal{C}(\mathbf{X}_s) \perp_{\mathbf{V}_{ss}^{-1}} \mathcal{C}(\mathbf{U}_s)$. Since $\mathcal{C}(\mathbf{X}_s) \perp_{\mathbf{V}_{ss}^{-1}} \mathcal{C}(\mathbf{U}_s)$ then $\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s = \mathbf{0}$ or equivalently $\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s = \mathbf{0}$ (Harville, 1997, p. 257). In particular, $\mathcal{C}(\mathbf{X}_s) \perp_{\mathbf{V}_{ss}^{-1}} \mathcal{C}(\mathbf{U}_s)$ implies $\mathbf{X}_i'\mathbf{V}_{ss}^{-1}\mathbf{U}_j = 0$ or $\mathbf{U}_j'\mathbf{V}_{ss}^{-1}\mathbf{X}_i = 0$ for i = 1, 2, …, p and j = 1, 2, …, q. In addition, since $\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s = \mathbf{0}$ then $\mathbf{X}_s$ is orthogonal to $\mathbf{U}_s$, denoted by $\mathbf{X}_s \perp \mathbf{U}_s$ (Harville, 1997, p. 257). Since $\mathbf{X}_s \perp \mathbf{U}_s$, this implies the column vectors of $\mathbf{X}_s$ are independent of the column vectors of $\mathbf{U}_s$. Let $\tilde{\mathbf{X}}_s$ = $\mathbf{K}_1\mathbf{X}_s$ and notice $\tilde{\mathbf{U}}_s'\tilde{\mathbf{X}}_s = \mathbf{0}$. Thus, (6.9) becomes

$$\boldsymbol{\beta}'\mathbf{X}_\mathbf{s}'\mathbf{K}_1'\left(\mathbf{I}_\mathbf{s}-\mathbf{P}_{\tilde{\mathbf{u}}_\mathbf{s}}\right)\mathbf{K}_1\mathbf{X}_\mathbf{s}\boldsymbol{\beta} - 2\,\mathbf{y}_\mathbf{s}'\mathbf{K}_1'\left(\mathbf{I}_\mathbf{s}-\mathbf{P}_{\tilde{\mathbf{u}}_\mathbf{s}}\right)\mathbf{K}_1\mathbf{X}_\mathbf{s}\boldsymbol{\beta}$$
$$= \boldsymbol{\beta}'\tilde{\mathbf{X}}_\mathbf{s}'\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta} - 2\,\mathbf{y}_\mathbf{s}'\mathbf{K}_1'\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}. \qquad (6.10)$$

Define the projection matrix

$$\mathbf{P}_{\tilde{\mathbf{X}}_\mathbf{s}} = \tilde{\mathbf{X}}_\mathbf{s}\left(\tilde{\mathbf{X}}_\mathbf{s}'\tilde{\mathbf{X}}_\mathbf{s}\right)^{-1}\tilde{\mathbf{X}}_\mathbf{s}'.$$

Then we can rewrite (6.11) as

$$\boldsymbol{\beta}'\tilde{\mathbf{X}}_\mathbf{s}'\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta} - 2\,\mathbf{y}_\mathbf{s}'\mathbf{K}_1'\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}$$
$$= \left(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}\right)'\left(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}\right) - 2\left(\mathbf{P}_{\tilde{\mathbf{X}}_\mathbf{s}}\mathbf{K}_1\mathbf{y}_\mathbf{s}\right)'\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}.$$

Since $\mathbf{C}_2 \equiv \mathbf{P}_{\tilde{\mathbf{X}}_\mathbf{s}}\mathbf{K}_1\mathbf{y}_\mathbf{s}$ is a constant with respect to $\boldsymbol{\beta}$, we have

$$\left(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}\right)'\left(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}\right) - 2\,\mathbf{C}_2'\left(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}\right) + \mathbf{C}_2'\mathbf{C}_2 - \mathbf{C}_2'\mathbf{C}_2.$$

Thus, the quadratic in the exponential of the distribution of $\pi(\boldsymbol{\beta}|\mathbf{y}_\mathbf{s})$ is

$$\mathbf{R}_\mathbf{s} \equiv \left(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}\right)'\left(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}\right) - 2\,\mathbf{C}_2'\left(\tilde{\mathbf{X}}_\mathbf{s}\boldsymbol{\beta}\right) + \mathbf{C}_2'\mathbf{C}_2$$
$$= \left(\boldsymbol{\beta} - \left(\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^{-1}\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{y}_\mathbf{s}\right)'\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}$$
$$\left(\boldsymbol{\beta} - \left(\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^{-1}\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{y}_\mathbf{s}\right).$$

This suggests that the marginal distribution $\pi(\boldsymbol{\beta}|\mathbf{y}_\mathbf{s})$ is

$$\pi(\boldsymbol{\beta}|\mathbf{y}_\mathbf{s})$$
$$= N\left[\boldsymbol{\beta}\,\middle|\left(\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^{-1}\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{y}_\mathbf{s}, \left(\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^{-1}\right]$$
$$\equiv N\left[\boldsymbol{\beta}\,\middle|\,\mu_{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}\right]. \qquad (6.11)$$

7. The Bayes Estimator

In this section we combine the results of Sections 4 and 5 and derive our Bayesian estimator of the population total.

While obtaining the marginal distribution $\pi(\boldsymbol{\beta}|\mathbf{y}_\mathbf{s})$ in the last section we assumed $\mathcal{C}(\mathbf{X}_\mathbf{s}) \perp_{\mathbf{V}_\mathbf{ss}^{-1}} \mathcal{C}(\mathbf{U}_\mathbf{s})$.

Consequently, (6.6) becomes

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta},\mathbf{y}_\mathbf{s}) =$$
$$N\left[\boldsymbol{\gamma}\,\middle|(\mathbf{U}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{U}_\mathbf{s})^{-1}\mathbf{U}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}(\mathbf{y}_\mathbf{s}-\mathbf{X}_\mathbf{s}\boldsymbol{\beta}), \left(\mathbf{U}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{U}_\mathbf{s}\right)^{-1}\right]$$
$$= N\left[\boldsymbol{\gamma}\,\middle|(\mathbf{U}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{U}_\mathbf{s})^{-1}\mathbf{U}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{y}_\mathbf{s}, \left(\mathbf{U}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{U}_\mathbf{s}\right)^{-1}\right]$$
$$= \pi(\boldsymbol{\gamma}\,|\,\mathbf{y}_\mathbf{s}) \qquad (7.1)$$

Substituting equations (5.1), (7.1) and (6.11) into formula (4.2) we obtain

$$\pi(\mathbf{y}_\mathbf{r}|\mathbf{y}_\mathbf{s}) =$$
$$\int_{\mathbb{R}^p}\int_{\mathbb{R}^q} f(\mathbf{y}_\mathbf{r}\,|\,\mathbf{y}_\mathbf{s},\boldsymbol{\gamma},\boldsymbol{\beta})\pi(\boldsymbol{\gamma}\,|\,\mathbf{y}_\mathbf{s})\pi(\boldsymbol{\beta}\,|\,\mathbf{y}_\mathbf{s})d\boldsymbol{\gamma}d\boldsymbol{\beta}$$
$$= \int_{\mathbb{R}^p}\int_{\mathbb{R}^q} N\Big(\mathbf{y}_\mathbf{r}\,\Big|(\mathbf{X}_\mathbf{r}\boldsymbol{\beta}+\mathbf{U}_\mathbf{r}\boldsymbol{\gamma})$$
$$+\mathbf{V}_\mathbf{rs}\mathbf{V}_\mathbf{ss}^{-1}\left(\mathbf{y}_\mathbf{s}-(\mathbf{X}_\mathbf{s}\boldsymbol{\beta}+\mathbf{U}_\mathbf{s}\boldsymbol{\gamma})\right), \mathbf{V}_\mathbf{rr}-\mathbf{V}_\mathbf{rs}\mathbf{V}_\mathbf{ss}^{-1}\mathbf{V}_\mathbf{sr}\Big)$$
$$\times N\left[\boldsymbol{\gamma}\,\middle|\left(\mathbf{U}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{U}_\mathbf{s}\right)^{-1}\mathbf{U}_s'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{y}_\mathbf{s}, \left(\mathbf{U}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{U}_\mathbf{ss}\right)^{-1}\right]$$
$$\times N\left[\boldsymbol{\beta}\,\middle|\left(\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^{-1}\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{y}_\mathbf{s}, \left(\mathbf{X}_\mathbf{s}'\mathbf{V}_\mathbf{ss}^{-1}\mathbf{X}_\mathbf{s}\right)^{-1}\right]$$
$$d\boldsymbol{\gamma}d\boldsymbol{\beta}. \qquad (7.2)$$

To obtain the posterior predictive mean we manipulate (7.2) to obtain the nested conditional means

$$E(\mathbf{y}_\mathbf{r}|\mathbf{y}_\mathbf{s}) = E_{\boldsymbol{\beta}}\left\{E_{\boldsymbol{\gamma}}\left[E_{\mathbf{y}_\mathbf{r}}\left(\mathbf{y}_\mathbf{r}\,|\mathbf{y}_\mathbf{s},\boldsymbol{\gamma},\boldsymbol{\beta}\right)|\,\mathbf{y}_\mathbf{s}\right]|\mathbf{y}_\mathbf{s}\right\}$$
$$= \left(\mathbf{X}_\mathbf{r}\mu_{\boldsymbol{\beta}}+\mathbf{U}_\mathbf{r}\mu_{\boldsymbol{\gamma}}\right)+\mathbf{V}_\mathbf{rs}\mathbf{V}_\mathbf{ss}^{-1}\left(\mathbf{y}_\mathbf{s}-\left(\mathbf{X}_\mathbf{s}\mu_{\boldsymbol{\beta}}+\mathbf{U}_\mathbf{s}\mu_{\boldsymbol{\gamma}}\right)\right).$$

Thus, our estimate of the population total under squared error loss is

$$\hat{T}_\mathrm{B} = \mathbf{1}_\mathbf{s}'\mathbf{y}_\mathbf{s} + \mathbf{1}_\mathbf{r}'E(\mathbf{y}_\mathbf{r}\,|\,\mathbf{y}_\mathbf{s})$$
$$= \mathbf{1}_\mathbf{s}'\mathbf{y}_\mathbf{s} + \mathbf{1}_\mathbf{r}'\Big\{\left(\mathbf{X}_\mathbf{r}\mu_{\boldsymbol{\beta}}+\mathbf{U}_\mathbf{r}\mu_{\boldsymbol{\gamma}}\right)$$
$$+\mathbf{V}_\mathbf{rs}\mathbf{V}_\mathbf{ss}^{-1}\left(\mathbf{y}_\mathbf{s}-\left(\mathbf{X}_\mathbf{s}\mu_{\boldsymbol{\beta}}+\mathbf{U}_\mathbf{s}\mu_{\boldsymbol{\gamma}}\right)\right)\Big\}. \qquad (7.3)$$

If we assume $\mathbf{V} = \boldsymbol{\pi}(\mathbf{I}-\boldsymbol{\pi})^{-1}\boldsymbol{\pi}$, $\mathbf{U} = \boldsymbol{\pi}\mathbf{1}$ then (7.3) can be shown to have the form of the general regression estimator (2.1) with an adjusted error term:

$$\hat{T}_\mathrm{B} = \mathbf{1}'\mathbf{X}\mu_{\boldsymbol{\beta}} + \mathbf{1}_\mathbf{s}'\boldsymbol{\pi}_\mathbf{s}^{-1}\left(\mathbf{y}_\mathbf{s}-\boldsymbol{\pi}_\mathbf{s}\mathbf{X}_\mathbf{s}\mu_{\boldsymbol{\beta}}\right).$$

Bibliography

Geisser, S. (1993), *Predictive Inference: An Introduction*, London: Chapman and Hall.

Ghosh, M. and Meeden, G. (1997), *Bayesian Methods for Finite Population Sampling*, London: Chapman and Hall.

Harville, D. A. (1997), *Matrix Algebra from A Statistician's Perspective*, New York: Springer-Verlag.

Nieuwenbroek, N. J. and Renssen R. H. (1997), "Aligning estimates for common variables in two or more sample surveys," *Journal of the American Statistical Association*, 92, 368-374.

Pfeffermann, D. and Royall, R. M. (1982), "Balanced samples and robust Bayesian inference in finite population sampling," *Biometrika*, 69, 401-409.

Sarndal, C. E., Swensson, B., and Wretman, J. (1992), *Model assisted Survey Sampling*, New Yok: Springer-Verlag.