A Bayes And Empirical Bayes Prediction For a finite Population Total Using Auxiliary Information

Mark S. Hamner[a], John W. Seaman, Jr.[b], Dean M. Young[b]
[a]*Department of Mathematics and Computer Science, Texas Woman's University, Denton, Texas*
[b]*Institute of Statistics, Baylor University, Waco, Texas*

Keywords: Superpopulation model; empirical Bayes estimator; general regression estimator; partitioned matrices;

1. Introduction

Let $P = \{1, 2,\ldots, N\}$ represent the set of labels of the units of a finite population (N known) and y_i represent a fixed value of interest for each unit $i \in P$. The focus of this paper is on the prediction of a finite population total $T = \sum_i y_i$ by sampling n units from a population of size N units. If s represents a sample of size n selected from $\boldsymbol{P}$ and r represent the remaining $N - n$ non-sampled units from P, then the population total can be written as $T = \sum_s y_i + \sum_r y_i$. Define $\mathbf{1}' = [\mathbf{1}'_s \mid \mathbf{1}'_r]$ and $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_s \\ \mathbf{y}_r \end{bmatrix}$ where $\mathbf{1}'_s$ is an $1 \times n$ vector of ones, $\mathbf{1}'_r$ is a $1 \times (N-n)$ vector of ones, $\mathbf{y}_s$ is an $n \times 1$ vector of sampled units, and $\mathbf{y}_r$ is an $(N-n) \times 1$ vector of the non-sampled units. Then we may write

$$T = \mathbf{1}'_s\,\mathbf{y}_s + \mathbf{1}'_r\,\mathbf{y}_r. \qquad (1.1)$$

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 that will be borrowed.

Classical theory models the data collection procedure with a sampling design, a probability function defined on the sample space, $\mathbf{S}$, 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.

Bayesian interpretations of classical designed-based estimators such as the Horvitz-Thompson estimator and the well-known ratio-estimator have been established in the literature. For example, the classical ratio estimator has been obtained through various Bayesian superpopulation models (see, for example, Ghosh and Meeden, 1997, Section 3.2.1 and Royal and Pfefferman, 1982, p. 402). The Horvitz-Thompson estimator has also been shown to have an empirical Bayesian analog; see Ghosh and Meeden (1997, Section 4.2) and a fully Bayesian analog; see Ghosh and Meeden (1997, Section 5.1). Another popular estimator for finite population prediction is the general regression estimator proposed by Cassel, Sarndal and Wretman (1976), Sarndal (1980), and Sarndal, Swensson, and Wretman, (1997). The general regression estimator is the Horvitz-Thompson estimator plus an adjustment term.

In this paper we define a projection matrix by constructing a multivariate error structure which allows us to develop a fully Bayesian estimator of the population total T. As a special case of our estimator we obtain a Bayesian interpretation of the general regression estimator. Since our estimator is a function of the mean of the posterior predictive distribution, we can find the posterior standard deviation to our fully Bayesian estimator. Consequently, we are able to construct interval estimates that have a strict probabilistic interpretation without reference to repeating sampling.

In Section 2 we introduce the general regression estimator and then derive it in matrix form. In Section 3 we introduce a superpopulation model and obtain an empirical Bayes estimator of the population total. We also demonstrate that a special case of our empirical Bayes estimator is the general regression estimator. In Section 4 we establish the Bayesian model and then provide the general form of the integrand that produces the predictive distribution. In Sections 5 and 6 we derive the distributions in the integrand of the predictive distribution. In Section 7 we combine the results of Sections 4 through 7 and derive our Bayesian estimator of the population total. We present the posterior standard deviation for the fully Bayesian estimator in Section 8 and end the paper with a brief discussion in Section 9.

2. General Bayesian Regression Estimator

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

As a means to possibly improve the basic Horvitz-Thompson estimator using auxiliary information, Sarndal et al. (1997) employ classical sampling design theory, using inclusion probabilities, and the regression model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \qquad (2.1)$$

where $\mathbf{X}$ is the model matrix, β is the unknown coefficient vector, and $\mathbf{e} \sim N(\mathbf{0,V})$. Equation (2.1) is used, however, only as a means to obtain an estimate of β. Hence, unbiasedness and variance expressions are derived under the sampling design. In short, Sarndal et al. (1997) do not assume that the regression model (2.1) generates the sample. Thus, the *general regression estimator* (GRE) derived in Sarndal et al. (1997) is *model assisted* but not model dependent. Sarndal et al. (p.225, 1997) define the *GRE* as

$$\hat{T}_{GRE} = \sum_{k=1}^{n} \frac{y_k}{\pi_k} + \sum_{j=1}^{p} \hat{\beta}_j \left(\sum_{k=1}^{N} X_{jk} - \sum_{k=1}^{n} \frac{X_{jk}}{\pi_k} \right), \qquad (2.2)$$

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 Horvitz-Thompson estimator plus an adjustment term. Using a regression model to estimate $\beta \equiv (\beta_1, \ldots, \beta_p)'$, Sardnal et al. (1997, p. 228) suggest the estimator

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

where $\mathbf{X}_k$ is a $p \times 1$ vector of known auxiliary information and σ_k^2 denotes the kth diagonal element of the variance matrix $\mathbf{V}$ in (2.1). Under a simple random sampling design in which $\boldsymbol{\pi} = \text{Diag}\left(\frac{N}{n}\right)$, $\hat{\beta}$ is the generalized least squares estimator $\hat{\beta}_s \equiv \left(\mathbf{X}_s' \Sigma^{-1} \mathbf{X}_s\right)^{-1} \mathbf{X}_s' \Sigma^{-1} \mathbf{y}_s$, where $\Sigma \in \ ^{>}_{p}$, the set of all $p \times p$ positive definite real-valued matrices. Now, the *GRE* (2.2) can be rewritten as

$$\begin{aligned} \hat{T}_{GRE} &= \mathbf{1}_s' \boldsymbol{\pi}_s^{-1} \mathbf{y}_s + \left(\mathbf{1}'\mathbf{X} - \mathbf{1}_s' \boldsymbol{\pi}_s^{-1} \mathbf{X}_s'\right) \hat{\beta}_s \\ &= \mathbf{1}'\mathbf{X}\hat{\beta}_s + \mathbf{1}_s' \boldsymbol{\pi}_s^{-1} \left(\mathbf{y}_s - \mathbf{X}_s' \hat{\beta}_s\right). \qquad (2.3) \end{aligned}$$

A similar matrix representation of the general regression estimator is presented by Mukhopadhyay (1993). Assuming $\Sigma = \sigma^2 \mathbf{I}$, we can express (2.3) as

$$\hat{T}_{\text{GRE}} = \mathbf{1}'\mathbf{X}\hat{\beta}_s + \frac{N}{n} \mathbf{1}_s' \mathbf{P}_{\mathbf{X}_s}^{\perp} \mathbf{y}_s,$$

where $\hat{\beta}_s = \left(\mathbf{X}_s' \mathbf{X}_s\right)^{-1} \mathbf{X}_s \mathbf{y}_s$, $\mathbf{P}_{\mathbf{X}_s} = \mathbf{X}_s \mathbf{X}_s^{+}$, is the orthogonal projection matrix onto $X(\mathbf{X}_s)$, the column space of $\mathbf{X}_s$, $\mathbf{X}_s^{+}$ is the Moore-Penrose inverse of $\mathbf{X}_s$, and $\mathbf{P}_{\mathbf{X}_s}^{\perp}$ is the orthogonal projection matrix onto the orthogonal complement of $X(\mathbf{X}_s)$.

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 (1.1). Our empirical Bayes estimator of the population total requires derivation of the mean of the posterior predictive distribution $E(\mathbf{y}_r|\mathbf{y}_S)$. Finally, we show that a special case of our empirical Bayes estimator is the general regression estimator.

Royall and Pfeffermann (1982) focus attention on necessary assumptions needed for robustness of their statistical procedures for predicting the population total (3.1) given $\mathbf{y}_S$. They consider their procedure robust if the posterior probability distribution of T is not greatly affected if the model is taken to be

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{U}\gamma + \mathbf{e} \qquad (3.1)$$

instead of (2.1), where $\mathbf{y}$, $\mathbf{X}$, β and $\mathbf{e}$, are defined as in (2.1) and $\mathbf{U}$ contains additional regressors with a fixed coefficient vector γ. Renssen and Niewenbroek (1997) consider estimating the population total T by using two or more surveys to obtain *common variables*, $\mathbf{U}$, for use in (3.1). They assume these additional regressors are observed in both surveys where the corresponding population totals are unknown. The variables in $\mathbf{U}$ are combined with the auxiliary variables in $\mathbf{X}$, which have known population totals. They then use these common variables as a tool to improve the estimate of the population total with what they call an adjusted general regression estimator.

Consider the superpopulation model (3.1) where $\mathbf{X} = \begin{bmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{bmatrix} \in \ _{N\times p}$ such that rank$(\mathbf{X}) = p$, with $\mathbf{X}_s \in \ _{n\times p}$ and $\mathbf{X}_r \in \ _{(N-n)\times p}$. Furthermore, assume $\mathbf{U} = \begin{bmatrix} \mathbf{U}_s \\ \mathbf{U}_r \end{bmatrix} \in \ _{N\times q}$ where rank$(\mathbf{U}) = q$ such that $\mathbf{U}_s \in \ _{N\times q}$ and $\mathbf{U}_r \in \ _{(N-n)\times q}$. Finally, let $\gamma \in \ _{q\times 1}$ and let $\mathbf{e} \sim N(\mathbf{0},\mathbf{V})$ with $\mathbf{V}$ known, where

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

To estimate the total (1.1), we utilize $E(\mathbf{y}_r \mid \mathbf{y}_s)$ by regressing $\mathbf{y}_s$ on $\mathbf{X}_s$, resulting in an empirical Bayes estimator of $\boldsymbol{\beta}$.

Thus, consider the regression model $\mathbf{y}_s = \mathbf{X}_s\boldsymbol{\beta} + \mathbf{e}$ with $\mathbf{e} \sim \mathrm{N}(\mathbf{0},\mathbf{V}_{ss})$, where $\mathbf{V}_{ss} \in {}^{>}_{p}$ is known. For an improper prior $p(\boldsymbol{\beta}) \propto$ constant, one can show that the posterior distribution of $\boldsymbol{\beta}$ is

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

Substituting into (3.1) we have

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}}_s + \mathbf{U}\boldsymbol{\gamma} + \mathbf{e}\ . \qquad (3.2)$$

Using the improper prior $p(\boldsymbol{\gamma}) \propto$ constant, the posterior predictive distribution has the form

$$p(\mathbf{y}_r|\mathbf{y}_s) = \int_q \mathrm{N}\left(\mathbf{y}_r \middle| \mathbf{X}_r\hat{\boldsymbol{\beta}}_s + \mathbf{U}_r\boldsymbol{\gamma}, \mathbf{V}_{rr}\right) \times \mathrm{N}\left(\boldsymbol{\gamma} \middle| (\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{e}_s, \left(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)^{-1}\right) d\boldsymbol{\gamma}, \qquad (3.3)$$

where $\mathbf{e}_s = (\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s)$. Using (3.4), we have

$$\begin{aligned} E(\mathbf{y}_r \mid \mathbf{y}_s) &= \int_q \int_{N-n} \mathbf{y}_r \mathrm{N}\left(\mathbf{y}_r \middle| \mathbf{X}_r\hat{\boldsymbol{\beta}}_s + \mathbf{U}_r\boldsymbol{\gamma}, \mathbf{V}_{rr}\right) \\ &\quad \times \mathrm{N}\left(\boldsymbol{\gamma} \middle| (\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{e}_s, \left(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)^{-1}\right) d\mathbf{y}_r d\boldsymbol{\gamma} \\ &= E_{\boldsymbol{\gamma}}\left[E_{\mathbf{y}_r}\left(\mathbf{y}_r \middle| \hat{\boldsymbol{\beta}}_s, \mathbf{U}_r, \boldsymbol{\gamma}, \mathbf{V}_{rr}\right) \middle| \hat{\boldsymbol{\beta}}_s, \mathbf{y}_s, \mathbf{U}_s, \mathbf{V}_{ss}\right] \\ &= \mathbf{X}_r\hat{\boldsymbol{\beta}}_s + \mathbf{U}_r(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right), \end{aligned}$$

where $E_{\mathbf{y}_r}(\cdot)$ and $E_{\boldsymbol{\gamma}}(\cdot)$ denote expectations with respect to the distributions of $\mathbf{y}_r$ and $\boldsymbol{\gamma}$, respectively. Thus, our empirical Bayes estimator of the population total is

$$\begin{aligned} \hat{T}_{EB} &= \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r' E(\mathbf{y}_r \mid \mathbf{y}_s) \\ &= \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r'\left[\mathbf{X}_r\hat{\boldsymbol{\beta}}_s + \mathbf{U}_r(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1} \times \mathbf{U}_s'\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right)\right]. \end{aligned} \qquad (3.4)$$

We now show that a special case of $\hat{T}_{EB}$ is the general regression estimator $\hat{T}_{GRE}$. Define $\boldsymbol{\pi}$ to be an $N \times N$ matrix whose diagonal elements, $\pi_{ii} \equiv \pi_i$, i = 1,2,…N, are inclusion probabilities and the off-diagonal elements are zero so that

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

and $\mathrm{tr}(\boldsymbol{\pi}) = n$. Next, assume $\mathbf{V} = \boldsymbol{\pi}(\mathbf{I} - \boldsymbol{\pi})^{-1}\boldsymbol{\pi}$ and $\mathbf{U} = \boldsymbol{\pi}\mathbf{1}$. Note that the covariance structure $\mathbf{V}$ is the matrix representation of the geometric distribution. Similar assumptions are presented by Ghosh and Meeden (1997, Section 4.2). Then, we have that

$$\begin{aligned} E(\boldsymbol{\gamma} \mid \mathbf{y}_s, \hat{\boldsymbol{\beta}}_s, \mathbf{U}_s, \mathbf{V}_{ss}) &= (\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right) \\ &= \left[\mathbf{1}_r'\boldsymbol{\pi}_r\mathbf{1}_r\right]^{-1}\mathbf{1}_s'\left(\mathbf{I}_s - \boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right). \end{aligned} \qquad (3.5)$$

Now $\mathrm{tr}(\boldsymbol{\pi}) = n$ implies $\mathbf{1}_r'\boldsymbol{\pi}_r\mathbf{1}_r = n - \mathbf{1}_s'\boldsymbol{\pi}_s\mathbf{1}_s$, so, from (3.5) our empirical Bayes estimator becomes

$$\begin{aligned} \hat{T}_{\mathrm{EB}} &= \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r'\left[\mathbf{X}_r\hat{\boldsymbol{\beta}}_s + \mathbf{U}_r(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{e}_s\right] \\ &= \mathbf{1}_s'\left[\mathbf{X}_s\hat{\boldsymbol{\beta}}_s + \mathbf{e}_s\right] + \mathbf{1}_r'\mathbf{X}_r\hat{\boldsymbol{\beta}}_s + \\ &\quad \mathbf{1}_r'\boldsymbol{\pi}_r\mathbf{1}_r\left[\mathbf{1}_r'\boldsymbol{\pi}_r\mathbf{1}_r\right]^{-1}\mathbf{1}_s'\left(\mathbf{I}_s - \boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\mathbf{e}_s \\ &= \mathbf{1}'\mathbf{X}\hat{\boldsymbol{\beta}}_s + \mathbf{1}_s'\boldsymbol{\pi}_s^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol{\beta}}_s\right) \\ &= \hat{T}_{\mathrm{GRE}}. \end{aligned}$$

4. The Superpopulation Model

In this section we develop a Bayesian model and then provide the general form of the integrand that will produce the predictive distribution.

Consider the superpopulation model

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

where $\mathbf{X} \in {}_{N\times p}$ with rank($\mathbf{X}$) = p, $\mathbf{U} \in {}_{N\times q}$ with rank($\mathbf{U}$) = q, $\boldsymbol{\beta} \in {}_{p\times 1}$, $\boldsymbol{\gamma} \in {}_{q\times 1}$, and $\mathbf{e} \sim \mathrm{N}(\mathbf{0},\mathbf{V})$. Assume $\mathbf{V} \in {}^{>}_{N}$ can 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 $\mathbf{V}_{ss} \in {}^{>}_{n}$, $\mathbf{V}_{rr} \in {}^{>}_{(N-n)}$, and $\mathbf{V}_{sr} \in {}_{(N-n)\times n}$ such that $\mathbf{V}_{rs} = \mathbf{V}_{sr}'$ and $(\mathbf{V}_{rr} - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{V}_{sr}) \in {}^{>}_{(N-n)}$.

Assume $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ each have improper uniform prior distributions. Using the super population model

(4.1), we obtain $E(\mathbf{y}_r \mid \mathbf{y}_s)$ to estimate $T_r = \mathbf{1}_r'\mathbf{y}_r$ in (1.1).

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

$p(\mathbf{y}_r|\mathbf{y}_s)=$

$$\frac{1}{f(\mathbf{y}_r)}\int_p\int_q f(\mathbf{y}_r|\mathbf{y}_s,\beta,\gamma)\,f(\mathbf{y}_s|\beta,\gamma)\,p(\beta,\gamma)\,d\gamma d\beta$$

$$=\int_p\int_q f(\mathbf{y}_r|\mathbf{y}_s,\gamma,\beta)\,p(\gamma|\beta,\mathbf{y}_s)\,p(\beta|\mathbf{y}_s)\,d\gamma d\beta, \quad (4.2)$$

where f denotes the appropriate density. Notice that, if $\mathbf{y}_r$ is independent of $\mathbf{y}_s$, then

$p(\mathbf{y}_r|\mathbf{y}_s) =$

$$\int_p\int_q f(\mathbf{y}_r|\gamma,\beta)\,p(\gamma|\beta,\mathbf{y}_s)\,p(\beta|\mathbf{y}_s)\,d\gamma d\beta.$$

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. To derive $f(\mathbf{y}_r \mid \mathbf{y}_s, \beta, \gamma)$, define $\mathbf{z}_r \equiv \mathbf{y}_r + \mathbf{M}\,\mathbf{y}_s$ and $\mathbf{z}_s \equiv \mathbf{y}_s$, where $\mathbf{M}$ is chosen so that $\mathbf{z}_r$ is uncorrelated with $\mathbf{z}_s$. That is, $\mathbf{M}$ must satisfy $Cov(\mathbf{z}_r, \mathbf{z}_s) = \mathbf{V}_{rs} + \mathbf{M}\mathbf{V}_{ss} = \mathbf{0}$, which yields $\mathbf{z}_r = \mathbf{y}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{y}_s$. Note that

$$\mathbf{z} = \begin{bmatrix} \mathbf{y}_s \\ \hline \mathbf{y}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{y}_{sr} \end{bmatrix}$$

and, therefore, $\mathbf{z} \sim N(\mu, \Sigma)$, where

$$\mu = \begin{bmatrix} \mu_{\mathbf{z}_s} \\ \hline \mu_{\mathbf{z}_r} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_s\beta + \mathbf{U}_s\gamma \\ \hline \left(\mathbf{X}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)\beta + \left(\mathbf{U}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)\gamma \end{bmatrix}$$

and

$$\Sigma \equiv \begin{bmatrix} \Sigma_{\mathbf{z}_s} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{z}_r} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{ss} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{rr} - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{V}_{sr} \end{bmatrix}.$$

Because $\mathbf{z}_s$ and $\mathbf{z}_r$ are uncorrelated and jointly multivariate normally distributed, they are independent and, therefore,

$$f(\mathbf{z}_s, \mathbf{z}_r) = N\left(\mathbf{z}_s|\mu_{\mathbf{z}_s}, \Sigma_{\mathbf{z}_s}\right) N\left(\mathbf{z}_r|\mu_{\mathbf{z}_r}, \Sigma_{\mathbf{z}_r}\right)$$
$$= N\left(\mathbf{z}_s|\mathbf{X}_s\beta + \mathbf{U}_s\gamma, \mathbf{V}_{ss}\right)$$
$$\times N\left(\mathbf{z}_r\left|\left(\mathbf{X}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)\beta + \left(\mathbf{U}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)\gamma,\right.\right.$$
$$\left.\times \mathbf{V}_{rr} - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{V}_{sr}\right). \quad (5.1)$$

Substituting $\mathbf{y}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{y}_s$ for $\mathbf{z}_r$ in (5.1) and noting that the Jacobian is one yields the quadratic term

$$\left\{\mathbf{y}_r - \left[(\mathbf{X}_r\beta + \mathbf{U}_r\gamma) + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - (\mathbf{X}_s\beta + \mathbf{U}_s\gamma)\right)\right]\right\}'$$
$$\times\left(\mathbf{V}_{rr} - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{V}_{sr}\right)^{-1}$$
$$\times\left\{\mathbf{y}_r - \left[(\mathbf{X}_r\beta + \mathbf{U}_r\gamma) + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - (\mathbf{X}_s\beta + \mathbf{U}_s\gamma)\right)\right]\right\}.$$

Thus,

$f(\mathbf{y}_r \mid \mathbf{y}_s, \beta, \gamma) =$

$$N\left(\mathbf{y}_r\left|(\mathbf{X}_r\beta + \mathbf{U}_r\gamma) + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - (\mathbf{X}_s\beta + \mathbf{U}_s\gamma)\right),\right.\right.$$
$$\left.\times \mathbf{V}_{rr} - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{V}_{sr}\right). \quad (5.2)$$

Similarly, substituting $\mathbf{y}_s$ for $\mathbf{z}_s$ in (5.1) yields

$$N(\mathbf{z}_s \mid \mu_{\mathbf{z}_s}, \Sigma_{\mathbf{z}_s}) = N(\mathbf{y}_s \mid \mathbf{X}_s\beta + U_s\gamma, V_{ss})$$
$$= f(\mathbf{y}_s \mid \beta, \gamma).$$

6. Posterior Distributions for β and γ

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

$$p(\gamma, \beta \mid \mathbf{y}_s) = p(\gamma \mid \beta, \mathbf{y}_s)\, p(\beta \mid \mathbf{y}_s)$$
$$= N(\mathbf{y}_s \mid \mathbf{X}_s\beta + \mathbf{U}_s\gamma, \mathbf{V}_{ss}). \quad (6.1)$$

Thus, the quadratic terms in the exponents of $p(\gamma \mid \beta, \mathbf{y}_s)$ and $p(\beta \mid \mathbf{y}_s)$ contained in (6.1) are

$$\left[\mathbf{y}_s - (\mathbf{X}_s\beta + \mathbf{U}_s\gamma)\right]'\mathbf{V}_{ss}^{-1}\left[\mathbf{y}_s - (\mathbf{X}_s\beta + \mathbf{U}_s\gamma)\right]$$
$$= \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s - \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta - \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma$$
$$-\beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s - \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s + \beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta$$
$$+\beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma + \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta$$
$$+\gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma. \quad (6.2)$$

From (6.2) we have that the quadratic term in $p(\gamma \mid \beta, \mathbf{y}_s)$ is

$$\gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma - \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma - \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s$$
$$+\beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma + \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta$$
$$= \gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma - 2\left(\mathbf{y}_s' - \beta'\mathbf{X}_s'\right)\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma. \quad (6.3)$$

Next, note that $\mathbf{V}_{ss}^{-1}$ has full-rank factorization

$$\mathbf{V}_{ss}^{-1} = \mathbf{K}_1'\mathbf{K}_1, \tag{6.4}$$

where $\mathbf{K}_1 \in \ ^{>}_{n}$ and let $\tilde{\mathbf{U}}_s = \mathbf{K}_1\mathbf{U}_s$. Using (6.4), we rewrite (6.3) as

$$\begin{aligned}
&\gamma'\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma - 2(\mathbf{y}_s' - \beta'\mathbf{X}_s')\mathbf{V}_{ss}^{-1}\mathbf{U}_s\gamma \\
&= \gamma'\mathbf{U}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{U}_s\gamma - 2(\mathbf{y}_s' - \beta'\mathbf{X}_s')\mathbf{K}_1'\mathbf{K}_1\mathbf{U}_s\gamma \\
&= \gamma'\tilde{\mathbf{U}}_s'\tilde{\mathbf{U}}_s\gamma - 2(\mathbf{y}_s' - \beta'\mathbf{X}_s')\mathbf{K}_1'\tilde{\mathbf{U}}_s\gamma \\
&= (\tilde{\mathbf{U}}_s\gamma)'(\tilde{\mathbf{U}}_s\gamma) - 2\left[\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1(\mathbf{y}_s - \mathbf{X}_s\beta)\right]'\tilde{\mathbf{U}}_s\gamma. \qquad (6.5)
\end{aligned}$$

Define

$$\mathbf{C}_1 = \mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1(\mathbf{y}_s - \mathbf{X}_s\beta). \tag{6.6}$$

Using (6.4) and (6.6) we can express the quadratic term in $p(\gamma \mid \beta, \mathbf{y}_s)$ as

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

Thus,

$$\begin{aligned}
p(\gamma|\beta,\mathbf{y}_r) &= \mathrm{N}\Big[\gamma\Big|(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}_s'\mathbf{V}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\beta), \\
&\qquad \times\left(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)^{-1}\Big]. \\
&\equiv \mathrm{N}\left[\gamma\middle|\mu_\gamma, \mathbf{V}_\gamma\right]. \qquad (6.7)
\end{aligned}$$

We next derive the quadratic term of the marginal distribution $p(\beta \mid \mathbf{y}_r)$. From (6.6) we have

$$\begin{aligned}
\mathbf{C}_1'\mathbf{C}_1 &= \left[\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1(\mathbf{y}_s - \mathbf{X}_s\beta)\right]'\left[\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1(\mathbf{y}_s - \mathbf{X}_s\beta)\right] \\
&= \beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\beta - 2\mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\beta \\
&\quad + \mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}\mathbf{y}_s. \qquad (6.8)
\end{aligned}$$

Using (6.3), we utilize the remaining term containing β in (6.1) to obtain the quadratic term in $p(\gamma \mid \beta, \mathbf{y}_s)$, which is

$$\begin{aligned}
&\beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta - \mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta - \beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s \\
&= \beta'\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta - 2\mathbf{y}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\beta. \\
&= \beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta - 2\mathbf{y}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta. \qquad (6.9)
\end{aligned}$$

Thus, from (6.8) and (6.9), the quadratic term in $\pi(\beta \mid \mathbf{y}_s)$ is

$$\begin{aligned}
&\beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta - \beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\beta \\
&\quad - 2\mathbf{y}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta + 2\mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\beta \\
&= \beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\beta - 2\mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\beta. \qquad (6.10)
\end{aligned}$$

Define $\dot{\mathbf{X}}_s = \mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s$. Because $\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}$ is a symmetric idempotent matrix, one can show that (6.10) becomes

$$\begin{aligned}
&\beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\beta - 2\mathbf{y}_s'\mathbf{K}_1'\mathbf{K}_1\mathbf{X}_s\beta \\
&= \beta'\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\beta - 2\mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\beta \\
&= \beta'\dot{\mathbf{X}}_s'\dot{\mathbf{X}}_s\beta - 2\mathbf{y}_s'\mathbf{K}_1'\dot{\mathbf{X}}_s\beta \\
&= \beta'\dot{\mathbf{X}}_s'\dot{\mathbf{X}}_s\beta - 2\mathbf{y}_s'\mathbf{K}_1'\mathbf{P}_{\dot{\mathbf{X}}_s}\dot{\mathbf{X}}_s\beta \\
&= (\dot{\mathbf{X}}_s\beta)'(\dot{\mathbf{X}}_s\beta) - 2\left(\mathbf{P}_{\dot{\mathbf{X}}_s}\mathbf{K}_1\mathbf{y}_s\right)'(\dot{\mathbf{X}}_s\beta).
\end{aligned}$$

We assume that the sample size n is greater than the combined rank($\mathbf{X}$) = p and rank($\mathbf{U}$) = q and since $\mathbf{C}_2 \equiv \mathbf{P}_{\dot{\mathbf{X}}_s}\mathbf{K}_1\mathbf{y}_s$ is a constant with respect to β, the quadratic term in $p(\beta \mid \mathbf{y}_s)$ becomes

$$\begin{aligned}
&(\dot{\mathbf{X}}_s\beta)'(\dot{\mathbf{X}}_s\beta) - 2\mathbf{C}_2'(\dot{\mathbf{X}}_s\beta) + \mathbf{C}_2'\mathbf{C}_2 \\
&= (\dot{\mathbf{X}}_s\beta - \mathbf{C}_2)'(\dot{\mathbf{X}}_s\beta - \mathbf{C}_2) \\
&= \left(\dot{\mathbf{X}}_s\beta - \mathbf{P}_{\dot{\mathbf{X}}_s}\mathbf{K}_1\mathbf{y}_s\right)'\left(\dot{\mathbf{X}}_s\beta - \mathbf{P}_{\dot{\mathbf{X}}_s}\mathbf{K}_1\mathbf{y}_s\right) \\
&= \left(\beta - \dot{\mathbf{X}}_s^{+}\mathbf{K}_1\mathbf{y}_s\right)'\dot{\mathbf{X}}_s'\dot{\mathbf{X}}_s\left(\beta - \dot{\mathbf{X}}_s^{+}\mathbf{K}_1\mathbf{y}_s\right) \\
&= \left[\beta - \left(\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{y}_s\right]' \\
&\quad \times\left[\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\right]^{-1} \\
&\quad \times\left[\beta - \left(\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{y}_s\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
p(\beta \mid \mathbf{y}_s) &= \mathrm{N}\left[\beta\middle|\dot{\mathbf{X}}_s^{+}\mathbf{K}_1\mathbf{y}_s, (\dot{\mathbf{X}}_s'\dot{\mathbf{X}}_s)^{-1}\right] \\
&= \mathrm{N}\Big[\beta\Big|\left(\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{y}_s, \\
&\qquad \times\left(\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\right)^{-1}\Big] \\
&\equiv \mathrm{N}\left[\beta\middle|\mu_\beta, \mathbf{V}_\beta\right]. \qquad (6.11)
\end{aligned}$$

The variable $\tilde{\mathbf{X}}_s$ has an interesting interpretation. In particular, by letting $\tilde{\mathbf{X}}_s = \mathbf{K}_1\mathbf{X}_s$ and recalling that $\tilde{\mathbf{U}}_s = \mathbf{K}_1\mathbf{U}_s$, then $\dot{\mathbf{X}}_s = \left(\mathbf{P}^{\perp}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\right)$ can be viewed as an error matrix, say $\mathbf{E}$, for projecting $\tilde{\mathbf{X}}_s$ onto the column space of $\tilde{\mathbf{U}}_s$. In other words, $\dot{\mathbf{X}}_s$ shows how the Bayesian paradigm handles the alignment or combination of the auxiliary information contained in $\tilde{\mathbf{X}}_s$ and $\tilde{\mathbf{U}}_s$ by creating a variable of "left-over" or residual information from projecting $\tilde{\mathbf{X}}_s$ onto the column space of $\tilde{\mathbf{U}}_s$. Also, notice that $\dot{\mathbf{X}}'_s\dot{\mathbf{X}}_s = \left(\mathbf{X}'_s\mathbf{K}'_1\mathbf{P}^{\perp}_{\tilde{\mathbf{u}}_s}\right)\left(\mathbf{P}^{\perp}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\right)$ because $\mathbf{P}^{\perp}_{\tilde{\mathbf{u}}_s}$ is idempotent. Thus, the covariance structure for $\boldsymbol{\beta}$ is $\mathbf{V}_{\boldsymbol{\beta}} = \dot{\mathbf{X}}'_s\dot{\mathbf{X}}_s = \mathbf{E}'\mathbf{E}$ which represents a multivariate estimate of the covariance matrix.

If we assume $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are conditionally independent, then

$$\mathrm{X}(\mathbf{X}_s) \perp_{\mathbf{V}_{ss}^{-1}} \mathrm{X}(\mathbf{U}_s), \tag{6.12}$$

and $\mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{X}_s = \mathbf{0}$ (Harville, 1997, p.257). In particular, because of (6.12) it follows that $\tilde{\mathbf{U}}'_s\tilde{\mathbf{X}}_s = \mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{X}_s = \mathbf{0}$. Thus, (6.10) becomes

$$\boldsymbol{\beta}'\mathbf{X}'_s\mathbf{K}'_1\mathbf{P}^{\perp}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\boldsymbol{\beta} - 2\mathbf{y}'_s\mathbf{K}'_1\mathbf{P}^{\perp}_{\tilde{\mathbf{u}}_s}\mathbf{K}_1\mathbf{X}_s\boldsymbol{\beta}$$
$$= \boldsymbol{\beta}'\tilde{\mathbf{X}}'_s\mathbf{P}^{\perp}_{\tilde{\mathbf{u}}_s}\tilde{\mathbf{X}}_s\boldsymbol{\beta} - 2\mathbf{y}'_s\mathbf{K}'_1\mathbf{P}^{\perp}_{\tilde{\mathbf{u}}_s}\tilde{\mathbf{X}}_s\boldsymbol{\beta}$$
$$= \boldsymbol{\beta}'\tilde{\mathbf{X}}'_s\tilde{\mathbf{X}}_s\boldsymbol{\beta} - 2\mathbf{y}'_s\mathbf{K}'_1\tilde{\mathbf{X}}_s\boldsymbol{\beta}$$
$$= \left(\tilde{\mathbf{X}}_s\boldsymbol{\beta}\right)'\left(\tilde{\mathbf{X}}_s\boldsymbol{\beta}\right) - 2\left(\mathbf{P}_{\tilde{\mathbf{X}}_s}\mathbf{K}_1\mathbf{y}_s\right)'\tilde{\mathbf{X}}_s\boldsymbol{\beta}.$$

The quadratic term in $p(\boldsymbol{\beta} \mid \mathbf{y}_s)$ can now be expressed as

$$\left(\tilde{\mathbf{X}}_s\boldsymbol{\beta}\right)'\left(\tilde{\mathbf{X}}_s\boldsymbol{\beta}\right) - 2\mathbf{C}'_2\left(\tilde{\mathbf{X}}_s\boldsymbol{\beta}\right) + \mathbf{C}'_2\mathbf{C}_2$$
$$= \left(\tilde{\mathbf{X}}_s\boldsymbol{\beta} - \mathbf{C}_2\right)'\left(\tilde{\mathbf{X}}_s\boldsymbol{\beta} - \mathbf{C}_2\right)$$
$$= \left(\tilde{\mathbf{X}}_s\boldsymbol{\beta} - \mathbf{P}_{\tilde{\mathbf{X}}_s}\mathbf{K}_1\mathbf{y}_s\right)'\left(\tilde{\mathbf{X}}_s\boldsymbol{\beta} - \mathbf{P}_{\tilde{\mathbf{X}}_s}\mathbf{K}_1\mathbf{y}_s\right)$$
$$= \left(\boldsymbol{\beta} - \left(\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s\right)'$$
$$\times\left(\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)$$
$$\times\left(\boldsymbol{\beta} - \left(\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s\right).$$

Hence, (6.12) implies that

$$p(\boldsymbol{\beta}|\mathbf{y}_s) = \mathrm{N}\left[\boldsymbol{\beta}\,\middle|\,\left(\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s, \left(\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\right]$$
$$\equiv \mathrm{N}\left[\boldsymbol{\beta}\,\middle|\,\boldsymbol{\mu}_{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}\right].$$

Also, because of (6.12), (6.7) becomes

$$p(\boldsymbol{\gamma}|\boldsymbol{\beta},\mathbf{y}_s) = \mathrm{N}\left[\boldsymbol{\gamma}\,\middle|\,(\mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}'_s\mathbf{V}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta}), \left(\mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)^{-1}\right]$$
$$= \mathrm{N}\left[\boldsymbol{\gamma}\,\middle|\,(\mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s, \left(\mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)^{-1}\right]$$
$$= p(\boldsymbol{\gamma} \mid \mathbf{y}_s).$$

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. Substituting equations (5.2), (6.7) and (6.11) into (4.2) we obtain

$$p(\mathbf{y}_r \mid \mathbf{y}_s) = \int_p\int_q f(\mathbf{y}_r \mid \mathbf{y}_s, \boldsymbol{\gamma}, \boldsymbol{\beta})\, p(\boldsymbol{\gamma} \mid \mathbf{y}_s)\, p(\boldsymbol{\beta} \mid \mathbf{y}_s)\, d\boldsymbol{\gamma}\, d\boldsymbol{\beta}$$
$$= \int_p\int_q \mathrm{N}\Big(\mathbf{y}_r \,\Big|\, (\mathbf{X}_r\boldsymbol{\beta} + \mathbf{U}_r\boldsymbol{\gamma})$$
$$+\mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - (\mathbf{X}_s\boldsymbol{\beta} + \mathbf{U}_s\boldsymbol{\gamma})\right), \mathbf{V}_{rr} - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{V}_{sr}\Big)$$
$$\times \mathrm{N}\Big[\boldsymbol{\gamma}\,\Big|\,(\mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{U}_s)^{-1}\mathbf{U}'_s\mathbf{V}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta}),$$
$$\times\left(\mathbf{U}'_s\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)^{-1}\Big]$$
$$\times \mathrm{N}\left[\boldsymbol{\beta}\,\middle|\,(\dot{\mathbf{X}}'_s\dot{\mathbf{X}}_s)^{-1}\dot{\mathbf{X}}'_s\mathbf{K}_1\mathbf{y}_s, (\dot{\mathbf{X}}'_s\dot{\mathbf{X}}_s)^{-1}\right] d\gamma\, d\boldsymbol{\beta}. \tag{7.1}$$

The posterior predictive mean is, therefore,

$$E(\mathbf{y}_r \mid \mathbf{y}_s) = E_{\boldsymbol{\beta}}\left\{E_{\boldsymbol{\gamma}}\left[E_{\mathbf{y}_r}\left(\mathbf{y}_r \mid \mathbf{y}_s, \boldsymbol{\gamma}, \boldsymbol{\beta}\right) \mid \mathbf{y}_s\right] \mid \mathbf{y}_s\right\}$$
$$= \left(\mathbf{X}_r\boldsymbol{\mu}_{\boldsymbol{\beta}} + \mathbf{U}_r\boldsymbol{\mu}_{\boldsymbol{\gamma}}\right) + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \left(\mathbf{X}_s\boldsymbol{\mu}_{\boldsymbol{\beta}} + \mathbf{U}_s\boldsymbol{\mu}_{\boldsymbol{\gamma}}\right)\right). \tag{7.2}$$

Using equation (7.2), our Bayes estimator of the population total is

$$\hat{T}_{\mathrm{B}} = \mathbf{1}'_s\mathbf{y}_s + \mathbf{1}'_r E(\mathbf{y}_r \mid \mathbf{y}_s)$$
$$= \mathbf{1}'_s\mathbf{y}_s + \mathbf{1}'_r\left\{\left(\mathbf{X}_r\mu_{\boldsymbol{\beta}} + \mathbf{U}_r\mu_{\boldsymbol{\gamma}}\right) + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\right.$$

$$\times\left(\mathbf{y}_s-\left(\mathbf{X}_s\mu_{\boldsymbol\beta}+\mathbf{U}_s\mu_{\boldsymbol\gamma}\right)\right)\Big\}. \qquad (7.3)$$

If we assume $\mathbf{V} = \boldsymbol\pi(\mathbf{I}-\boldsymbol\pi)^{-1}\boldsymbol\pi$ and $\mathbf{U} = \boldsymbol\pi\mathbf{1}$, then we can show (7.3) has the exact form of the regression estimator (2.1)

$$\hat{T}_B = \mathbf{1}'\mathbf{X}\boldsymbol\mu_{\boldsymbol\beta} + \mathbf{1}_s'\boldsymbol\pi_s^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\boldsymbol\mu_{\boldsymbol\beta}\right),$$

where $\boldsymbol\mu_{\boldsymbol\beta} = \left\{\mathbf{X}_s'\mathbf{Q}\mathbf{X}_s\right\}^{-1}\mathbf{X}_s'\mathbf{Q}\mathbf{y}_s$ such that

$$\begin{aligned}\mathbf{Q} &= \mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\\ &= \boldsymbol\pi_s^{-1}\Big\{\left(\mathbf{I}_s-\boldsymbol\pi_s\right)-\left(\mathbf{I}_s-\boldsymbol\pi_s\right)\mathbf{1}_s\\ &\quad\times\left[\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol\pi_s\right)\mathbf{1}_s\right]^{-1}\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol\pi_s\right)\Big\}\boldsymbol\pi_s^{-1}. \qquad (7.4)\end{aligned}$$

Recall that Sardnal et al. (1997, p. 228) suggest estimating β with $\hat{\boldsymbol\beta}_s \equiv \left(\mathbf{X}_s'\boldsymbol\Sigma^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\boldsymbol\Sigma^{-1}\mathbf{y}_s$. Note that $\hat{\boldsymbol\beta}_s$ has the exact form as $\boldsymbol\mu_{\boldsymbol\beta}$ above, but in our Bayesian context, $\boldsymbol\Sigma^{-1}$ is defined by (7.4).

Under the more strict assumption where $\boldsymbol\beta$ and $\boldsymbol\gamma$ are independent, i.e. $X(\mathbf{X}_s) \perp_{\mathbf{V}_{ss}^{-1}} X(\mathbf{U}_s)$, equation (7.2) becomes

$$E(\mathbf{y}_r \mid \mathbf{y}_s) = \mathbf{X}_r\hat{\boldsymbol\beta} + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}\right)$$

where $\hat{\boldsymbol\beta} = \left(\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s$. In particular, our Bayes estimator (7.3) becomes the population total predictor found in Royall and Pfeffermann (1982, equation 1, p.402; that is

$$\hat{T}_{\mathrm{B}} = \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r'\left[\mathbf{X}_r\hat{\boldsymbol\beta} + \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\boldsymbol\beta}\right)\right].$$

The above estimator has the form of the best linear unbiased predictor (Royall, 1976a). Furthermore, when $\boldsymbol\beta$ and $\boldsymbol\gamma$ are independent, then the Bayes estimator for the population total under the assumption $\mathbf{V} = \boldsymbol\pi(\mathbf{I}-\boldsymbol\pi)^{-1}\boldsymbol\pi$ with $\mathbf{U} = \boldsymbol\pi\mathbf{1}$ is the general regression estimator (2.1) with an adjusted error term, that is

$$\hat{T}_B = \mathbf{1}'\mathbf{X}\hat{\boldsymbol\beta} + \mathbf{1}_s'\boldsymbol\pi_s^{-1}\left(\mathbf{y}_s - \boldsymbol\pi_s\mathbf{X}_s\hat{\boldsymbol\beta}\right),$$

where

$$\hat{\boldsymbol\beta} = \left(\mathbf{X}_s'\boldsymbol\pi_s^{-1}(\mathbf{I}_s-\boldsymbol\pi)\boldsymbol\pi_s^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\boldsymbol\pi_s^{-1}(\mathbf{I}_s-\boldsymbol\pi)\boldsymbol\pi_s^{-1}\mathbf{y}_s.$$

Notice that difference between the Bayesian estimator above to the general regression estimator (2.1) is the fraction $\boldsymbol\pi_s$, which is the inclusion probability multiplied to $\mathbf{X}_s\boldsymbol\mu_{\boldsymbol\beta}$. Note that $\boldsymbol\mu_{\boldsymbol\beta}$ is equal to the weighted least squares estimator $\hat{\boldsymbol\beta}_s$.

8. Dispersion of $\hat{T}_B$

In this section we derive the posterior standard deviation for the fully Bayesian estimator $\hat{T}_B$ given in (7.3). We have

$$\begin{aligned}Var(T) &= Var\left(\mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_r'\mathbf{y}_r \mid \mathbf{y}_s\right)\\ &= \mathbf{1}_r'\,Var(\mathbf{y}_r \mid \mathbf{y}_s)\mathbf{1}_r\\ &= \mathbf{1}_r'\Big[\left(\mathbf{V}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{V}_{sr}\right)\\ &+ \left(\mathbf{U}_r - \mathbf{V}_{rs}\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)\left(\mathbf{U}_s'\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)^{-1}\left(\mathbf{U}_r - \mathbf{V}_{sr}\mathbf{V}_{ss}^{-1}\mathbf{U}_s\right)'\\ &+ \left(\mathbf{X}_r - \mathbf{V}_{sr}\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)\left(\mathbf{X}_s'\mathbf{K}_1'\mathbf{P}_{\tilde{\mathbf{u}}_s}^{\perp}\mathbf{K}_1\mathbf{X}_s\right)^{-1}\\ &\times\left(\mathbf{X}_r - \mathbf{V}_{sr}\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)'\Big]\mathbf{1}_r.\end{aligned}$$

9. Discussion

In this paper we have developed an empirical Bayes estimator, $\hat{T}_{EB}$, and a fully Bayesian estimator, $\hat{T}_B$, that yield the classical general regression estimator $\hat{T}_{\mathrm{GRE}}$ as a special case. The fully Bayesian procedure recognizes the uncertainty of all prior parameters with "non-informative" or improper priors whereas the empirical Bayesian procedure estimates one of the parameters with a classical approach. The fully Bayesian procedure explicitly shows how the estimate of the parameter $\boldsymbol\beta$ is weighted with the inclusion probabilities that are used extensively in classical designed-based finite population prediction (see equation (7.4)). One of the advantages of using the fully Bayesian procedure for inference on the population total is the ability to clearly measure the posterior variation on the unknown quantity of interest *T*. Consequently, we can obtain the posterior variation on the unknown quantity of interest *T* under the same assumptions that provided our special case Bayesian analog to the classical general regression estimator $\hat{T}_{\mathrm{GRE}}$. This provides us with a Bayesian interpretation of the general regression estimator specified by Sarndal et al. (p.225, 1997). That is, we are able to construct interval estimates that have a strict probabilistic interpretation without reference to repeating sampling.

References

Bolfarine, H., Zacks, S. 1992. *Prediction theory for finite populations.* Springer Series in Statistics. Springer-Verlag, New York,

Cassel, C.M., Särndal, C.E., Wretman, J.H. 1976. Some results on generalized difference estimation and generalized regression estimation for finite populations. With comments by T. M. F. Smith and a reply by the authors. *Biometrika* 63**,** no. 3, 615-620.

Geisser, S. 1993. *Predictive inference. An introduction.* Monographs on Statistics and Applied Probability. Chapman and Hall, New York.

Ghosh, M., Meeden, G. 1997. *Bayesian Methods For Finite Population Sampling.* Chapman Hall, New York.

Harville, D.A. 1997. *Matrix Algebra From A Statistician's Perspective.* Springer-Verlag, New York.

Mukhopadhyay, N. 1993. Estimation of a finite population total under regression models: A review. *The Indian Journal Of Statistics* 55, series B, 141-155.

Renssen, R.H., Nieuwenbroek, N.J. 1997. Aligning estimates for common variables in two or more sample surveys. *J. Amer. Statist. Assoc.* 92, 368-374.

Royall, R.M. 1976. The linear least-squares prediction approach to two-stage sampling. *J. Amer. Statist. Assoc.* 71**,** no. 355, 657-664.

Royall, R.M., Pfeffermann, D. 1982. Balanced samples and robust Bayesian inference in finite population sampling. *Biometrika* 69, no. 2, 401-409.

Särndal, C.E. 1980. Two model-based inference arguments in survey sampling. *Austral. J. Statist.* 22, no. 3, 341-348.

Sarndal, C.E., Swensson, B., Wretman, J. 1997. *Model Assisted Survey Sampling.* Springer-Verlag, New York.