Texas Project NExT Journal
Volume 2, Number 1, Pages 11-20
Article electronically published on December 22, 2004

EMPIRICAL BAYES AND BAYES PREDICTION OF FINITE POPULATION TOTAL USING AUXILIARY INFORMATION

Mark S. Hamner[a], John W. Seaman, Jr.[b], Dean M. Young[b]

Abstract: 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. 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 classical approach to relating observed with unobserved population units. To assist in the population prediction we assume that a *p*-dimensional vector is known for each of the *N* population units. Known as auxiliary information, these *p*-dimensional vectors are used to obtain a very well-known classical estimator for finite population prediction known as the general regression estimator suggested by [1] and [5]. The general regression estimator is the popular design-based Horvitz-Thompson estimator plus an adjustment term. Sarndal [6] employ classical sampling design theory, using inclusion probabilities, and a regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathbf{X}$ is the auxiliary information matrix, $\boldsymbol{\beta}$ is the unknown coefficient vector, and $\mathbf{e} \sim N(\mathbf{0},\mathbf{V})$. The regression model, however, is used only as a means to obtain an estimate of $\boldsymbol{\beta}$. Hence, unbiasedness and variance expressions for the general regression estimator are derived under the classical sampling design approach. In contrast, a superpopulation model provides the stochastic structure for Bayesian inferential purposes. The superpopulation model establishes the main relations between the observed and unobserved units of $\mathbf{y}$. Using a specified superpopulation model, we derive an existing Bayesian estimator in matrix form and show the matrix mathematics involved in order to attain a Bayesian analog to a certain classical estimator.

1. The sample survey problem: Motivating the research

The need for obtaining information for statistical prediction is increasing in the modern times of the web and data warehousing. A particular area of prediction concerns a specific set of *N* identifiable elements in a finite population with unit labels $P = \{1, 2,\ldots, N\}$. For example, a university having budgetary concerns may begin by looking at the population of currently enrolled students and try to predict the total number of credit hours to expect from these students the next academic year. Predicting the total number of credit hours will help them make budget plans since student credit hours, not

Received by the editors October 18, 2004.
Key words and phrases. Superpopulation model; empirical Bayes estimator; general regression estimator; partitioned matrices

headcount, determine the university's state funding. Indeed, universities can identify all their students and generally have access to them either in person, mail, email or phone but the logistics of contacting all of them may be very time consuming and expensive. In such cases, inferences made from a random sample must suffice. Through a sample survey, which is a subset of measurements from the population, data is collected and we can employ a statistical estimator to make an inference about the population total. Often in the area of finite population prediction prior data exists that can be used to aid in the prediction. For our example, knowing cumulative hours, previous semester hours taken, class status (freshman, sophomore, etc.) and other *a priori* patterns may provide additional information with regards to enrollment status and potential hourly class load. Accordingly, inference for finite population sampling utilizes prior information either explicitly or implicitly.

In this paper we derive an existing Bayesian estimator in matrix form and show the mathematics involved in order to attain a Bayesian analog to a certain classical estimator. The derivations we employ are interesting in themselves, as they illustrate some of the underlying suppositions needed to connect the Bayesian approach to that of the classical approach in finite population total prediction.

In Section 2 we introduce a classical estimator called the Horvitz Thompson estimator (HTE) that uses a sampling design as a means to draw its stochastic properties. The Bayesian approach uses a superpopulation model for this purpose. The superpopulation defines the probability model that generates the values of interest, say y_i, for $i \in P$. Accordingly the auxiliary information may be an explicit part of the model that generates the values of interest. Thus, the superpopulation model approach considers the values of the population elements as random variables having joint distributions. We discuss how a synthesis often exists between the sampling design and superpopulation approaches. In Section 3 we derive Ghosh and Meeden's [3] finite population prediction result in a matrix context by specifying a particular superpopulation model. We show the appropriate matrix assumptions needed to derive the HTE from the matrix representation of Ghosh and Meeden's results, section 4.2 in [3]. In Section 4, we introduce the classical general regression estimator and then present it in a more convenient form using matrices. We conclude by proposing that our matrix representations can be used to facilitate the derivation of a fully Bayesian form of the general regression estimator.

2. Desirability of a Bayesian interpretation of important sample survey estimators

A well-known classical estimator for the population total $t = \sum_{k=1}^{N} y_k$ that uses the sampling-design method is the HTE estimator. The sampling design assigns a probability $p(s)$ to each sample $s \in S$, where S is the set of all possible samples. Let I_k be the inclusion function defined by

$$I_k = \begin{cases} 1 \text{ if } k \in s \\ 0 \text{ otherwise.} \end{cases}$$

The HTE is

$$\hat{t}_{HT} = \sum_{k=1}^{N} I_k \pi_k^{-1} y_k \ ,$$

where $\pi_k = E(I_k) = \sum_{k=1}^{N} I_k p(s)$ denotes the inclusion probability of element $k \in P$ into the sample s. Note that the HTE is unbiased with respect to p. Alternatively, for a given sample of size-n, the HTE has the form $\hat{t}_{HT} = \sum_{k=1}^{n} \pi_k^{-1} y_k$.

The inability of classical theory to provide reasonable answers in various situations prompted the development of the superpopulation approach. For a humorous illustration of how the HTE can fail to give a reasonable answer see [1] or section 5.2 in [3].

Interestingly, Ghosh [3] present a step-wise Bayesian estimator that, with specific assumptions, provides a Bayesian justification for the HTE which is something of a special case. However, the aforementioned problem of the classical HTE is avoided under minimized Bayes risk for their special-case Bayes estimator.

Even though the two approaches fundamentally differ in their interpretations, a synthesis between the Bayesian approach and the classical approach often exists, in that operationally equivalent estimators are obtainable through both approaches. For example, the classical ratio estimator has been obtained through various Bayesian superpopulation models (see, for example, Section 3.3.1 in [3] and [4] (p. 402)). The HTE has also been shown to have an empirical Bayesian analog; see Section 4.2 in [3], and a fully Bayesian analog; see Section 5.1 in [3]. Often the synthesis between approaches is revealed upon rewriting estimators in matrix form, as we shall see below.

3. Matrix derivation the Ghosh and Meeden result

In this section we begin by presenting the results of an empirical Bayesian estimator that is presented by Ghosh and Meeden, Section 5.1in [3]. Next, we will derive the results of Ghosh and Meeden in a matrix context by specifying a particular superpopulation model.

Ghosh [3], present an empirical Bayes estimator of the finite population mean for a quantity of interest, y_1 , y_2 , …, y_N. In this section we briefly present their empirical Bayes estimate of the population mean and then present it in a more convenient form using matrices. Whereas Ghosh and Meeden establish their results without assuming a distribution for the unknown quantity, we will specify a specific superpopulation model for the quantity of interest, y_1, y_2,…, y_N. Use of our matrix representation facilitates subsequent derivations.

Ghosh and Meeden consider a superpopulation model conditional on θ, in which y_1 , y_2 , …, y_N are independently distributed with $E(y_i) = \theta a_i$ and $V(y_i) = \sigma_i^2$. They assume a_i and σ_i^2 are known constants. Let s represent a sample of size n selected from P and denote by r the remaining $N - n$ non-sampled units from P. Then the finite population mean μ is estimated by

$$E(\mu|y_s) = \sum_{i\in s} y_i + \theta \sum_{j\in r} a_j \ . \qquad (3.1)$$

Their empirical Bayes estimator requires estimation of the unknown parameter θ. They use the following weighted least squares estimate of θ

$$\hat{\theta}_{LE} = \frac{\sum_{i \in s} a_i \sigma_i^{-2} y_i}{\sum_{i \in s} a_i^2 \sigma_i^{-2}} . \tag{3.2}$$

Using the weighted least squares estimate of θ they obtain the empirical Bayes estimator (EBE) of μ

$$\hat{\mu}_{EB} = N^{-1}\left[\sum_{i \in s} y_i + \left(\frac{\sum_{i \in s} a_i \sigma_i^{-2} y_i}{\sum_{i \in s} a_i^2 \sigma_i^{-2}}\right)\sum_{j \in r} a_j\right]. \tag{3.3}$$

Ghosh and Meeden note that other interesting estimators are obtainable from equation (3.3). In particular, by letting $a_i = \pi_i$, $\sigma_i^2 = \pi_i^2 / (1 - \pi_i)$, and having $\sum_{i=1}^{N} \pi_i = n$ "one obtains the celebrated Horvitz-Thompson estimator" $\sum_{i \in s} \pi_i^{-1} y_i$ [3] (p.164).

We now rewrite the results of Ghosh and Meeden in matrix form by assuming the following superpopulation model

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

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

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_s \\ \hline \mathbf{y}_r \end{bmatrix}$$

such that $\mathbf{y}_s$ 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. Assume $\mathbf{A}$ is a known $N \times N$ diagonal matrix

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

such that $\mathbf{A}_s$ 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 $a_{n+1}, a_{n+2}, \ldots, a_N$. Furthermore, $\mathbf{1}$ is an $N \times 1$ vector of ones

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

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 $\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \mathbf{V})$ where $\mathrm{N}(\mathbf{u}, \mathbf{W})$ 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

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

where $\mathbf{V}_{ss}$ is an $n \times n$ diagonal matrix with diagonal elements σ_i^2 for i = 1, 2, …, n, and 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 we first note that the population total can written as

$$\begin{aligned} T &= 1'\mathbf{Y} \\ &= \left[\mathbf{1}'_s \mid \mathbf{1}'_r\right]\left[\begin{array}{c} \mathbf{y}_s \\ \hline \mathbf{y}_r \end{array}\right] \\ &= \mathbf{1}'_s\mathbf{y}_s + \mathbf{1}'_r\mathbf{y}_r \end{aligned}$$

Thus, the population average can be written as

$$\begin{aligned} \mu &= N^{-1}\left[T\right] \\ &= N^{-1}\left[\mathbf{1}'_s\mathbf{y}_s + \mathbf{1}'_r\mathbf{y}_r\right] \end{aligned}$$

and 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 $\mathrm{E}(\mathbf{y}_r|\mathbf{y}_s)$ for estimation of $\mathbf{y}_r$. This posterior expectation can be obtained by the nested expectations

$$\mathrm{E}(\mathbf{y}_r|\mathbf{y}_s) = E_\theta\left[E_{\mathbf{y}_r}\left(\mathbf{y}_r \mid \mathbf{y}_s, \theta\right)\middle|\mathbf{y}_s\right]$$

where $\mathrm{E}_{\mathbf{y}_r}(\cdot)$ is the expectation with respect to the likelihood of $\mathbf{y}_r$ and $\mathrm{E}_\theta(\cdot)$ is the expectation with respect to the posterior of θ. Under model $\mathbf{Y} = \mathbf{A1}\theta + \mathbf{e}$,

$$E_{\mathbf{y}_r}\left(\mathbf{y}_r \mid \mathbf{y}_s, \theta\right) = \mathbf{A}_r\mathbf{1}_r\theta.$$

Now, instead of taking the expectation $\mathrm{E}_\theta(\mathbf{A}_r\mathbf{1}_r\theta)$, an empirical Bayes result ensues by estimating the unknown parameter θ using a weighted least squares estimate. Let $\mathbf{X}_s = \mathbf{A}_s\mathbf{1}_s$, then the least squares estimate of θ, equation (3.2), in matrix form is

$$\begin{aligned} \hat{\theta}_{LE} &= \left(\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}'_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s \\ &= \left(\mathbf{1}'_s\mathbf{A}'_s\mathbf{V}_{ss}^{-1}\mathbf{A}_s\mathbf{1}_s\right)^{-1}\mathbf{1}'_s\mathbf{A}'_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s \\ &= \left(\mathbf{1}'_s\mathbf{A}_s\mathbf{V}_{ss}^{-1}\mathbf{A}_s\mathbf{1}_s\right)^{-1}\mathbf{1}'_s\mathbf{A}_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s. \end{aligned} \tag{3.4}$$

Thus, an empirical Bayes estimate of the population total is

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

$$=\mathbf{1}_s'\mathbf{y}_s+\mathbf{1}_r'\mathbf{A}_r\mathbf{1}_r\left(\mathbf{1}_s'\mathbf{A}_s\mathbf{V}_{ss}^{-1}\mathbf{A}_s\mathbf{1}_s\right)^{-1}\mathbf{1}_s'\mathbf{A}_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s \tag{3.5}$$

Using equation (3.5), we can rewrite equation (3.3) in matrix form as

$$\begin{aligned}\hat{\mu}_{EB} &= N^{-1}\left[\hat{T}_{EB}\right]\\ &= N^{-1}\left[\mathbf{1}_s'\mathbf{y}_s+\mathbf{1}_r'\mathbf{A}_r\mathbf{1}_r\hat{\theta}_{LE}\right]\\ &= N^{-1}\left\{\mathbf{1}_s'\mathbf{y}_s+\mathbf{1}_r'\mathbf{A}_r\mathbf{1}_r\left[\left(\mathbf{1}_s'\mathbf{A}_s\mathbf{V}_{ss}^{-1}\mathbf{A}_s\mathbf{1}_s\right)^{-1}\mathbf{1}_s'\mathbf{A}_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s\right]\right\}.\end{aligned} \tag{3.6}$$

Next we will use the matrix form of the empirical Bayes estimate, equation (3.6) of the population mean and derive the HTE using the assumptions noted by Ghosh and Meeden. Let π_i represents the inclusion probability for an individual k from a finite population of size N where i = 1, 2,…, N. Define π to be an $N \times N$ matrix whose diagonal elements, $\pi_{ii} \equiv \pi_i$, i = 1, 2,…N, are the inclusion probabilities and the off-diagonal elements are zero. We can write π as

$$\pi = \left[\begin{array}{c|c}\pi_s & \mathbf{0}\\ \hline \mathbf{0} & \pi_r\end{array}\right]$$

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

$$\begin{aligned}n &= \mathbf{1}'\mathbf{\Pi}\mathbf{1}\\ &= \left[\mathbf{1}_s' \;\vdots\; \mathbf{1}_r'\right]\left[\begin{array}{c|c}\pi_s & \mathbf{0}\\ \hline \mathbf{0} & \pi_r\end{array}\right]\left[\begin{array}{c}\mathbf{1}_s\\ \hline \mathbf{1}_r\end{array}\right]\\ &= \mathbf{1}_s'\pi_s\mathbf{1}_s+\mathbf{1}_r'\pi_r\mathbf{1}_r\end{aligned}$$

which implies

$$\mathbf{1}_r'\pi_r\mathbf{1}_r = n-\mathbf{1}_s'\pi_s\mathbf{1}_s\ . \tag{3.7}$$

Let $\mathbf{A} = \pi$, $\mathbf{V} = \pi(\mathbf{I}-\pi)^{-1}\pi$ where $\mathbf{I}$ is an $N \times N$ 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. It is interesting to note that the assumption for $\mathbf{V}$ above is the matrix representation for the for the covariance structure of a geometric distribution. Using equation (3.7), then equation (3.4) becomes

$$\begin{aligned}\hat{\theta}_{LE} &= \left(\mathbf{1}_s'\mathbf{A}_s\mathbf{V}_{ss}^{-1}\mathbf{A}_s\mathbf{1}_s\right)^{-1}\mathbf{1}_s'\mathbf{A}_s\mathbf{V}_{ss}^{-1}\mathbf{y}_s\\ &= \left\{\mathbf{1}_s'\boldsymbol{\pi}_s\left[\boldsymbol{\pi}_s\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)^{-1}\boldsymbol{\pi}_s\right]^{-1}\boldsymbol{\pi}_s\mathbf{1}_s\right\}^{-1}\mathbf{1}_s'\boldsymbol{\pi}_s\left[\boldsymbol{\pi}_s\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)^{-1}\boldsymbol{\pi}_s\right]^{-1}\mathbf{y}_s\end{aligned}$$

$$= \left[\mathbf{1}_s'\boldsymbol{\pi}_s\boldsymbol{\pi}_s^{-1}\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\boldsymbol{\pi}_s\mathbf{1}_s\right]^{-1}\mathbf{1}_s'\boldsymbol{\pi}_s\boldsymbol{\pi}_s^{-1}\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\mathbf{y}_s$$
$$= \left[\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_{\mathrm{s}}\right)\mathbf{1}_s\right]^{-1}\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\mathbf{y}_s$$
$$= \left[\mathbf{1}_s'\mathbf{1}_s-\mathbf{1}_s'\boldsymbol{\pi}_s\mathbf{1}_s\right]^{-1}\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\mathbf{y}_s$$
$$= \left[n-\mathbf{1}_s'\boldsymbol{\pi}_s\mathbf{1}_s\right]^{-1}\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\mathbf{y}_s$$
$$= \left[\mathbf{1}_r'\pi_r\mathbf{1}_r\right]^{-1}\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\mathbf{y}_s\,. \tag{3.8}$$

With equation (3.5), equation (3.7) and the assumptions made for **A** and **V**, the empirical Bayes estimator (3.6) becomes

$$\hat{\mu}_{EB} = N^{-1}\left[\hat{T}_{EB}\right]$$
$$= N^{-1}\left[\mathbf{1}_s'\mathbf{y}_s+\mathbf{1}_r'\boldsymbol{\pi}_r\mathbf{1}_r\hat{\theta}_{LE}\right]$$
$$= N^{-1}\left[\mathbf{1}_s'\mathbf{y}_s+\mathbf{1}_r'\boldsymbol{\pi}_r\mathbf{1}_r\left[\mathbf{1}_r'\pi_r\mathbf{1}_r\right]^{-1}\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\mathbf{y}_s\right]$$
$$= N^{-1}\left[\mathbf{1}_s'\mathbf{y}_s+\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\mathbf{y}_s\right]$$
$$= N^{-1}\left[\mathbf{1}_s'+\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\right]\mathbf{y}_s$$
$$= N^{-1}\left[\mathbf{1}_s'\boldsymbol{\pi}_s\boldsymbol{\pi}_s^{-1}+\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\boldsymbol{\pi}_s^{-1}\right]\mathbf{y}_s$$
$$= N^{-1}\left[\mathbf{1}_s'\boldsymbol{\pi}_s+\mathbf{1}_s'\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\right]\boldsymbol{\pi}_s^{-1}\mathbf{y}_s$$
$$= N^{-1}\mathbf{1}_s'\left[\boldsymbol{\pi}_s+\left(\mathbf{I}_s-\boldsymbol{\pi}_s\right)\right]\boldsymbol{\pi}_s^{-1}\mathbf{y}_s$$
$$= N^{-1}\mathbf{1}_s'\mathbf{I}_s\boldsymbol{\pi}_s^{-1}\mathbf{y}_s$$
$$= N^{-1}\mathbf{1}_s'\boldsymbol{\pi}_s^{-1}\mathbf{y}_s\,. \tag{3.9}$$

Notice that equation (3.9) implies an empirical Bayes estimate of the population total is

$$\hat{T}_{EB} = \mathbf{1}_s'\boldsymbol{\pi}_s^{-1}\mathbf{y}_s\,. \tag{3.10}$$

Equation (3.10) is the matrix representation of the HTE, sometimes referred to as the π-*estimator*, for the population total.

In general, by letting $\mathbf{X}_s = \mathbf{A}_s\mathbf{1}_s$ and $\mathbf{X}_r = \mathbf{A}_r\mathbf{1}_r$ equation (3.5) becomes

$$\hat{T}_{EB} = \mathbf{1}_s'\mathbf{y}_s+\mathbf{1}_r'E_{\mathbf{y}_r}\left(\mathbf{y}_r \mid \mathbf{y}_s,\theta\right)$$
$$= \mathbf{1}_s'\mathbf{y}_s+\mathbf{1}_r'\mathbf{X}_r\left(\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\mathbf{V}_{ss}^{-1}\mathbf{y}_s \tag{3.11}$$

Equation (3.11) is a matrix representation of the results presented by Ghosh and Meeden, equation (3.3). In terms of application, our matrix representation can easily be implemented using any software that facilitates simple matrix algebra.

4. GRE result in Matrix form

In this section we introduce the general regression estimator and then present it in a more convenient form using matrices. As a means to possibly improve the basic π-estimator using auxiliary information, Sarndal [6] employ classical sampling design theory, using inclusion probabilities, and the regression model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ where $\mathbf{e} \sim N(\mathbf{0},\mathbf{V})$. The latter is used, however, only as a means to obtain an estimate of β. Hence, unbiasedness and variance expressions are derived under the sampling design. In short, Sarndal [6] do not assume that the regression model generated the sample. Thus, the *general regression estimator* (GRE) derived in Sarndal [6] is *model assisted* but not model dependent. Sarndal [6] (p.225) define the GRE as

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

where y_k is a variable of interest 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 $\beta \equiv (\beta_1,\ldots,\beta_p)'$, Sarndal [6] (p. 228) suggest the following estimate

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

where $\mathbf{X}_k$ is a $p \times 1$ vector of known auxiliary information. Without loss of generality we will rewrite $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_s \\ \hline \mathbf{X}_r \end{bmatrix} \tag{4.1}$$

such that rank($\mathbf{X}$) = p, with $\mathbf{X}_s \in \mathbb{R}_{n\times p}$, and $\mathbf{X}_r \in \mathbb{R}_{(N-n)\times p}$. Still, under a simple random sampling design in which $\pi = \left(\frac{n}{N}\right)^{-1} \mathbf{I}, \hat{\beta}$ is the 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 Σ is a $p \times p$ positive definite covariance matrix. Now, using the notation established in equation (4.1) and in section section-3, we rewrite the GRE in matrix form as

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

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

$$= \mathbf{1}_s'\pi_s^{-1}\mathbf{y}_s + \mathbf{1}'\mathbf{X}\hat{\beta}_s - \mathbf{1}_s'\pi_s^{-1}\mathbf{X}_s\hat{\beta}_s$$

$$= \mathbf{1}'\mathbf{X}\hat{\beta}_s + \mathbf{1}_s'\pi_s^{-1}\left(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s\right). \tag{4.2}$$

Equation (4.2) is the matrix representation of the alternative expression for the general regression estimator presented by Sarndal [6] (p. 231). Looking at equation (4.2) we note that the GRE uses the sample data to establish the estimate for β. Yet, the left part of the summand in equation 4.3 uses the auxiliary information, $\mathbf{X}$, that is known for the entire population. Thus, the left part of the summand in equation (4.2) basically consists of a sum of N "guesses" where for each $i \in P$ the "guess", call it $\hat{y}_i$, is $\mathbf{X}_i\ \hat{\beta}_s$. For the sampled data, $\mathbf{y}_s$, we note that

$$\mathbf{y}_s = \mathbf{X}_s\hat{\beta}_s + \left(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s\right) \tag{4.3}$$

where $\left(\mathbf{y}_s - \mathbf{X}_s\hat{\beta}_s\right)$ is the error vector associated with the prediction of $\mathbf{y}_s$ with $\mathbf{X}_s\hat{\beta}_s$. We see that equation (4.3) gives insight to the fundamental workings of equation (4.2). Equation (4.2) is predicting the total for all y_i, $i \in P$, and even though there are only n errors, the right part of the summand in equation (4.2) weights these errors with π_s^{-1}. Under simple random sampling equation (3.7) implies that $\text{tr}(\pi_s) = \frac{1}{N}$. This means that each error value, $\left(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}_s\right)$, gets weighed by $\frac{N}{n}$. Thus, equation (4.2) adds the weighted errors, where $\text{tr}(\pi_s^{-1}) = \text{N}$, to the sum of the N "guesses", $\hat{y}_i$ where $i \in P$. More specifically, assuming the elements of $\mathbf{Y}$ are independently and identically distributed, under simple random sampling design we can express equation (4.1) 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 projection matrix onto $\mathcal{C}(\mathbf{X}_s)$, the column space of $\mathbf{X}_s$, $\mathbf{X}_s^{+}$ is the Moore-Penrose inverse of $\mathbf{X}_\mathbf{s}$, and $\mathbf{P}_{\mathbf{X}_s}^{\perp}$ is the orthogonal projection matrix onto the orthogonal complement of $\mathcal{C}(\mathbf{X}_s)$. In a matrix context, we see that $\mathbf{P}_{\mathbf{X}_s}^{\perp}\ \mathbf{y}_s$ is weighted residual "information" that is left over from projecting $\mathbf{y}_s$ onto the column space of $\mathbf{X}_s$. This weighted information is then added to the estimate of the population total $\mathbf{1}'\mathbf{X}\hat{\beta}_s$.

5. Conclusion

In Section-3 we derived the empirical Bayes estimator of Ghosh [3] by specifying a particular superpopulation model. We then showed the mathematical detail needed in order to show how the HTE can be derived from the empirical Bayesian estimator using matrices. Related to the HTE is another classical population total estimator called the general regression estimator. This estimator is the HTE plus an adjustment term. To our knowledge, the literature has yet to present a direct empirical Bayesian or fully Bayesian analog to the general regression estimator. However, our matrix representations can be used to facilitate the derivation of a fully Bayesian form of the general regression estimator.

References

[1] D. Basu. An essay on the logical foundations of survey, part-one. In: Godambe, V.P., and Sprott, D.A. (eds), Foundations of Statistical Inference. Toronto: Holt, Rinehart and Winston, 1971.

[2] C.M.Cassel, C.E. Särndal, and J.H. Wretman. 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, (1976), 615-620.

[3] M. Ghosh and G Meeden. Bayesian methods for finite population sampling. Chapman Hall, New York, 1997.

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

[5] C.E. Särndal. Two model-based inference arguments in survey sampling. Austral. J. Statist. 22, no. 3, (1980), 341-348.

[6] C.E. Särndal, B. Swensson, and J. Wretman. Model assisted survey sampling. Springer-Verlag, New York, 1992.

[a]*Department of Mathematics and Computer Science, Texas Woman's University, P.O. Box 425886, Denton, TX 76204-5886, USA.*
Email address: *mahmner@twu.edu*

[b]*Department of Statistical Sciences, Baylor University, P.O. Box 7140, Waco, TX, 76798-7140. USA*