Division of Research
Graduate School of Business Administration
The University of Michigan

April 1982

FINITE POPULATION SAMPLING WITH MULTIVARIATE AUXILIARY INFORMATION

Working Paper No. 302

Roger L. Wright

The University of Michigan

FOR DISCUSSION ONLY

ABSTRACT

This paper examines finite population estimation and sample design from a robust, model-based viewpoint. The paper introduces a new class of multivariate regression estimators that integrates several model-based procedures and clarifies the role of weighted least squares and analysis of residuals in sampling. A new model-based procedure is suggested for designing an efficiently stratified sampling plan.

KEY WORDS: Balanced sampling, Regression estimators, Robustness, Stratification, Superpopulation models, Unequal probability sampling.

Author's Footnote:

Roger L. Wright is Associate Professor of Statistics, Graduate School of Business Administration, The University of Michigan, Ann Arbor, MI 48109.
This paper was prepared with the support of the U.S. Department of Energy, Grant No. DE-FG02-80ER10125, the Electric Power Research Institute, and Consumers Power Company. However, any opinions, findings, conclusions, or recommendations expressed herein are those of the author and do not necessarily reflect the views of the sponsors.

The author wishes to thank K.R.W. Brewer, Tore Dalenius, Leslie Kish, Carl Sarndal, and Allen Spivey for valuable comments on earlier versions of this paper.

1. INTRODUCTION

Various strategies have been proposed for combining sample data with available auxiliary information. One approach is to adapt linear model prediction theory to the finite population context, giving a highly efficient BLU estimator if the model is correctly specified but one that can be seriously biased if the assumed model is inaccurate [Brewer (1963), Hansen, Madow, and Tepping (1978), Royall (1970), and Smith (1976)]. An alternative procedure is the generalized regression estimators which feature greater robustness to model misspecification while retaining much of the efficiency of the best linear unbiased (BLU) estimator [Cassel, Sarndal and Wretman (1976) and Sarndal (1980 a,b)]. Recently, Brewer (1979) suggested a robust estimator that blends aspects of the BLU and generalized regression estimators but is restricted to a single auxiliary variable. Isaki and Fuller (1982) gave a closely related multivariate estimator.

This paper introduces a large class of robust estimators that utilize multivariate auxiliary information in finite population sampling. This new class includes both BLU and generalized regression estimators, as well as many more conventional sampling estimators. New results are given that suggest the pervasiveness of the generalized regression estimators within the class of robust estimators. A by-product of the analysis is added insight into the relationship between survey sampling and applied regression analysis, especially concerning the widely encountered problems of weighting and analysis of residuals.

The paper also examines model-based sample design, and a new procedure is suggested for constructing efficiently stratified sampling plans that preserve the robustness of the generalized regression estimators. This procedure is a natural extension of Neyman allocation to the model-based case, and it provides an alternative to the balanced sampling approach of Royall and Herson (1973 a,b), to nonrandom model-based designs, and to other stratified sampling designs such as those of Andrews, Kish, and Cornell (1980), Dalenius and Hodges (1959), Singh (1975), and Rao (1977).

The paper discusses model-based estimation in Section 2, first in the ratio case involving a single auxiliary variable and then in the more general multivariate situation. Sample design is discussed in Section 3, and to simplify the presentation, the derivation of several key results is deferred to Section 4.

2. MODEL-BASED ESTIMATION

2.1 The Ratio Case

The model-based approach blends elements of finite population sampling and linear statistical models. Assume that a sample of n units is to be randomly selected from a population of N units labeled $I = 1,\ldots,N$. The probability of obtaining each possible sample s is denoted by p(s), and the probability that unit I is included in the selected sample is denoted π_I. In general, the inclusion probabilities may vary from unit to unit but are known from the sampling plan.

Consider a superpopulation model ξ, under which the target variable of interest, say y_I, is related to an auxiliary variable x_I following a simple zero-intercept regression equation:

$$y_I = \beta x_I + u_I. \qquad (1)$$

The usual assumption that each residual u_I has zero expectation under ξ is employed; thus, $E_\xi(u_I) = 0$. The auxiliary variable x_I in (1) is assumed to be positive and known for each unit in the population.

The model ξ is generally taken to be heteroscedastic, with varying residual standard deviation σ_I associated with each unit in the population. Our analysis will assume that the σ_I are known, but in practice we have used available data to estimate a functional relationship between σ_I and suitable auxiliary variables, using a technique developed by Harvey (1976). In addition, it is assumed that $u_1,\ldots,u_N$ are uncorrelated. Thus, under ξ, we have $E_\xi(u_I) = 0$, $E_\xi(u_I^2) = \sigma_I^2$, and $E_\xi(u_I u_J) = 0$ for $I \neq J$.

There is considerable confusion about estimation when varying probabilities of inclusion and varying residual standard deviations are involved. In light of the heteroscedasticity of ξ, many analysts would employ a weighted least squares (WLS) procedure with weights determined by σ_I. On the other hand, most survey samplers would emphasize the sampling plan and recommend WLS with weights determined by π_I and possibly x_I. Table 1 shows five different estimators of the population total $Y = \sum_{I=1}^{N} y_I$ that have been recommended in the recent statistical literature.

(Table 1 about here)

A framework for examining these alternatives is produced by embedding them within a class of estimators that is amenable to analysis. In fact, each of the estimators shown in Table 1 can be written in the form

$$\hat{Y}_{QR} = \hat{\beta}X + \sum_{I\varepsilon s} r_I\hat{u}_I, \qquad (2)$$

where $\hat{\beta} = \sum_{I\varepsilon s} q_I x_I y_I \,/\, \sum_{I\varepsilon s} q_I x_I^2$

and $\hat{u}_I = y_I - \hat{\beta}x_I$

with $q_I > 0$ and $r_I \geq 0$.

Here $\hat{\beta}X$ is the naive regression-based estimator using the finite population total X, and the second term is a correction based on the sample residuals.

The q_I are weights used to calculate the regression coefficient $\hat{\beta}$, and r_I is a weight used to extend the observed residual. Table 2 shows how q_I and r_I can be chosen for each of the estimators shown in Table 1.

(Table 2 about here)

One reason why so many alternative ratio-type estimators have been proposed is that there are two conflicting bases for statistical inference in survey sampling. Under model-based inference, the sampling distribution of an estimator $\hat{Y}$ is considered to be induced by the joint distribution of the residuals u_I of the model ξ. In terms of the expectation E_ξ taken with respect to ξ, $\hat{Y}_{QR}$ is an unbiased predictor of the random variable $Y = \sum_{I=1}^{N} y_I$ for any choice of q_I and r_I. However, Royall has shown that the best linear unbiased (BLU) predictor of Y, denoted $\hat{Y}_{BLU}$, uses the model-based WLS estimator $\hat{\beta}_{BLU}$ given by choosing $q_I = \sigma_I^{-2}$, and places unit weight on the observed sample residuals. Although $\hat{Y}_{BLU}$ completely ignores the sampling plan, it would be preferred by many statisticians if they were certain of the accuracy of the model ξ. The major objection to $\hat{Y}_{BLU}$ is that it can be seriously biased if ξ is even moderately inaccurate.

To protect against such dependence on ξ--i.e., to provide robustness--survey samplers have traditionally emphasized design-based inference. Here the y_I are assumed to take unknown but fixed values throughout the population, but the estimator $\hat{Y}$ is considered to have a sampling distribution induced by the sampling plan. Estimators that are design-unbiased have a desirable robustness since they are less dependent on the accuracy of a model such as ξ.

Design unbiasedness can be provided by linking the choice of weights used in the estimator to the inclusion probabilities of the sampling plan. Consider an estimator $\hat{Y}_R = \sum_{I\varepsilon s} r_I y_I$, and let $E_p(\hat{Y}_R)$ denote its design-based expected value.

$$E_p(\hat{y}_R) = \sum_s p(s) \sum_{I\varepsilon s} r_I y_I$$

$$= \sum_{I=1}^{N} r_I y_I \sum_{s\varepsilon S_I} p(s), \text{ where } S_I = \{s | I\varepsilon s\},$$

$$= \sum_{I=1}^{N} r_I \pi_I y_I .$$

This implies that $\hat{Y}_R$ is a design-unbiased estimator of Y if and only if for all $y_1,\ldots,y_N$,

$$\sum_{I=1}^{N} r_I \pi_I y_I = \sum_{I=1}^{N} y_I ;$$

or, equivalently, if and only if $\pi_I > 0$ and $r_I = \pi_I^{-1}$ for all I. So $\hat{Y}_R$ is a design-unbiased estimator of Y if and only if it is the Horvitz-Thompson estimator $\hat{Y}_{HT}$ for a sampling plan having positive inclusion probabilities for all units in the population.

Unfortunately, design-based analysis of the estimator $\hat{Y}_{QR}$ is complicated by the nonlinearity in $\hat{\beta}$. However, it is useful to examine approximate design unbiasedness in large samples from even larger populations. An estimator $\hat{Y}$ is said to be an asymptotically design-unbiased (ADU) estimator of Y if and only if for all $y_1,\ldots,y_N$

$$\lim_{(n,N)\to\infty} E_p(\hat{Y}) = Y. \qquad (3)$$

The sense in which this limit is taken follows Brewer (1979) and Sarndal (1980a) and is described in Section 4. Two results are shown:

Result 1. $\hat{Y}_{QR}$ is ADU for Y if and only if there is some constant λ such that

$$1-r_I\pi_I = q_I \pi_I x_I \lambda \text{ for all } I=1,\ldots,N. \qquad (4)$$

Result 2. $\hat{Y}_{QR}$ is ADU for Y if and only if for all $y_1, \ldots, y_N$

$$\sum_{I \varepsilon s} r_I \hat{u}_I = \sum_{I \varepsilon s} \pi_I^{-1} \hat{u}_I. \qquad (5)$$

Result 1 gives an algebraic condition on π_I, q_I, and r_I that guarantees that $\hat{Y}_{QR}$ is ADU for Y. Using Table 2, Result 1 shows that $\hat{Y}_{HTR}$, $\hat{Y}_{CR}$, $\hat{Y}_{GR}$, and $\hat{Y}_{RO}$ are all ADU for Y, but $\hat{Y}_{BLU}$ is not. The generalized regression estimator $\hat{Y}_{GR}$ was developed to retain the use of $\hat{\beta}_{BLU}$ as in $\hat{Y}_{BLU}$, while offering the ADU robustness provided by the more conventional design-based estimators $\hat{Y}_{HTR}$ and $\hat{Y}_{CR}$. Brewer's $\hat{Y}_{RO}$ was apparently introduced in order to retain the $r_I = 1$ used in $\hat{Y}_{BLU}$ while offering conventional robustness. So Result 1 helps to sort out features of these alternative estimators.

Result 1 also shows how to construct a large family of robust, model-based estimators, i.e., estimators that are model-based but still ADU. Specifically, $\hat{Y}_{QR}$ is ADU for Y for any choice of q_I as long as $r_I = \pi_I^{-1}$.

Result 2 provides a converse to the previous statement; it states that whenever $\hat{Y}_{QR}$ is ADU for Y, any choice of r_I is equivalent to choosing $r_I = \pi_I^{-1}$. In other words, as long as two estimators in the class (2) are ADU in our sense and use identical q_I, they are identical estimators; there is nothing to be gained from any choice of $r_I \neq \pi_I^{-1}$ except computational simplicity.

These results are consistent with the sound practice that has developed in applied regression analysis of examining the sample residuals for information about model misspecification. In using regression analysis for finite population inference, we find that robustness can be obtained by adding a Horvitz-Thompson-like residual correction $\sum_{I \varepsilon s} \pi_I^{-1} \hat{u}_I$ to the regression-based estimate $\hat{\beta}X$.

To summarize so far, the choice of estimator in the ratio case reduces to two questions: (1) Do we want the robustness provided by an ADU estimator or can we place faith in the model ξ and use $\hat{Y}_{BLU}$? (2) If we want an ADU estimator $\hat{Y}_{QR}$, how should we choose the weight q_I to be used in the WLS regression? The answer to the first question is highly dependent on the purpose and context of the project and the credibility of the model. The second question will be addressed subsequently by examining the asymptotic variance of $\hat{Y}_{QR}$.

2.2 Estimation with Multivariate Auxiliary Information

The preceding results extend easily to the multiple regression case. For added generality assume that the population parameter of interest is $a'y = \sum_{I=1}^{N} a_I y_I$, where a_I is known for each unit in the population. For example, the a_I may indicate a subpopulation of interest. Under ξ, we now assume that y_I is related to a vector $X_I' = (x_I, \ldots x_{Ik})$ of $k \geq 1$ auxiliary variables following a linear regression equation of the form

$$y_I = \beta_1 x_{I1} + \ldots + \beta_k x_{Ik} + u_I, \tag{6}$$

with $E_\xi(u_I) = 0$, $E_\xi(u_I^2) = \sigma_I^2$, and $E_\xi(u_I u_J) = 0$ for $I \neq J$. For any choice of $r_I \geq 0$ and any $q_I \geq 0$ such that $(\sum_{I\varepsilon s} q_I X_I X_I')^{-1}$ exists for any possible s, define a QR estimator as

$$a'\hat{y}_{QR} = \sum_{I=1}^{N} a_I X_I'\hat{\beta} + \sum_{I\varepsilon s} r_I a_I \hat{u}_I, \text{ where} \tag{7}$$

$$\hat{\beta} = (\sum_{I\varepsilon s} q_I X_I X_I')^{-1} \sum_{I\varepsilon s} q_I X_I y_I$$

(a WLS estimator), and

$$\hat{u}_I = y_I - X_I'\hat{\beta}.$$

Royall (1976) has shown that under ξ, the BLU model-based predictor of $a'y$ is $a'\hat{y}_{BLU}$, given by choosing $q_I = \sigma_I^{-2}$ and $r_I = 1$. Sarndal (1980 a,b) has considered generalized regression estimators that use various q_I and $r_I = \pi_I^{-1}$. Isaki and Fuller (1982) have studied the estimator given by $q_I = \pi_I^{-2}$ and $r_I = 0$.

To examine the robustness of QR estimators, define $a'\hat{y}_{QR}$ to be ADU for $a'y$ if and only if for all $y_1,\ldots,y_N$

$$\lim_{(n,N)\to\infty} E_p(a'\hat{y}_{QR}) = a'y.$$

Then the two results given for the ratio case can be generalized as follows:

<u>Result 3</u>. $a'\hat{y}_{QR}$ is ADU for $a'y$ if and only if there is some vector $\lambda = (\lambda_1 \cdots \lambda_k)'$ such that for all $I = 1,\ldots,N$

$$(1-r_I\pi_I)a_I = q_I\pi_I X_I'\lambda. \qquad (8)$$

<u>Result 4</u>. $a'\hat{y}_{QR}$ is ADU for $a'y$ if and only if for all $y_1,\ldots,y_N$

$$\sum_{I\varepsilon s} r_I a_I \hat{u}_I = \sum_{I\varepsilon s} \pi_I^{-1} a_I \hat{u}_I. \qquad (9)$$

Result 3 shows that $a'\hat{y}_{QR}$ is ADU for all population parameters $a'y$ if and only if $r_I = \pi_I^{-1}$, as in the generalized regresson estimators. The BLU model-based predictor, using $q_I = \sigma_I^{-2}$ and $r_I = 1$, is ADU only if $(\pi_I^{-1}-1)\ a_I\sigma_I^2$ is equal to a linear combination of the auxiliary variables in X_I throughout the population. The Isaki and Fuller estimator, using $q_I = \pi_I^{-2}$ and $r_I = 0$, is ADU if and only if $a_I\pi_I$ is a linear combination of X_I throughout the population.

Result 4 shows that any ADU QR estimator (7) is identical to the generalized regression estimator that uses the same q_I with $r_I = \pi_I^{-1}$. Proofs of Results 3 and 4 are given in Section 4.

3. ROBUST MODEL-BASED SAMPLE DESIGN

In sample design, as in estimation, the challenge is to develop an approach that takes advantage of the model (6) but is not totally dependent on the model's accuracy, i.e., a robust model-based approach to sample design. Suppose for a moment that (6) is indeed considered to be unquestionably accurate. Then the sample plan ought to be designed to minimize the model-based expected variance of $a'\hat{y}_{BLU}$. Royall has shown that this leads to non-random sampling. In particular, under fairly common circumstances, the optimal design under ξ is to select the n units that have the largest residual standard deviation (σ_I) in the population. This is unacceptable in many applications because the slightest inaccuracy in ξ will produce substantial but almost undetectable biases.

For a robust model-based approach, consider any estimator $a'\hat{y}_{QR}$ that is ADU for the parameter of interest $a'y$. At the planning stage, it is reasonable to utilize both the model ξ and the proposed sample plan to evaluate the anticipated performance of the estimator. Nonlinearity makes exact analysis difficult, but for many purposes a large-sample approximation is satisfactory. So we define the asymptotic variance of $a'\hat{y}_{QR}$ to be

$$V(a'\hat{y}_{QR}) = \lim_{(n,N)\to\infty} E_\xi E_p(a'\hat{y}_{QR} - a'y)^2.$$

Result 5. If $a'\hat{y}_{QR}$ is ADU for $a'y$, then

$$V(a'\hat{y}_{QR}) = \sum_{I=1}^{N} a_I^2(\pi_I^{-1} - 1)\sigma_I^2. \qquad (10)$$

Note that the asymptotic variance of $a'\hat{y}_{QR}$ does not depend on the choice of q_I, provided that the estimator is ADU. Alternative choices of q_I give estimators that have similar sampling distributions in large samples. This implies that if a sample design is based on Result 5, it will be applicable to the entire class of robust QR estimators based on the model ξ. This class includes virtually all of the standard survey sampling procedures.

Result 5 can be written more attractively. Define the <u>asymptotic standard error</u>, se, to be $V(a'\hat{y}_{QR})^{1/2}$, and rescale the inclusion probabilities of the sample plan as

$$w_I = \pi_I/\text{mean}(\pi) \qquad (11)$$
$$= N\pi_I/n,$$

where mean(π) denotes the finite population mean

$$N^{-1} \sum_{I=1}^{N} \pi_I.$$

Then (10) becomes

$$se = \frac{N}{\sqrt{n}} \sqrt{\text{mean}(a^2\sigma^2/w) - (n/N)\text{mean}(a^2\sigma^2)}. \qquad (12)$$

Equation (12) also provides qualitative insights that are useful for planning. As is usual in sampling, equation (12) shows that the standard error increases in proportion to the total number (N) of population units, and decreases in proportion to the square root of the sample size. The term "$(n/N)\text{mean}(a^2\sigma^2)$" generalizes the conventional finite population correction factor and is often negligible.

The remaining term in (12), "$\text{mean}(a^2\sigma^2/w)$," reflects the interaction of the parameter of interest, the residual standard deviations assumed in the model ξ, and the inclusion probabilities of the sample plan. As Brewer and Sarndal have both noted, an efficient sampling plan can be developed by choosing the π_I to minimize this term. Indeed the Cauchy-Schwartz inequality implies that

$$\left\{\sum_{I=1}^{N} a_I{}^2\sigma_I{}^2/w_I\right\}\left\{\sum_{I=1}^{N} w_I\right\} \geq \left\{\sum_{I=1}^{N} (|a_I|\sigma_I/\sqrt{w_I})(\sqrt{w_I})\right\}^2 ;$$

or equivalently, that

$$\text{mean}(a^2\sigma^2/w) \geq \text{mean}^2(|a|\sigma).$$

Here the lower bound is achieved if and only if w_I is proportional to $|a_I|\sigma_I$, or specifically if

$$w_I = |a_I|\sigma_I/\text{mean}(|a|\sigma), \text{ and} \tag{13}$$

$$\pi_I = n|a_I|\sigma_I/N \text{ mean}(|a|\sigma).$$

A sampling plan is said to be best for a'y under ξ if and only if it satisfies (13). For a best sampling plan,

$$se = \frac{N}{\sqrt{n}}\sqrt{\text{mean}^2(|a|\sigma)-(n/N)\text{mean}(a^2\sigma^2)}. \tag{14}$$

These results can be summarized in terms of the relevance of each unit in the population to the parameter of interest, where the relevance of unit I is defined to be the quantity $|a_I|\sigma_I$. Then (13) implies that a best sampling plan selects each unit with probability proportional to its relevance. Moreover, (14) shows how to calculate the standard error of a best sampling plan from the distribution of relevance in the finite population.

There are sometimes good reasons to consider a sampling plan that is not best. The efficiency (eff) of any such plan can be defined to be the ratio n_b/n, where n is the sample size required to achieve a certain standard error using the plan under consideration, and n_b is the sample size required to achieve the same standard error with the best sampling plan. Suppose that w_I describes the plan under consideration. Then

$$\text{eff} = \text{mean}^2(|a|\sigma)/\text{mean}(a^2\sigma^2/w). \tag{15}$$

In particular, the efficiency of a simple random sampling plan is $(1 + CV^2)^{-1}$, where CV is the finite population coefficient of variation of $|a_I|\sigma_I$.

Equation (13) can be regarded as a generalization of Neyman allocation for stratified sampling. Consider the special case in which the population parameter of interest is the population total, so that all $a_I = 1$; and suppose strata can be defined such that the σ_I are constant within strata. Then (13) gives the Neyman allocation. Thus, stratification with Neyman allocation is a best sampling plan in this special situation.

More generally, stratification is a very useful technique for developing convenient sampling plans that are highly efficient, i.e., nearly best for any population parameter and any model ξ. In general, relevance, $|a_I|\sigma_I$, can vary almost continuously through the population. However, strata can be constructed so that the units within each stratum are nearly equally relevant. In such a case, a stratified sampling plan based on the mean within-strata relevance is highly efficient. This principle provides a direct model-based method of constructing strata.

To see this precisely, consider any specific stratification of the population into H strata with N_h units in stratum h. Let CV_h be the coefficient of variation of $|a_I|\sigma_I$ within stratum h, so that

$$1 + CV_h^2 = \text{mean}_h(a^2\sigma^2)/\text{mean}_h^2(|a|\sigma). \tag{16}$$

Here mean_h denotes the population mean within stratum h. A stratification is said to be <u>strong</u> if $\varepsilon = \max_{1 \leq h \leq H} (CV_h)$ is small. For a strong stratification, the sample size n_h allocated to each stratum h should be

$$n_h = nN_h \,\text{mean}_h(|a|\sigma)/N \,\text{mean}(|a|\sigma). \tag{17}$$

In other words, with a strong stratification, the sampling fractions n_h/N_h should be proportional to the mean relevance of the units within each stratum; or equivalently, the total sample size should be allocated to each stratum in proportion to the total relevance of the units within each stratum.

Any strong stratification with allocation following (17) will be highly efficient. In fact, (17) implies that for any unit I in stratum h,

$$w_I = \text{mean}_h(|a|\sigma)/\text{mean}(|a|\sigma); \tag{18}$$

then (15) and (16) give

$$\text{eff} = \frac{\text{mean}\ (|a|\sigma)}{N^{-1}\sum_{h=1}^{H} N_h(1+CV_h^2)\text{mean}_h(|a|\sigma)} \tag{19}$$

$$\geq (1+\varepsilon^2)^{-1}.$$

It is often convenient to construct strata with an equal number of sample units in all strata. With the allocation in (17), the n_h will be equal as long as the total relevance of units is constant from stratum to stratum. So a suitable stratification can be constructed by sorting the population in order of relevance, and then dividing the population into H strata containing H^{-1} of the total relevance. For example, to form ten strata, each stratum should contain 10 percent of the total relevance. The efficiency of the design can be made as high as desired by increasing H, but as few as ten strata are often adequate.

4. DERIVATION OF KEY RESULTS

4.1 Notation

The purpose of this section is to provide a derivation of Results 1-5. The analysis builds on Brewer (1979) and Sarndal (1980a).

Additional vector notation is useful. Define $a = [a_1 \ldots a_N]'$, $y = [y_1 \ldots y_N]'$, and $u = [u_1 \ldots u_N]'$. Let $X = [X_1 \ldots X_N]'$, the (Nxk) matrix of auxiliary information. Also, define the following (NxN) diagonal matrices: $\Sigma = \text{diag}(\sigma_I^2)$; $\Pi = \text{diag}(\pi_I)$; $Q = \text{diag}(q_I)$; $R = \text{diag}(r_I)$; and $\Delta = \text{diag}(\delta_I)$, where $\delta_I = 1$ if $I \in s$ and $\delta_I = 0$ if $I \notin s$. In this notation model ξ is

$$y = X\beta + u,\ E_\xi(u) = 0,\ E_\xi(uu') = \Sigma, \text{ assumed known.} \tag{20}$$

A sampling strategy is characterized by the triplet (Π, Q, R). To estimate the population characteristic $a'y$, we use the estimator $a'\hat{y}_{QR}$ with

$$\begin{aligned} \hat{y}_{QR} &= X\hat{\beta} + R\Delta\hat{u}, \\ \hat{\beta} &= (X'Q\Delta X)^{-1} X'Q\Delta y, \text{ and} \\ \hat{u} &= y - X\hat{\beta}. \end{aligned} \tag{21}$$

It is assumed that the sample size $n = \sum_{I=1}^{N} \pi_I$ is fixed, and that $X'Q\Delta X$ is nonsingular for all s with nonzero probability of occurrence.

A substantial advantage of model-based analysis lies in the strong links that are established with linear statistical inference. With the added definitions

$$\begin{aligned} \hat{C} &= X(X'Q\Delta X)^{-1}X'Q, \text{ and} \\ \hat{T} &= \hat{C} + R - R\Delta\hat{C}, \end{aligned} \tag{22}$$

we have $X\hat{\beta} = \hat{C}\Delta y$, $\hat{y}_{QR} = \hat{T}\Delta y$, $\hat{C}\Delta X = X$, and

$$\begin{aligned} I - \hat{T}\Delta &= (I-R\Delta)(I-\hat{C}\Delta) \\ &= (I-\Delta) + (I-\hat{T})\Delta. \end{aligned}$$

The prediction error $a'y - a'\hat{y}_{QR} = a'(I-\hat{T}\Delta)y$ reduces to $a'(I-\hat{T}\Delta)u$ under ξ, since $(I-\hat{C}\Delta)X = 0$. This implies that $a'\hat{y}_{QR}$ is a ξ-unbiased predictor of $a'y$, with the mean squared error

$$\begin{aligned} &a'(I-\hat{T}\Delta)\Sigma(I-\hat{T}'\Delta)a \\ &= a'(I-\Delta)\Sigma(I-\Delta)a + a'(I-\hat{T})\Sigma\Delta(I-\hat{T}')a. \end{aligned} \tag{23}$$

As Royall and others have shown, this is minimized by using a BLU prediction strategy, (Π, Σ^{-1}, I), which is, of course, conditional on the sample s.

4.2 ADU Estimators

In dealing with finite population sampling, care must be exercised in defining the context of asymptotic analysis. Various approaches can be utilized for letting the sizes of the sample and population both increase while the sampling fraction remains more or less fixed. We will follow a formulation introduced by Brewer (1979) and used by Sarndal (1980a).

For the asymptotic analysis, the population of interest is assumed to consist of $N^* = mN$ units composed of m blocks of N units. Each of these m blocks is assumed to have an identical matrix X of auxiliary information. For model-based analysis, (20) is used to generate m independent realizations of the vector y, say y_j, $j = 1,\ldots,m$. However, to make the definition of ADU independent of the model, in this subsection the y_j are assumed to be identical copies of some y.

The vector a is assumed to be identical across blocks so that the population parameter of interest, say $a'y^*$, can be written as $\sum_{j=1}^{m} a'y_j$, or simply $ma'y$ in this subsection. Similarly it is assumed that the strategy (Π,Q,R) is identical across blocks, and in particular that a sample of size $n^* = mn$ is selected with first-order inclusion probabilities following Π within each block. The matrix Δ_j indicates the units in the sample from block j, and the estimator $a'\hat{y}^*$ is formed following (21) as

$$a'\hat{y}^* = \sum_{j=1}^{m} a'\hat{y}_j, \text{ where} \tag{24}$$

$$\hat{y}_j = X\hat{\beta}^* + R\Delta_j\,\hat{u}_j,$$

$$\hat{\beta}^* = \left(\sum_{j=1}^{m} X'Q\Delta_j X\right)^{-1} \sum_{j=1}^{m} X'Q\Delta_j\, y_j, \text{ and}$$

$$\hat{u}_j = y_j - X\hat{\beta}^*.$$

We also define

$$\hat{\Pi}^* = m^{-1} \sum_{j=1}^{m} \Delta_j, \tag{25}$$

$$\hat{C}^* = X(X'Q\hat{\Pi}^*X)^{-1}X'Q, \text{ and}$$

$$\hat{T}^* = \hat{C}^* + R - R\hat{\Pi}^*C^*,$$

so that

$$a'\hat{y}^* = \sum_{j=1}^{m} a'\hat{T}^*\Delta_j y_j = ma'\hat{T}^*\hat{\Pi}^*y \quad \text{if the } y_j \text{ are identical.}$$

With this formulation, the strategy (Π,Q,R) is said to be ADU for the characteristic a if and only if for all y, $E_p(a'\hat{y}^*/m)$ converges to a'y as m increases to infinity. As m increases, $\hat{\Pi}^*$ converges almost surely to Π. The assumption that $X'Q\Delta X$ is nonsingular for all samples that can occur under Π implies that $\hat{C}^*$ is bounded and converges almost surely to $C = X(X'Q\Pi X)^{-1}X'Q$. So $\hat{T}^*$ converges almost surely to $T = C + R - R\Pi C$ and $E_d(a'\hat{y}^*/m)$ converges to $a'T\Pi y$. Thus,

<u>Lemma</u>. The strategy (Π,Q,R) is asymptotically design unbiased (ADU) for the characteristic a if and only if $a'(I-T\Pi)y = 0$ for all $y \in R^N$.

It is helpful to note that in its derivation, this lemma describes each block of the population, but once the derivation is complete, the lemma can be considered to describe the entire population of interest.

An immediate consequence of the lemma is that for any strategy (Π,Q,R) that is ADU for a, $\pi_I = 0$ implies $a_I = 0$. Any unit with both $\pi_I = 0$ and $a_I = 0$ is clearly irrelevant and can be eliminated from the population. Because we are primarily interested in ADU strategies, it is assumed henceforth that $\Pi > 0$.

An algebraic characterization of ADU strategies can be developed from the identity $I-T\Pi = (I-R\Pi)(I-C\Pi)$. Suppose initially that $Q > 0$, so that $Q\Pi$ defines an inner product over R^N. In this case $C\Pi = X(X'Q\Pi X)^{-1}X'Q\Pi$

is the orthogonal projector onto the linear manifold M(X) spanned by the column vectors of X, and I-CΠ is the projector onto the linear manifold orthogonal to M(X) with respect to the inner product QΠ.

Since

$$a'(I-T\Pi)y = a'(I-R\Pi)(Q\Pi)^{-1}Q\Pi(I-C\Pi)y,$$

(Π,Q,R) is ADU for a if and only if $(Q\Pi)^{-1}(I-R\Pi)a \in M(X)$, or equivalently, $(I-R\Pi)a = Q\Pi x$ for some $x \in M(X)$. The restriction $Q > 0$ can easily be relaxed, giving

<u>Theorem 1</u>. A strategy (Π,Q,R) is ADU for the characteristic a if and only if $(I-R\Pi)a = Q\Pi x$ for some $x \in M(X)$.

This gives Results 1 and 3 of Section 2.

Two strategies, (Π,Q_1,R_1) and (Π,Q_2,R_2), are said to be <u>equivalent for a</u> if and only if they produce identical estimates of a'y for all y and all samples with positive probability of occurrence. Using (21) and (22), two strategies that employ identical Q are equivalent if and only if

$$a'(R_1 - R_2)\Delta\hat{u} = a'(R_1 - R_2)\Delta(I - \hat{C}\Delta)y$$
$$= 0 \text{ for all s and all y.}$$

Using an argument similar to the proof of Theorem 1, this is true if and only if $(R_1-R_2)a = Qx$ for some $x \in M(X)$. However, Theorem 1 shows that a strategy (Π,Q,R) is ADU for a if and only if $(\Pi^{-1}-R)a = Qx$, $x \in M(X)$. This proves

<u>Theorem 2</u>. A strategy (Π,Q,R) is ADU for a if and only if (Π,Q,R) and the generalized regression strategy (Π,Q,Π^{-1}) are equivalent for a.

This gives Results 2 and 4 of Section 2.

4.3 Efficiency of ADU Strategies

Within the class of ADU strategies, a useful planning criterion is the asymptotic variance of $a'\hat{y}_{QR}$, denoted $V(a'\hat{y}_{QR})$. Here $V(a'\hat{y}_{QR})$ is defined to be the asymptotic expectation, with respect to both design and model, of the mean square prediction error of $a'\hat{y}_{QR}$. To develop the asymptotic analysis we must return to the assumption that the population comprises m blocks, as in the previous subsection, but with y_j independently generated within each block following (20). In this case, there are m independent u_j with $E_\xi(u_j) = 0$ and $E_\xi(u_j u_j') = \Sigma$, $j = 1,\ldots,m$. To examine the square error $(a'y^*-a'\hat{y}^*)^2$, use (25) to note that

$$\sum_{j=1}^{m} y_j - \hat{y}_j = \sum_{j=1}^{m} (I-\hat{T}^*\Delta_j)u_j, \text{ since}$$

$$\begin{aligned}\sum_{j=1}^{m} (I-\hat{T}^*\Delta_j)X &= m(I-\hat{T}^*\hat{\Pi}^*)X \\ &= m(I-R\hat{\Pi}^*)(I-\hat{C}^*\hat{\Pi}^*)X \\ &= 0.\end{aligned}$$

A derivation similar to that of (23) implies

$$\begin{aligned}m^{-1}E_\xi(a'y^*-a'\hat{y}^*)^2 &= m^{-1} E_\xi[\sum_{j=1}^{m} a'(y_j-\hat{y}_j)]^2 \\ &= m^{-1}\sum_{j=1}^{m} a'(I-\hat{T}^*\Delta_j)\ \Sigma\ (I-\hat{T}^{*\prime}\Delta_j)a \\ &= a'(I-\hat{\Pi}^*)\Sigma a + a'(I-\hat{T}^*)\hat{\Pi}^*\Sigma(I-\hat{T}^{*\prime})a.\end{aligned}$$

Now the asymptotic design-based expectation can be evaluated as in the previous subsection, giving

$$\lim_{m\to\infty} m^{-1}E_d E_\xi(a'y^*-a'\hat{y}^*)^2 = a'(I-\Pi)\Sigma a + a'(I-T)\Pi\Sigma(I-T')a. \tag{26}$$

Given that (Π,Q,R) is ADU for a, $a'T\Pi y = a'y$ for all $y \in R^N$; so (26) simplifies to $a'(\Pi^{-1}-I)\Sigma a$.

Note that this expression represents a summation over a single block, so that the corresponding summation over the entire population is m times larger. Thus, in terms of the entire population we have

Theorem 3. If (Π,Q,R) is ADU for the characteristic a, then the asymptotic variance of (Π,Q,R) for a is

$$\begin{aligned} V(a'\hat{y}) &= a'(\Pi^{-1}-I)\Sigma a \\ &= \sum_{I=1}^{N} a_I^2(\pi_I^{-1} - 1)\sigma_I^2. \end{aligned} \tag{27}$$

This gives Result 5 of Section 3.

Theorem 3 has been proven previously for $a_I = 1$ with specific choices of q_I and r_I. Brewer (1979) considered the case $k = 1$, $q_I = (\pi_I x_I)^{-1}(1 - \pi_I)$, and $r_I = 1$, as discussed in Section 2.1. Sarndal (1981a) obtained the result for $q_I = \pi_I^{-1}$, σ_I^{-2}, and $\pi_I^{-1}\sigma_I^{-2}$. Isaki and Fuller (1982) obtained (27) for $q_I = \pi_I^{-2}$ when π_I follows (13), using a somewhat different asymptotic argument.

REFERENCES

Anderson, D. W., L. Kish and R. G. Cornell, (1980), "On Stratification, Grouping, and Matching," Scandinavian Journal of Statistics, 7, 61-66.

Brewer, K. R. W., (1963), "Ratio Estimation in Finite Populations: Some Results Deductible from the Assumption of an Underlying Stochastic Process," Australian Journal of Statistics, 5, 93-105.

_____, (1979), "A Class of Robust Sampling Designs for Large-Scale Surveys," Journal of the American Statistical Association, 74, 911-915.

Cassel, C. M., C. E. Sarndal and J. H. Wretman, (1976), "Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations," Biometrika, 63, 615-620.

_____, _____, _____, (1977), Foundation of Inference in Survey Sampling, New York: John Wiley & Sons.

Cochran, W. G., (1961), "Comparison of Methods for Determining Stratum Boundaries," Bulletin of the International Statistical Institute, 38, 345-358.

Dalenius, T. and J. L. Hodges, Jr., (1959), "Minimum Variance Stratification," Journal of the American Statistical Association, 54, 88-101.

Hajek, J., (1971), "Comment On an Essay by D. Basu," in Foundations of Statistical Inference, eds. V. P. Godambe and D. A. Sprott. Toronto: Holt, Rinehart & Winston.

Hansen, M. H., W. G. Madow and B. J. Tepping, (1978), "On Inference and Estimation from Sample Surveys," Proceedings of the Survey Research Section, American Statistical Association, 82-107.

Harvey, A. C., (1976), "Estimating Regression Models with Multiplicative Heteroscedasticity," Econometrica, 44, 461-464.

Isaki, C. T. and W. A. Fuller, (1982), "Survey Design under a Regression Superpopulation Model," Journal of the American Statistical Association, 77, 89-96.

Rao, T. J., (1977), "Optimum Allocation of Sample Size and Prior Distributions: A Review," International Statistical Review, 45, 173-179.

Royall, R. M., (1970), "On Finite Population Sampling Theory under Certain Linear Regression Models," Biometrika, 57, 377-387.

_____, (1971), "Linear Regression Models in Finite Population Sampling Theory," in Foundations of Statistical Inference, eds. V. P. Godambe and D. A. Sprott. Toronto: Holt, Rinehart & Winston.

_____, (1976), "The Linear Least Squares Prediction Approach to Two-Stage Sampling," Journal of the American Statistical Association, 71, 657-664.

_____ and J. Herson, (1973a), "Robust Estimation in Finite Populations," Journal of the American Statistical Association, 68, 880-889.

_____ and _____, (1973b), "Robust Estimation in Finite Population, II: Stratification on a Size Variables," Journal of the American Statistical Association, 68, 891-893.

Sarndal, C. E., (1980a), "On π-Inverse Weighting versus Best Linear Unbiased Weighting in Probability Sampling," Biometrika, 67, 639-650.

_____, (1980b), "A Two-way Classification of Regression Estimation Strategies in Probability Sampling," The Canadian Journal of Statistics, 8, 165-177.

Singh, R. (1975), "On Optimal Stratification for Proportional Allocation," Sankhya, 37, C, Pt. 1, 109-115.

Smith, T. M. F., (1976), "The Foundations of Survey Sampling, A Review," Journal of the Royal Statistical Society, A, 139, Part 2, 183-204.

Table 1. Various Estimators Proposed for the Ratio Model

a. Horvitz-Thompson Ratio (Hajek, 1971)

$$\hat{Y}_{HTR} = \hat{\beta}_{HT}\, X, \text{ where } \hat{\beta}_{HT} = \hat{Y}_{HT}/\hat{X}_{HT},$$

$$\text{with } \hat{Y}_{HT} = \sum_{I \varepsilon s} \pi_I^{-1} y_I \quad \text{and } X = \sum_{I=I}^{N} x_I.$$

b. Combined Regression through the Origin

$$\hat{Y}_{CR} = \hat{Y}_{HT} + \hat{\beta}_{CR}\,(X - \hat{X}_{HT}),$$

$$\text{with } \hat{\beta}_{CR} = \sum_{I \varepsilon s} \pi_I^{-1} x_I y_I \Big/ \sum_{I \varepsilon s} \pi_I^{-1} x_I^2.$$

c. Best Linear Unbiased (Royall, 1970)

$$\hat{Y}_{BLU} = \sum_{I \varepsilon s} y_I + \hat{\beta}_{BLU} \sum_{I \notin s} x_I,$$

$$\text{with } \hat{\beta}_{BLU} = \sum_{I \varepsilon s} \sigma_I^{-2}\, x_I y_I \Big/ \sum_{I \varepsilon s} \sigma_I^{-2} x_I^2.$$

d. Generalized Regression (Cassel, Sarndal, and Wretman, 1976)

$$\hat{Y}_{GR} = \hat{Y}_{HT} + \hat{\beta}_{BLU}\,(X - \hat{X}_{HT}).$$

e. Another Robust Estimator (Brewer, 1979)

$$\hat{Y}_{RO} = \sum_{I \varepsilon s} y_I + \hat{\beta}_{RO} \sum_{I \notin s} x_I,$$

$$\text{where } \hat{\beta}_{RO} = \sum_{I \varepsilon s} q_I x_I y_I \Big/ \sum_{I \varepsilon s} q_I x_I^2,$$

$$\text{with } q_I = (\pi_I x_I)^{-1}(1-\pi_I).$$

Table 2. Choice of q_I and r_I for the Estimators in Table 1

Estimator		q_I	r_I
a.	$\hat{Y}_{HTR}$	$(\pi_I x_I)^{-1}$	0
b.	$\hat{Y}_{CR}$	π_I^{-1}	π_I^{-1}
c.	$\hat{Y}_{BLU}$	σ_I^{-2}	1
d.	$\hat{Y}_{GR}$	σ_I^{-2}	π_I^{-1}
e.	$\hat{Y}_{RO}$	$(\pi_I x_I)^{-1}(1-\pi_I)$	1