

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

August 1982

ROBUST ESTIMATION IN SURVEY SAMPLING
USING MULTIVARIATE AUXILIARY INFORMATION

Working Paper No. 309

Roger L. Wright
The University of Michigan

Carl-Erik Sarndal
The University of Montreal and
The National Central Bureau of Statistics
of Sweden

FOR DISCUSSION PURPOSES ONLY

None of this material is to be quoted or
reproduced without the expressed permission
of the Division of Research.

ABSTRACT

This paper considers multivariate regression estimators for finite population totals and other linear functions. Estimators are identified that are asymptotically design unbiased and preserve certain aspects of the model-based prediction approach. The design-based variance of these estimators is approximated using a generalization of the Yates-Grundy expression. The approach is illustrated in an electric power load research application.

KEY WORDS: Generalized regression, Prediction approach, Asymptotic design unbiased, Design-based inference, Model-based inference

1. ALTERNATIVE FORMS OF ASYMPTOTICALLY DESIGN UNBIASED ESTIMATORS

Suppose we want to estimate the linear function $T = \sum_1^N a_I Y_I$ of the character Y for a finite population $U = \{1, \dots, I, \dots, N\}$, where a_I is a known constant and Y_I is the unknown Y -value attached to the unit labelled I . The population total ($a_I = 1$, all I) or mean ($a_I = N^{-1}$, all I) are important special cases of T . The Y_I are observed for $I \in s$, where $s \subset U$ denotes a sample. Several estimation approaches exist that exploit (multivariate) auxiliary information. These can be divided into two main groups, those that appeal to principles of design inference (DESI) and those that appeal to principles of model inference (MODI).

In DESI, it is the randomized selection of the sample s , according to a given sampling design, $p(s)$, that is the source of the inference. Some recent contributions in this trend are Brewer (1979), Cassel, Sarndal and Wretman (1976), Isaki and Fuller (1982), Sarndal (1980a,b), Wright (1982).

In DESI, estimators are usually required to be asymptotically design unbiased (ADU; see below) and the quality of the estimate (given the design) is measured by the design variance, $V_p(\hat{T})$. For confidence intervals, an approximately design unbiased variance estimate $\hat{V}_p(\hat{T})$ is created; the resulting interval $\hat{T} \pm z_{1-\alpha/2} \{\hat{V}_p(\hat{T})\}^{1/2}$ has the property of covering the true point T for roughly $100(1-\alpha)\%$ of all samples s drawn repeatedly under the specified sampling design.

By contrast, in MODI the argument is conditional on s and makes reference to an assumed model ξ . If an estimator is model unbiased, its quality is measured by its model variance, $E_\xi(\hat{T} - T)^2 = V_\xi(\hat{T})$. Having found an

estimated model variance, one can construct MODI confidence intervals that are interpreted via repeated realizations of the model. Usually ξ is a linear model stating that $\underline{Y} = \underline{X}\underline{\beta} + \underline{e}$ with $E_{\xi}(\underline{e}) = 0$, $V_{\xi}(\underline{e}) = \sigma^2 \underline{I}$. This point of view has been developed in a series of papers by Royall and coworkers (e.g., Royall (1970), Royall and Herson (1973), Royall and Cumberland (1981)).

We favor the DESI outlook, and our paper concentrates on that approach. It is nevertheless fitting to observe: (a) that the estimators proposed by DESI as well as by MODI advocates can be represented within a single class, called the QR class, (Wright, 1982), and (b) that some DESI writers (e.g., Brewer, 1979) are attracted by the particular form of the usual MODI estimator, yet manage to conform to DESI requirements, including ADU-ness, by an appropriate choice of "free constants."

The QR class of estimators of $T = \sum_1^N a_I Y_I$ is given by

$$\hat{T} = \sum_1^N a_I \hat{Y}_I + \sum_S r_I a_I e_I \quad (1.1)$$

where $\hat{Y}_I = \underline{x}_I' \hat{\underline{\beta}}$,

$$e_I = Y_I - \hat{Y}_I, \quad (1.2)$$

$$\hat{\underline{\beta}} = (\sum_S q_I \underline{x}_I \underline{x}_I')^{-1} \sum_S q_I \underline{x}_I Y_I, \quad (1.3)$$

and $\underline{x}_1, \dots, \underline{x}_N$ are known auxiliary p-vectors. The constants r_I and $q_I > 0$ are open to choice. The typical estimator \hat{T} is evidently suggested by the model ξ such that

$$E_{\xi}(Y_I) = \underline{x}_I' \underline{\beta}. \quad (1.4)$$

In addition, we assume under ξ that

$$V_{\xi}(Y_I) = \sigma_I^2$$

and that the N random variables Y_I are mutually independent. The σ_I^2 are assumed known except for a multiplicative constant.

The QR class contains three important subclasses (Sarndal, 1980b) defined by three simple choices of the r_I :

Simple projection estimators (SPRO) defined by setting $r_I = 0$ for all I , yielding

$$\hat{T} = \sum_1^N a_I \hat{Y}_I, \quad (1.5)$$

Linear prediction estimators (LPRE) defined by letting $r_I = 1$ for all I , yielding

$$\hat{T} = \sum_s a_I Y_I + \sum_{\bar{s}} a_I \hat{Y}_I \quad (1.6)$$

where $\bar{s} = U - s$, and

Generalized regression estimators (GREG) defined by setting $r_I = \pi_I^{-1}$, yielding

$$\hat{T} = \sum_1^N a_I \hat{Y}_I + \sum_s a_I e_I / \pi_I. \quad (1.7)$$

Here π_I is the inclusion probability $P(I \in s)$, assumed to be nonzero.

We argue that the statistician (and his client) may feel attracted to a certain subclass on

- (1) theoretical grounds,
- (2) cosmetic grounds.

By theoretical grounds we mean the basic properties of the estimator. These differ in DESI, with its ADU-ness requirement, and in MODI, where model-based properties are appealed to. By cosmetic grounds we mean that the form of the estimator matters; many may feel, for example, that the LPRE form is both simple and realistic, being composed of a sum of observed Y 's plus a sum of predicted Y 's for nonsampled units. The simplicity of the SPRO form may

also appeal to many. Although the GREG form is not complicated, let us for sake of argument consider that the SPRO and LPRE subclasses are "cosmetic."

Returning to theoretical considerations, we note that the LPRE subclass includes the best linear unbiased (BLU) prediction estimator based on the MODI approach under ξ , (1.4). The BLU-LPRE estimator is obtained by letting $r_I = 1$ and $q_I = \sigma_I^{-2}$ for all I. Although the BLU-LPRE estimator has desirable MODI properties under ξ , it has been shown to be vulnerable to serious bias when the assumed model is inaccurate, (e.g., Hansen, Madow, and Tepping, 1978).

Under the DESI approach, an important basic property is asymptotic design unbiasedness. An estimator \hat{T} is said to be ADU for $T = \sum_1^N a_I Y_I$ if

$$\lim_{(n,N) \rightarrow \infty} [E_p(\hat{T}) - T] = 0.$$

DESI writers consider ADU-ness desirable for reasons of robustness. Whether the model (1.4) is true or not, an ADU estimator is correct "on the average" in large samples drawn repeatedly from proportionately larger populations. We do not enter into detail here about the limit argument. Somewhat different versions are found in Brewer (1979), Isaki and Fuller (1982) Sarndal (1980a), and Wright (1982), although the differences do not seem to be consequential.

The latter author shows that within the QR class, \hat{T} is ADU for T if and only if there exists a p-vector $\underline{\lambda}$ such that

$$a_I(1 - r_I \pi_I) = \pi_I q_I x_I' \underline{\lambda} \quad \text{for } I = 1, \dots, N. \quad (1.8)$$

Several important conclusions follow from (1.8): (a) The GREG estimator, being defined by the choice $r_I = \pi_I^{-1}$, is ADU regardless of the choice of q_I . (b) We can have ADU-ness (thus DESI) and cosmetic form (SPRO or LPRE) at the same time by choosing the q_I to satisfy (1.8) for a particular $\underline{\lambda}$. (c) Let us define $\underline{u} = (u_1, \dots, u_N)'$ with

$$u_I = a_I(\pi_I^{-1} - r_I)q_I^{-1}; \quad I = 1, \dots, N. \quad (1.9)$$

Then \hat{T} is ADU if and only if $\underline{u} \in M(\underline{X})$, where $M(\underline{X})$ is the linear manifold spanned by the columns of $\underline{X} = (\underline{x}_1, \dots, \underline{x}_N)'$.

Conclusion (c) implies that we can provide ADU-ness for any choice of q_I by expanding \underline{X} . If, for a predetermined choice of r_I and q_I , $\underline{u} \notin M(\underline{X})$, we can augment \underline{X} by the column \underline{u} . Then surely $\underline{u} \in M(\underline{X}_A)$, where \underline{X}_A is \underline{X} augmented by the column \underline{u} . This may in fact lead to an estimator with a smaller variance, since \underline{u} is an additional explanatory variable.

An important additional result for all ADU estimators within the QR class is the following (Sarndal, 1980a; Wright, 1982): If \hat{T} given by (1.1) is ADU, then its asymptotic expected variance,

$$\lim_{(n,N) \rightarrow \infty} E_{\xi} E_P (\hat{T} - T)^2,$$

is given by

$$\sum_1^N a_I^2 (\pi_I^{-1} - 1) \sigma_I^2. \quad (1.10)$$

This expression does not depend on the choice of r_I and q_I , provided these quantities satisfy the ADU requirement (1.8).

This result is important for planning a survey. The expression (1.10) is easily seen to be minimized by a design whose inclusion probabilities are given by

$$\pi_I = n |a_I| \sigma_I / \sum_1^N |a_I| \sigma_I. \quad (1.11)$$

Wright (1982) has suggested a procedure that approximates the optimum. It involves estimating the σ_I and stratifying the population into strata of units I with similar values of $|a_I| \hat{\sigma}_I$, where $\hat{\sigma}_I$ is the estimate of σ_I .

2. ESTIMATING THE POPULATION TOTAL

In this section we consider the estimation of the population total $\sum_{I=1}^N Y_I$ and comment on the ADU proposals already in the literature.

GREG form. As already observed, the GREG estimator is ADU, whatever the choice of q_I . The term GREG was first used by Cassel, Sarndal and Wretman (1976) who considered the GREG estimator arising when $q_I = \sigma_I^{-2}$ under the simple ratio model such that $E_{\xi}(Y_I) = \beta x_I$ and $V_{\xi}(Y_I) = \sigma_I^2 = \sigma^2 x_I$. Sarndal (1980a,b) also studied the GREG form with $q_I = \pi_I^{-1}$ and $q_I = \pi_I^{-1} \sigma_I^2$, for the general linear model (1.4). He notes (1980a) that GREG takes on the cosmetic SPRO form if there exist λ such that

$$\sigma_I^2 = x_I' \lambda; \quad I = 1, \dots, N.$$

This is the case for any homoscedastic model including an intercept, and for any model (with or without intercept) if $\sigma_I^2 \propto x_{Ij}$ for some variable x_j in the model. The GREG form was extended for use in situations with nonresponse by Cassel, Sarndal and Wretman (1979).

What q_I should be chosen for the GREG? One may study the question with reference to second order efficiency (in terms of expected variance), but no definite answer is available.

LPRE and SPRO forms. Brewer (1979) argued that it makes good sense to require the LPRE form (1.6) and ADU-ness. Working under the simple ratio model $E_{\xi}(Y_I) = \beta x_I$, he consequently posed the question: Can one obtain an ADU estimator of the form (1.6)? His answer, in the affirmative, was that this requires that

$$q_I \propto (\pi_I^{-1} - 1) x_I^{-1}; \quad I = 1, \dots, N$$

a choice which clearly satisfies the general condition (1.8).

If we extend Brewer's wish to maintain ADU-ness and the LPRE form to a general linear model $E_{\xi}(Y_I) = \underline{x}_I' \underline{\beta}$, we find that the ADU condition (1.8) is satisfied if we fix any constant vector $\underline{\lambda}$ such that $\underline{x}_I' \underline{\lambda} > 0$ for all I and take

$$q_I \propto (\pi_I^{-1} - 1) / \underline{x}_I' \underline{\lambda}; \quad I = 1, \dots, N. \quad (2.1)$$

Similarly, we see that the SPRO is ADU if we fix any $\underline{\lambda}$ and take

$$q_I \propto \pi_I^{-1} / \underline{x}_I' \underline{\lambda}; \quad I = 1, \dots, N. \quad (2.2)$$

Since any choice of $\underline{\lambda}$ in (2.1) or (2.2) will produce ADU-ness, we may legitimately ask: What $\underline{\lambda}$ should be chosen? This is an open question, just as the choice of q_I in the GREG case mentioned above.

The augmented X-matrix argument. We observed in the previous section that, having fixed the r_I and the q_I , we can always achieve ADU-ness simply by augmenting the \underline{X} -matrix. We draw the consequences of this argument for the SPRO and LPRE forms, where ADU-ness is not "automatic," as it is for GREG.

It follows from (1.9) that the SPRO is ADU if we augment the \underline{X} matrix by a single column $\underline{u}_1 = (u_{11}, \dots, u_{N1})'$, where

$$u_{I1} = \pi_I^{-1} q_I^{-1}, \quad I = 1, \dots, N \quad (2.3)$$

provided \underline{u}_1 is not already in the linear manifold $M(\underline{X})$ spanned by the columns of \underline{X} .

By the same argument, the LPRE is ADU if we augment the \underline{X} matrix by the single column $\underline{u}_2 = (u_{12}, \dots, u_{N2})'$ where

$$u_{I2} = (\pi_I^{-1} - 1) q_I^{-1}, \quad I = 1, \dots, N$$

provided \underline{u}_2 is not already in $M(\underline{X})$.

It also follows that we can have an ADU estimator expressible in either SPRO or LPRE form if \underline{X} is augmented by the two (non-identical) columns \underline{u}_1 and

$\underline{u}_3 = (u_{13}, \dots, u_{N3})'$, where

$$u_{I3} = q_I^{-1}, \quad I = 1, \dots, N \quad (2.4)$$

provided neither \underline{u}_1 nor \underline{u}_3 is already in $M(\underline{X})$.

Isaki and Fuller (1982) were concerned in particular with the SPRO estimator together with the choice $q_I \propto \pi_I^{-2}$, where it is understood that the π_I have already been determined optimally according to (1.11), that is

$$\pi_I = n\sigma_I / \sum_{I=1}^N \sigma_I.$$

In agreement with (2.3) they observed that the SPRO is ADU if the \underline{X} -matrix is augmented by the column $(\pi_1, \dots, \pi_N)'$. Moreover, confirming (2.4) they observed that the resulting estimator also has LPRE form if we further add the column $(\pi_1^2, \dots, \pi_N^2)'$.

3. VARIANCE ESTIMATION

In this section, we are concerned with estimating the design variance $V_p(\hat{T})$ of any ADU estimator \hat{T} in the QR class. Here T is the linear function $\sum_{I=1}^N a_I y_I$. That is, we seek $\hat{V}_p(\hat{T})$ such that we can construct a confidence interval

$$\hat{T} \pm z_{1-\alpha/2} \{\hat{V}_p(\hat{T})\}^{1/2} \quad (3.1)$$

that carries the repeated sampling interpretation.

We start by considering the GREG subclass. Several important results concerning the GREG form follow from Sarndal (1982). Consider the estimate $\hat{\beta}$ given by (1.2) for a certain fixed choice of q_I . Let

$$\underline{B} = \left(\sum_1^N q_I \pi_I \underline{x}_I \underline{x}_I' \right)^{-1} \sum_1^N q_I \pi_I \underline{x}_I Y_I$$

be the expression to which $\hat{\underline{\beta}}$ converges in design probability as $(n, N) \rightarrow \infty$, and define "residuals from the census fit" by

$$\epsilon_I = Y_I - \underline{x}_I' \underline{B}.$$

It can then be shown by Taylor expansion techniques that the design variance of \hat{T} is approximately

$$V_p(\hat{T}) = \sum_{I < J}^N \sum_{I < J}^N \pi_{IJ} g_{IJ} \left[\frac{a_I \epsilon_I}{\pi_I} - \frac{a_J \epsilon_J}{\pi_J} \right]^2 \quad (3.2)$$

where

$$\pi_{IJ} = P(I \in s \text{ and } J \in s), \text{ and}$$

$$g_{IJ} = (\pi_I \pi_J - \pi_{IJ}) \pi_{IJ}^{-1} \text{ for } \pi_{IJ} \neq 0,$$

$$= 0 \text{ otherwise.}$$

We recognize (3.2) as being the variance of the Horvitz-Thompson estimator in $a_I \epsilon_I$, $\sum_s a_I \epsilon_I / \pi_I$, which is design unbiased for the total $\sum_1^N a_I \epsilon_I$.

The result (3.2) comes as no great surprise, for the estimation error of the GREG form can be expressed as

$$\hat{T} - T = \sum_s a_I e_I / \pi_I - \sum_1^N a_I e_I,$$

where e_k given by (1.2) is the sample proxy for ϵ_I . As an estimate of (3.2), the Yates-Grundy expression in $a_I \epsilon_I$ suggests itself. However, since ϵ_I is unknown, we replace it by the computable residual e_I , thus arriving at

$$\hat{V}_p(\hat{T}) = \sum_{\substack{I < J \\ I, J \in s}} g_{IJ} \left[\frac{a_I e_I}{\pi_I} - \frac{a_J e_J}{\pi_J} \right]^2 \quad (3.3)$$

Monte Carlo experiments for a variety of situations have shown that the resulting confidence interval (3.1) gives a coverage rate in repeated samples that approximates the intended $100(1-\alpha)\%$ rather well, even with rather modest sample sizes. There is a clear tendency, however, for the actual coverage rate to fall on the short side of $100(1-\alpha)\%$, indicating that the normal score $z_{1-\alpha/2}$ produces intervals that are too short on the average.

Sarndal (1982) considered the case where $q_I = \pi_I^{-1} \sigma_I^{-2}$, but his argument carries through for any q_I . The variance (3.2) and its estimate (3.3) will of course depend on the q_I actually chosen, and additional work is needed to settle the question of a "best" choice.

We now extend these results to any ADU estimator in the QR class. Wright (1982) showed that the general estimator \hat{T} given by (1.1), for a fixed choice of the r_I and the q_I , is ADU if and only if it is identical to the GREG estimator for the same q_I . In other words, for a fixed choice of the r_I , when \hat{T} given by (1.1) is ADU, it is numerically equal, for every sample s and for any Y_I -values, to the GREG with the same q_I . But we have just given a variance estimation procedure for the GREG with arbitrary q_I . It follows that we have a procedure for variance estimation for any \hat{T} that is ADU.

Example. Brewer's (1979) estimator of $\sum_1^N Y_k$ for the model $E_\xi(Y_k) = \beta x_k$ is ADU and of the LPRE form. It can be written as

$$\hat{T} = \left(\sum_1^N x_I \right) \hat{\beta} + \sum_S e_I \quad (3.4)$$

where $e_I = Y_k - \hat{\beta} x_k$ and

$$\hat{\beta} = \left\{ \sum_S (\pi_I^{-1} - 1) Y_I \right\} / \left\{ \sum_S (\pi_I^{-1} - 1) x_I \right\}$$

As is easily checked, and consistent with the result of Wright, an expression for \hat{T} that is equivalent for all s and all $\underline{y} = (Y_1, \dots, Y_N)'$ is

$$\hat{T} = (\sum_1^N x_I) \hat{\beta} + \sum_s e_I / \pi_I.$$

But this is of the GREG form. The approximate design variance of Brewer's \hat{T} (3.4) is therefore given by (3.2) with all $a_I = 1$ and $\epsilon_I = Y_I - Bx_I$,

where

$$B = \{ \sum_1^N (1 - \pi_I) Y_I \} / \{ \sum_1^N (1 - \pi_I) x_I \}$$

The estimated design variance is given by (3.3) with $a_I = 1$.

4. AN ILLUSTRATION

The preceding estimation techniques can be illustrated in an application drawn from electric utility load research. Many large power companies estimate the time-of-day usage of electricity within various populations of customers through special metering of samples of customers. These estimates are used in rate analysis and system planning. Brandenburg and Higgins (1974) describe the application and the standard methodology. Wright (1981) has introduced some of the techniques discussed in this paper. In these load research studies, the target variable Y_I is often a system peak demand, which is the use of electricity by customer I during a prespecified period of high system-wide consumption of power. In our example, Y_I measures usage during twelve monthly hours of system peak consumption throughout a specific year.

A rich data base of auxiliary information is available in the administrative customer records. One important auxiliary variable, x_{I1} is the customer's total usage of electricity during the year. For certain commercial and industrial customers, another variable, x_{I2} called maximum non-coincident demand (MND) can be extracted from billing information. In our

example, maximum noncoincident demand is available throughout the target population; it measures each customer's peak half-hour rate of electricity use during the year.

Our sample consists of 175 customers selected from a population of 20,600 customers, following the stratified sampling plan shown in Table 1. Table 1 also shows the population totals of x_{I1} and x_{I2} that are used in the example.

Table 1 about here

Two separate analyses have been performed, one using x_{I1} to estimate the population total $T = \sum_1^N Y_I$, and the other using both x_{I1} and x_{I2} to estimate T . In each case, we have calculated various estimates that satisfy the ADU condition (1.8) either through suitable choice of q_I , as in (2.1) and (2.2), or by augmenting the X-matrix as in (2.3) and (2.4). In several of these procedures we have used an estimate of σ_I based on a multiplicative heteroscedasticity model of Harvey (1976).

Under Model One, we assume

$$E_{\xi}(Y_I) = \beta x_{I1}, \tag{4.1}$$
$$\sigma_I = \sigma_0 x_{I1}^{\gamma}.$$

Given our sample and the assumption of normality, we used seven iterations of the Harvey algorithm to calculate the maximum likelihood estimates

$$\hat{Y}_I = 2.042 * 10^{-3} x_{I1}, \tag{4.2}$$
$$\hat{\sigma}_I = 3.378 * 10^{-3} x_{I1}^{.8703}.$$

Table 2 shows various estimates of T using (1.5) - (1.7) with several choices of q_I . Line (1) of Table 2 follows (2.2) while line (2) uses (2.1). The choice of q_I in line (3) yields the best linear unbiased LPRE estimator

conditional on the model (4.1) and $\hat{\sigma}_I$ from (4.2). The choice of q_I in line (4) is suggested in Sarndal (1980a).

Lines (5) - (7) are obtained following various augmentation techniques. In particular, we augment the X -matrix by u_{I1} , u_{I2} , or both u_{I1} and u_{I3} , using (2.3) - (2.4) with $q_I = \hat{\sigma}_I^{-2}$ as in the BLU approach. For illustrative purposes the population totals of u_{I1} , u_{I2} and u_{I3} are assumed to be $142.6 * 10^9$, $141.9 * 10^9$, and $0.7 * 10^9$ respectively. In practice it is not always easy to obtain the population totals of these artificial variables.

Table 2 about here

Note that the GREG estimates in Table 2 are all quite similar. Of course they are identical to the corresponding SPRO or LPRE estimates whenever the latter are ADU. However the SPRO and LPRE estimates in line (3) are somewhat lower than the ADU estimates.

The last column of Table 2 shows the standard errors of the GREG estimates, calculated using (3.3). These standard errors are also applicable to the SPRO and LPRE estimates when the latter are ADU. As expected there is little variation among these standard errors.

Under Model Two, we assume

$$E_{\xi}(Y_I) = \beta_0 + \beta_1 x_{I1} + \beta_2 x_{I2}, \tag{4.3}$$

$$\sigma_I = \sigma_0 \begin{matrix} \gamma_1 & \gamma_2 \\ x_{I1} & x_{I2} \end{matrix},$$

and we obtain the maximum likelihood estimates

$$\begin{aligned}\hat{Y}_I &= -13.15 + 7.319 * 10^{-3} x_{I1} + 5.182 x_{I2}, \\ \hat{\sigma}_I &= 1.825 * 10^5 x_{I1}^{-1.369} x_{I2}^{2.237}.\end{aligned}\tag{4.4}$$

Table 3 shows several estimates of T based on this model. In rows (1) and (2), q_I is chosen following (2.2) and (2.1) respectively, with $\lambda = (1 \ 0 \ 0)'$. The remaining rows are analagous to the corresponding rows of Table 2. To calculate the estimates in rows (5) - (7), we have assumed that the population totals of u_{I1} , u_{I2} and u_{I3} are now $70.1 * 10^9$, $69.6 * 10^9$, and $0.5 * 10^9$ respectively.

Table 3 about here

The GREG estimates in Table 3 are very similar, displaying substantially smaller differences than in Table 2. Again, whenever the SPRO and LPRE are ADU, they are identical to the corresponding GREG estimate. Moreover the non-ADU estimates are very similar to the ADU estimates, perhaps reflecting the validity of Model Two. Finally we note that the standard errors in Table 3 are substantially smaller than those shown in Table 2, indicating that the use of both auxiliary variables is quite advantageous.

REFERENCES

- Brandenburg, L. and C. E. Higgins, Jr., (1974), "Stratified Random Sampling Methods for Class Load Surveys for Electric Utilities," Applied Statistics for Load Research, Vol. 3, New York: Association of Edison Illuminating Companies.
- Brewer, K. R. W., (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.
- _____, _____, _____, (1979), "Prediction Theory for Finite Populations when Model-based and Design-based Principles Are Combined," Scandinavian Journal of Statistics, 6, 97-106.
- Hansen, M. H., G. W. 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, 461-465.
- Isaki, C. T. and W. A. Fuller, (1982), "Survey Design under a Regression Superpopulation Model," Journal of the American Statistical Association, 77, 89-96.
- Royall, R. M., (1970), "On Finite Population Sampling Theory under Certain Linear Regression Models," Biometrika, 57, 377-387.
- _____ and J. Herson, (1973), "Robust Estimation in Finite Populations," Journal of the American Statistical Association, 68, 880-889.
- _____ and W. G. Cumberland, (1981), "An Empirical Study of Prediction Theory in Finite Population Sampling: Simple Random Sampling and the Ratio Estimator," Journal of the American Statistical Association, 76, 66-77.
- 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.
- _____, (1982), "Implications of Survey Design for Generalized Regression Estimation of Linear Functions," to appear in Journal of Statistical Planning and Inference.

Wright, R. L., (1981), "Electric Utility Load Research Using Model-Based Statistical Sampling," Proceedings of the 1981 DOE Statistical Symposium, New York: Brookhaven National Laboratory.

_____, (1982), "Finite Population Sampling with Multivariate Auxiliary Information," School of Business Administration, The University of Michigan.

Table 1. Sampling Plan

Stratum	Sample Size	Number of Customers	Total Usage 10^6 KWH	Total MND 10^3 KW
1	32	19,800	2,370	694
2	51	682	532	134
3	92	118	197	49
Total	175	20,600	3,099	877

Table 2. Various Estimates of T Under Model One.

Approach	q_I	Added Variables	SPRO 10^3 KWH	LPRE 10^3 KWH	GREG 10^3 KWH	SE 10^3 KWH
(1) Ratio	$\pi_I^{-1} x_{I1}^{-1}$	-	6659.7	6631.9	6659.7	198.7
(2) Adjusted Ratio	$(\pi_I^{-1} - 1) x_{I1}^{-1}$	-	6697.4	6667.2	6667.2	199.9
(3) BLU	$\hat{\sigma}_I^{-2}$	-	6326.7	6320.3	6592.7	191.4
(4) π^{-1} *BLU	$\pi_I^{-1} \hat{\sigma}_I^{-2}$	-	7049.3	6996.5	6737.9	214.2
(5) AUG1	$\hat{\sigma}_I^{-2}$	u_{I1}	6741.9	6737.3	6741.9	224.8
(6) AUG2	$\hat{\sigma}_I^{-2}$	u_{I2}	6747.9	6743.3	6743.3	225.4
(7) AUG3	$\hat{\sigma}_I^{-2}$	u_{I1}, u_{I3}	6591.1	6591.1	6591.1	202.2

Table 3. Various Estimates of T Under Model Two.

Approach	q_I	Added Variables	SPRO 10^3 KWH	LPRE 10^3 KWH	GREG 10^3 KWH	SE 10^3 KWH
(1) π^{-1}	π_I^{-1}	-	6529.6	6530.2	6529.6	153.0
(2) Adjusted π^{-1}	π_I^{-1-1}	-	6529.6	6529.0	6529.0	152.5
(3) BLU	$\hat{\sigma}_I^{-2}$	-	6541.5	6539.7	6535.2	152.8
(4) π^{-1} *BLU	$\pi_I^{-1} \hat{\sigma}_I^{-2}$	-	6445.5	6455.5	6491.6	152.6
(5) AUG1	$\hat{\sigma}_I^{-2}$	u_{I1}	6532.8	6530.9	6532.8	152.7
(6) AUG2	$\hat{\sigma}_I^{-2}$	u_{I2}	6535.3	6533.5	6533.5	152.7
(7) AUG3	$\hat{\sigma}_I^{-2}$	u_{I1}, u_{I3}	6517.1	6517.1	6517.1	152.7