SHRINKAGE WEIGHTS FOR UNEQUAL PROBABILITY SAMPLES

Theodora Cohen, Bruce D. Spencer
Theodora Cohen, Searle R&D, Skokie, IL 60091
Key Words: Estimation, Weighting, Unequal probability sampling, Superpopulation models;

1. Introduction

A longstanding question in making inferences from unequal probability samples is whether to use an unweighted or other model-based estimator, say Z_m, or whether to use an approximately unbiased estimator Z_u that uses sampling weights reflecting the unequal selection probabilities. (Moments are defined with respect to the sampling design unless otherwise noted.) An unweighted estimator of a population mean will often have smaller variance than a weighted estimator but it will have bias proportional to the correlation between the characteristic of interest and the sample weights (Rao 1966). For many sampling strategies, the variances of Z_m and Z_u alike decrease to zero as the sample size increases, but although the bias of Z_u is zero or approaches zero the bias of Z_m does not. In such cases, for sufficiently large samples, Z_u will have smaller mean square error. On the other hand, for small samples Z_m may have a smaller mean square error (Cochran 1977, p.296-297). If one could know the mean square errors of Z_m and Z_u one could easily choose the optimal one. Fortunately, it is possible to use the sample itself to estimate the mean square errors, as DuMouchel and Duncan (1983) proposed in a different context.

DuMouchel and Duncan were concerned with estimation of a regression equation and showed that if the assumptions about the error terms in a linear regression model were appropriate for the population from which the sample was drawn, then the regression coefficients could be estimated consistently without sample weights. They proposed testing a null hypothesis that the error terms in the model were uncorrelated with the sample weights. Large empirical correlations were grounds for rejecting the null hypothesis and hence for disbelieving that the regression model was correctly specified. Taking an *analytic* approach, DuMouchel and Duncan were only interested in estimating the regression coefficients if the model was correctly specified; faced with evidence of misspecification they would try to improve the model specification.

In this paper we take a *descriptive* approach of trying to estimate the population mean, whether or not a particular superpopulation model "holds" or is correctly specified. Thus, we may recast the DuMouchel-Duncan procedure as a test-based estimator that equals Z_u if the empirical correlation between the weights and the variable of interest is large and that equals Z_m otherwise. For many sampling problems, the DuMouchel-Duncan estimator has a smaller mean square error than either Z_m or Z_u. However, the DuMouchel-Duncan procedure is not continuous, in the sense that a small change in the data can shift the estimator from Z_m to Z_u or conversely. We derive a continuously weighted average $AZ_u + (1 - A)Z_m$ with A chosen to minimize mean square error.

Under rather general conditions, as the sample size gets large the bias can overwhelm the variance and so A tends to 1. Consequently, the estimator converges to Z_u and, under general conditions, is consistent. This property is not shared by the alternative approaches to modifications of sample weighting as proposed by Rizzo (1990) and Stokes (1990); see section 5. In practice the optimal value of A is not known but must be estimated from the data, as discussed below.

2. General Properties of the Estimator

Let Z_u be an unbiased estimator with variance $V(Z_u) = V_u$ and let Z_m be an estimator with bias $B(Z_m) = B_m$ and variance $V(Z_m) = V_m$. Let C be the covariance between Z_u and Z_m. Define

$$(2.1)\quad A = \begin{cases} A_m/(A_u + A_m) & \text{if } A_m \geq 0 \\ 1 & \text{if } A_u < 0 \\ 0 & \text{otherwise} \end{cases}$$

with $A_m = V_m + B_m^2 - C$ and $A_u = V_u - C$. The following lemma shows that Z_A defined by $Z_A = AZ_u + (1 - A)Z_m$ has smaller bias than Z_m, smaller variance than Z_u, and smaller mean square error (MSE) than Z_m or Z_u.

Lemma 2.1 If $A_u \geq 0$ then

(i) $|B(Z_A)| \leq |B(Z_m)|$
(ii) $MSE(Z_A) \leq MSE(Z_m)$
(iii) $MSE(Z_A) \leq MSE(Z_u)$
(iv) $V(Z_A) \leq V(Z_u)$

with strict inequality if $0 < A < 1$.
The proof is given in the appendix.

The mean square error of Z_A is
$MSE(Z_A) = A^2(V_u) + 2A(1 - A)C + (1 - A)^2 MSE_m$.

The estimator Z_A depends on quantities not usually known: V_u, V_m, B_m^2, and C. A simple approach in practice is to estimate A by using unbiased estimators of these quantities.
Thus, let $a_m = \hat{MSE}_m - \hat{C}$ and $a_u = \hat{V}_u - \hat{C}$ where $\hat{MSE}_m = \hat{V}_m + \hat{B}_m^2$ and

(2.2) $\hat{B}_m^2 = (Z_u - Z_m)^2 - \hat{V}_u - \hat{V}_m + 2\hat{C}$.

We define an estimator of A,
$a = \min\{(a_m/(a_m + a_u), 1\}$ if $a_m \geq 0$ and $a_m + a_u > 0$
$= 0$ otherwise.
We now define the *two-stage estimator* Z_a by

(2.3) $Z_a = aZ_u + (1 - a)Z_m$.

3. Estimation of the Mean in PPS Sampling

As an application of the preceding results, we consider estimation of the population mean from an unequal probability sample.

Under *PPS with-replacement sampling (PPSWR)* unit i has probability $P_i > 0$ of being selected in each of n independent draws, $\Sigma_{i=1}^{N} P_i = 1$. The Hansen-Hurwitz (1943) estimator

$$\bar{y}_{HH} = \Sigma_{i=1}^{n} y_i/(Nnp_i),$$

where the y_i's and p_i's are the values corresponding to the sampled units, is unbiased for this sampling design and has variance (Cochran, 1977, p.252)

$$V_D(\bar{y}_{HH}) = \Sigma_{i=1}^{N} P_i(Y_i/(NP_i) - \bar{Y})^2/n$$

where the Y_i are the population values.

An alternative to the exact design-unbiased estimator is the π-inverse estimator

(3.1) $\bar{y}_{PI} = (\Sigma_{i=1}^{n} y_i/p_i)/(\Sigma_{i=1}^{n} 1/p_i)$,

which is asymptotically design-unbiased (Särndal,1980).

Model-dependent estimators are motivated by assuming various superpopulation models. The *general regression model*, ξ, specifies that $Y_1, \ldots, Y_N$ are independently distributed and

$$E_\xi(\tilde{Y}_i|X_1 \ldots X_N) = \mu + \beta X_i,$$
$$V_\xi(Y_i|X_1 \ldots X_N) = \sigma^2 v(X_i) \quad (i = 1, \ldots, N),$$

where μ, β and σ^2 are fixed but unknown, v(.) is a known function and X_i are fixed positive numbers that are known for the sampled units. We assume that $v(z) = z^g$ where g is known. The population values, $\tilde{Y}_1, \ldots Y_N$, are assumed to be a realization of $\tilde{Y}_1, \ldots, Y_N$.

Brewer (1963) and Royall (1970) have shown that under model ξ the best linear unbiased (ξ-BLU) estimator of the population mean (i.e., minimum variance among all linear unbiased estimators) is

$$\bar{y}_m = \bar{X}(\Sigma_{i=1}^{n} w_i y_i x_i)/(\Sigma_{i=1}^{n} w_i x_i^2)$$

with $w_i = 1/v(x_i)$. Table 1 shows that, for certain variance functions v, the ξ-BLU estimator assumes a simple form, such as the sample mean $\bar{y}$, the regression estimator $\bar{y}_{reg} = \bar{X}(\Sigma_{i=1}^{n} y_i x_i)/(\Sigma_{i=1}^{n} x_i^2)$, the ratio estimator, $\bar{y}_R = \bar{X}(\Sigma_{i=1}^{n} y_i)/(\Sigma_{i=1}^{n} x_i)$, or $\bar{y}_u = \bar{X}\Sigma_{i=1}^{n}(y_i/x_i)/n$, which equals the (design-unbiased) Hansen-Hurwitz estimator $\bar{y}_{HH}$ if $X_i \propto P_i$ and the sample is PPSWR. The ratio of the ξ-expected MSEs of the two estimators $\bar{y}_m$ and $\bar{y}_{HH}$ is $\gamma + V_\xi(B(\bar{y}_m)/E_\xi V(\bar{y}_{HH})$, with $\gamma = E_\xi V(\bar{y}_m)/E_\xi V(\bar{y}_{HH})$.

Table 1. Special Cases of the General Regression Model, ξ

model specification					average-variance ratios	
model name	g	μ	β	BLUE $\bar{y}_m$	$E_\xi V(\bar{y}_m)/$ $E_\xi V(\bar{y}_{HH})$	$E_\xi V(\bar{y}_{HH})/$ $E_\xi V(\bar{y}_{PI})$
ξ_a	0	0	0	$\bar{y}$	≤ 1	≥ 1
ξ_b	0	> 0	0	$\bar{y}$	≤ 1	> 1
ξ_c	0	0	> 0	$\bar{y}_{reg}$	$\leq$ or > 1	$\leq$ or > 1
ξ_d	1	0	0	$\bar{y}_R$	$\leq$ or > 1	≤ 1
ξ_e	2	0	0	$\bar{y}_u$	$= 1$	< 1

When $\mu = \beta = 0$ and v(.) is constant, $E_\xi V_D(\bar{y}_{HH})$ and $E_\xi V_D(\bar{y}_{PI})$ are similar in size.

For the remainder of this paper, $\bar{y}_u$ will refer to $\bar{y}_{HH}$ (models ξ_a, ξ_c and ξ_d) or $\bar{y}_{PI}$ (model ξ_b) and $\bar{y}_m$ will refer to the optimal model-based estimator depending on which model we are considering ($\bar{y}$ for models ξ_a and ξ_b, $\bar{y}_{reg}$ for model ξ_c and $\bar{y}_R$ for model ξ_d).

It follows from the Lemma in section 2 that the estimator $\bar{y}_A = A\bar{y}_u + (1 - A)\bar{y}_m$ has smaller design-mean square error than both $\bar{y}_u$ and $\bar{y}_m$ for any design, where A is defined as in (2.1) with $V_m = V_D(\bar{y}_m)$, $V_u = V_D(\bar{y}_u)$, $B_m = Bias(\bar{y}_m)$ and $C = Cov_D(\bar{y}_m, \bar{y}_u)$. In addition $\bar{y}_A$ has smaller bias than $\bar{y}_m$ and smaller variance than $\bar{y}_u$ under the same conditions.

4. Illustration of The Two-Stage Estimator

To evaluate the performance of $\bar{y}_a$ we performed two sets of simulations.

Simulation I. Twelve different populations of Y's were generated, one for each of the four models $\xi_a,...\xi_d$ in Table 1 and each of three sets of sampling probabilities.

The three sets of sampling probabilities were generated from a Dirichlet distribution. Let $Z_1,...Z_N \approx$ iid gamma with parameter α with density function $f(z) = z^{\alpha-1}e^{-z}/\Gamma(\alpha)$ and let $P_i = Z_i/Z$ where $Z = \Sigma_{i=1}^{N} Z_i$. The joint distribution of the P_i's is Dirichlet with parameter $\eta = (\alpha,....\alpha)$. Three different values of α, 0.5, 1 and 3, were used to generate the Z_i's. The Dirichlet distribution is the N-variate analogue to the bivariate Beta distribution.

By letting the parameters vary we can study the performance of $\bar{y}_a$ when the shape of the histogram of the sampling probabilities is skewed ($\alpha = 0.5$), flat ($\alpha = 1.0$) and normally shaped ($\alpha = 3.0$). The ε_i's in $\xi_a,...\xi_d$ were generated from a N(0,1) distribution.

Five thousand samples of size 50, twenty-five hundred of size 100, one thousand of size 250 and five hundred of size 500 were selected from each of the twelve populations, using a PPSWR sampling design.

The ratios of the mean square errors obtained from the simulations were used to compare the two-stage estimator (2.3) to the design-unbiased estimator and the model estimator. The results are summarized in Table 2.

The first row of Table 2 gives the overall ratio of the mean square errors of the estimators. The last three rows of the table give the ranges of the ratios classified by model, parameter of the distribution of the sampling weights and sample size.

Table 2 Ratios of Mean Square Errors, Models ξ_a-ξ_d

	$MSE(\bar{y}_u)/MSE(\bar{y}_a)$	$MSE(\bar{y}_m)/MSE(\bar{y}_a)$
Overall	3.36	0.94
Model	1.14 - 5.40	0.50 - 1.69
α	1.52 - 4.46	0.86 - 0.98
n	2.76 - 3.79	0.82 - 1.12

$\hat{B}_m$ given by (2.2) is a function of the design-unbiased estimator. When the data are generated from the model, $\bar{y}_u$ is subject to relatively large variation and $V(\hat{B}_m)$ is relatively large. Since $\bar{y}_m$ is unbiased with respect to the models used to generate the data in this simulation, we can replace $\hat{MSE}(\bar{y}_m)$ with $\hat{V}(\bar{y}_m)$. In this case the overall ratio of $MSE(\bar{y}_m)$ and $MSE(\bar{y}_a)$ equals 1.06 with values ranging from 0.92 to 1.63.

Simulation II. The efficiency of the estimators was compared using data from the National Educational Longitudinal Study of 1988 - NELS88 (Spencer et al.,1990) which was conducted by the National Opinion Research Center. The population surveyed is composed of all the known public and private schools in the United States which have eighth grades. Within certain major strata the schools were sampled roughly proportional to the estimated number of students enrolled in eighth grade at the school. The population size is 38,866. The shape of the distribution of the sampling probabilities is skewed with a long right tail, similar to a Dirichlet with $\alpha = 0.5$.

Three variables from the NELS88 study (percent of white students in the school, percent of black students and percent of hispanic students) were used to study the properties of $\bar{y}_a$ under each of the four models $\xi_a,...\xi_d$.

The sample sizes and number of samples selected were the same as for the first simulation.

The ratios of the mean square errors obtained from the simulations were used to compare the two-stage estimator (2.3) to the design-unbiased estimator and the model estimator. The results are summarized in Table 3.

The first row of Table 3 gives the overall ratio of the mean square errors of the estimators. The last three rows of the table give the ranges of the ratios classified by model, variables and sample size.

Table 3 Ratios of Mean Square Errors, NELS88

	$MSE(\bar{y}_u)/MSE(\bar{y}_a)$	$MSE(\bar{y}_m)/MSE(\bar{y}_a)$
Overall	1.05	14.41
Model	0.74 - 1.18	7.37 - 19.30
Variables	0.90 - 1.27	6.83 - 24.84
n	0.98 - 1.21	3.23 - 32.45

Unlike the first simulation, none of the models considered fits the data well and $\bar{y}_m$ is biased. In this case, $\hat{MSE}(\bar{y}_m)$ is of the same order of magnitude as $MSE(\bar{y}_m)$.

By estimating the components of A we lose the "minimum mean square error" property which $\bar{y}_A$ has, but the results of the two simulations show that $\bar{y}_a$ works well as a compromise estimator.

5. Related Estimators

The property of design-consistency is not shared by a different class of estimators which is derived assuming that the sample is given and the unobserved population values are predicted under an assumed model. The resulting estimator can be represented as

$$T = (1/N)\Sigma_{i=1}^{n} y_i + (1 - n/N)U,$$

where U is a predictor of $(\Sigma_{i=1}^{N} Y_i - \Sigma_{i=1}^{n} y_i)/(N - n)$. The observed sample is sometimes used to estimate the unknown parameters of the distribution of **Y** under the assumed model. One approach (e.g. Cassel, Särndal and Wretman, 1977) is to choose U to minimize the ξ-expected mean square error of T, which is equivalent to minimizing the ξ-mean square error of T when the sampling design does not depend on the sampled y-values. A more general approach (Rizzo, 1989) is to include the sample indicator, the inclusion probabilities as well as the joint inclusion probabilities into the model, as random variables. The sample indicator is a vector indicating which elements are selected into the sample. Under this approach the sample indicator and the variable of interest **Y** have a joint distribution. The predictive estimator under the simplest specification of the model ($E_\xi(Y_i|\pi_i) = \gamma\pi_i$) is

$$\bar{y}_P = (1/N)[\ \hat{\gamma}\ (\Sigma_{i=1}^{n}\Pi_i - \Sigma_{i=1}^{n}\pi_i) + \Sigma_{i=1}^{n} y_i)$$

where $\hat{\gamma} = (\Sigma_{i=1}^{n}\pi_i y_i)/(\Sigma_{i=1}^{n}\pi_i^2)$. As mentioned above, this estimator does not converge to the unbiased estimator as the sample size increases. Since the unbiased estimator converges almost surely to the population mean, it follows that $\bar{y}_P$ is not design-consistent.

Another approach which leads to an estimator similar to $\bar{y}_a$ is described by Stokes (1990). The population consists of H strata of M_i units each, i = 1,....,H. The model assumed is $E(Y_{ij}|\mu_i) = \mu_i$ and $Var(Y_{ij}|\mu_i) = \tau^2$ where $E(\mu_i) = \mu$, $Var(\mu_i) = \sigma^2$, the Y_{ij}'s and μ_i's are independent and m units are sampled from each stratum. The BLU estimator of the population mean $\bar{Y} = \Sigma_{i=1}^{H} W_i\bar{Y}_i$ (where $W_i = M_i/(\Sigma_{i=1}^{H} M_i)$ and $\bar{Y}_i$ is the mean in stratum i) is

$$\hat{T} = (1/(Hm))\Sigma_{i=1}^{H}\Sigma_{j=1}^{m} w_i y_{ij}$$

where $w_i = B + (1 - B)(HW_i)$, $B = (\tau^2/m)/(\sigma^2 + \tau^2/m)$. $\hat{T}$ can also be written as $\hat{T} = B\bar{y} + (1 - B)\bar{y}_{st}$ where $\bar{y}_{st} = \Sigma_{i=1}^{H} W_i\bar{y}_i$ is the conventional design-unbiased estimator and $\bar{y}$ is the unweighted sample mean. $\hat{T}$ is called a shrinkage estimator because its weights ($w_i = B + (1 - B)W_i$) are a weighted average of the weights of the conventional estimator ($w_i = W_i$) and 1/H. The amount of shrinkage is determined by the relative variability of $\bar{y}_i$ and μ_i. When the variance within the strata is small compared to the variance between the strata, the estimator is almost identical to the conventional estimator. On the other hand, if the variance between the strata is small compared to the variance within the strata, then all the units in the sample are weighted approximately equally. As in the case of $\bar{y}_A$, this estimator can only be calculated if B is known. Stokes (1990) describes the performance of the estimator when B is estimated from the sample. $\hat{T}$ is design-consistent when the number of strata is kept fixed and the sample size in each stratum increases. When the number of strata increases as well, $\hat{T}$ does not converge to a design-unbiased estimator. If in our approach we replace the PPSWRX sampling design with a stratified design as described above, the resulting estimator is

$$\bar{y}_A = A\bar{y}_{st} + (1 - A)\bar{y}$$

where $1 - A = \delta(\Sigma_{i=1}^{H} W_i^2 - 1/H)/(\delta(\Sigma_{i=1}^{H} W_i^2 - 1/H) + (\Sigma_{i=1}^{H}\bar{Y}_i/H - \bar{Y})^2)$ with $\delta = Var_\xi(\bar{y}_i)$. This differs from the weight B given to $\bar{y}$ in $\hat{T}$ which can be written as $B = Var(\bar{y}_i)/(Var(\bar{y}_i) + \sigma^2)$. Unlike $\hat{T}$, $\bar{y}_A$ does not require any model assumptions and W_i and $\bar{Y}_i$ can be correlated.

DuMouchel and Duncan (1983) give a method of choosing between the randomization-based and the model-based approach. Their method can be applied to our case by assuming that the model under consideration is the general regression model with $\beta = 0$ and $g = 0$ and the sampling design is PPSWRX. Let $X_i = 1$, i = 1,....,n and $w_i = 1/p_i$. Then the least squares estimate of the regression coefficient is $\hat{b} = \bar{y}$ and the weighted least square estimate is $\hat{b}_w = (\Sigma_{i=1}^{n} y_i/p_i)/(\Sigma_{i=1}^{n} 1/p_i)$ which is equivalent to $\bar{y}_{PI}$ given by (3.1).

A test of whether $\bar{y}$ or $\bar{y}_{HH}$ should be used is based on the difference $\hat{\Delta}_y = \bar{y} - \bar{y}_{PI}$.

We compared the compromise estimator with the estimator obtained using DuMouchel and Duncan's method of selecting the "optimal" estimator. We generated three data sets from Model ξ_a ($\alpha = 1$) with correlations of 0, 0.1 and

0.5 between Y_i and P_i. Then we selected one thousand samples from each data set using the same PPSWR sampling method as in section 7. For each sample we calculated the F statistic for testing $E(\hat{\Delta}_y) = 0$ and its p-value denoted by p.

The DuMouchel - Duncan estimator is defined as follows

$$\bar{y}_D = \begin{cases} \bar{y}_u & \text{if } p < 0.05 \\ \bar{y} & \text{otherwise.} \end{cases}$$

We also calculated the mean square error of the two-stage estimator as well as that of the DuMouchel-Duncan estimator. The results are given in Table 4.

TABLE 4 Mean Square Errors of the Two-Stage Estimator and the DuMouchel-Duncan Estimator.

	CORRELATION COEFFICIENT		
ESTIMATOR	0	0.1	0.5
Two-Stage	0.0262	0.0336	0.0528
DuMouchel-Duncan	0.0468	0.0552	0.1181

The power of the F - test as calculated from the simulation is 0.113 and 0.463 for a correlation coefficient of 0.1 and 0.5 respectively.

In conclusion, $\bar{y}_a$ protects us against misspecification of the model and it has appreciably smaller mean square error than $\bar{y}_u$ in those cases where the model fits.

6. Conclusions

This paper introduces a method of developing new estimators which are optimal under both a given sampling design and superpopulation model. The compromise estimator, $\bar{y}_A$, is better than both the unbiased and the optimal model estimator under the fixed population approach as well as the superpopulation approach for PPSWRX sampling and all the models considered, however it is not of practical use. The two-stage estimator, $\bar{y}_a$, can be used to estimate the population mean any time we are willing to tolerate some bias in exchange for smaller mean square error. When the model fits the data exactly, $\bar{y}_a$ approaches the model-optimal estimator which has smaller design-mean square error than the unbiased estimator. When the model is misspecified $\bar{y}_a$ approaches the unbiased estimator which in this case has smaller variance than the mean square error of the model-based estimator. In all other cases, $\bar{y}_a$ approaches $\bar{y}_A$ which is better than both the unbiased estimator and the model-based estimator.

The mean square error of $\bar{y}_a$ can be considerably smaller than the variance of the unbiased estimator if the model fits the data well, and than the mean square error of the model estimator when the model is misspecified.

Appendix

Proof of Lemma 2.1

(i) If $A_m < 0$ then $A = 0$ and if $A_u \geq 0$ then $A = A_m/(A_u + A_m) \leq 1$. Thus $0 \leq A \leq 1$. Combining these two facts we have $0 \leq A \leq 1$. It follows that $|B_D(Z_A)| = (1 - A)|B_m| \leq |B_m|$ with strict inequality if $A < 1$. Thus (i) is established.

(ii) If $A = 0$ then $Z_m = Z_A$ and (ii) holds.

If $A > 0$ then

$MSE(Z_A) = A^2(A_m + A_u) - 2AA_m + A_m + C$
$= A_m(1 - A) + C < A_m + C = MSE_D(Z_m)$

since $A \leq 1$. Thus (ii) is established.

(iii) Note that $MSE(Z_u) = A_u + C$.

If $A > 0$ then $MSE(Z_u) - MSE(Z_A) = A_u - A_mA_u/(A_u + A_m) = A_u^2/(A_u + A_m) > 0$.

If $A = 0$ then $A_m \leq 0$ and $A_u - A_m > 0$.

Thus, $MSE(Z_A) = MSE(Z_m) < MSE(Z_u)$.

(iv) $MSE(Z_A) \geq V(Z_A)$ with strict inequality if $A > 0$, and $MSE(Z_u) = V(Z_u)$. The result follows from (iii).#

The framework for the asymptotic analysis is that of Brewer (1979). The original population of N units is reproduced (k-1) times. A sample is selected from each of the resulting k populations using the same sample-selection procedure (i.e. the same P_i's) for each one. The k populations are aggregated to a population of size $N_k = kN$ units with a population total $Y_k = kY$. The k samples are aggregated to a sample of n_k units. k is allowed to tend to infinity. We assume that all of the P_i are greater than zero and the variances and covariances of sample averages and functions of sample averages approach zero as n_k approaches infinity.

Proposition A.1

$E(a) = A + O(n^{-1})$ and

(A.1) $Var(a) = \{(C - \widehat{MSE}_m)^2 Var(\hat{V}_u) + 2(C - MSE_m)(V_u - C)Cov(\hat{V}_u, \widehat{MSE}_m) + 2(C - MSE_m)(MSE_m - V_u) Cov(\hat{V}_u, \hat{C}) + (V_u - C)^2 Var(\widehat{MSE}_m) + 2(V_u - C)(MSE_m - V_u)Cov(\widehat{MSE}_m, \hat{C}) + (MSE_m - V_u)Var(\hat{C})\} /(n(MSE(\bar{y}_m - \bar{y}_u))^4) + O(n^{-3/2})$.

Proof

First, note that all the components of a can be expressed as functions of sample averages. Second, if $MSE(\bar{y}_m) + Var(\bar{y}_u) - 2Cov(\bar{y}_m,\bar{y}_u) > 0$ then a satisfies condition (1) of Theorem 14.5-2 in Bishop, Fienberg and Holland (1975, p.486). Note that $MSE(\bar{y}_m) + Var(\bar{y}_u) - 2Cov(\bar{y}_m,\bar{y}_u) = MSE(\bar{y}_m - \bar{y}_u) > 0$. Third, a is bounded by definition. Thus, we can apply the theorem mentioned above and see that $E(a) = A + O(n^{-1})$ and (A.1) holds as well.#

Proposition A.2

The asymptotic expected value of $\bar{y}_a$ to order n^{-1} is

(A.2) $E(\bar{y}_a) = \bar{Y} + O(n^{-1})$

and the asymptotic mean square error of $\bar{y}_a$ to order $n^{-3/2}$ is $MSE(\bar{y}_a) = [B^2Var(a) - 2ABCov(a,\bar{y}_u) - 2(1 - A)BCov(a,\bar{y}_m) + A^2Var(\bar{y}_u) + 2A(1 - A)Cov(\bar{y}_u,\bar{y}_m) + (1 - A)^2MSE(\bar{y}_m)]/n$, where Var(a) is given by (A.1).

Proof

Both equations above are derived using the same theorem as in Proposition A.1. Note that $\bar{y}_a < \bar{y}_u + \bar{y}_m$, and this expression is bounded. Thus, $E(\bar{y}_a) = \bar{Y} + (1 - A)B + O(n^{-1})$ and $(1 - A)$ is $o(n^{-1})$ and B is O(1) so that (A.2) holds.#

Proposition A.3

$n^{-1/2}(\bar{y}_a - \bar{y}_u) \to 0$ in probability as $n \to \infty$.

Proof $a = O_p(n^{-1})$, $\bar{y}_m = \bar{Y} + B + O_p(n^{-1/2})$, $\bar{y}_u = \bar{Y} + O_p(n^{-1/2})$ and $B = O(1)$. Thus $(\bar{y}_m - \bar{y}_u) = O_p(1)$ and

$n^{-1/2}(\bar{y}_a - \bar{y}_u) = n^{-1/2}(1 - a)(\bar{y}_m - \bar{y}_u) = O_p(n^{-1/2})O_p(1) = O_p(n^{-1/2})$.#

BIBLIOGRAPHY

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975) Discrete Multivariate Analysis: Theory and Practice. Cambridge: The MIT Press.

Brewer, K.R.W. (1979) A Class of Robust Sampling Designs for Large-Scale Surveys. JASA 74: 911 - 915.

Brewer, K.R.W. and Hanif, M. (1983) Sampling with Unequal Probabilities. New York: Springer-Verlag.

Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1977) Foundations of Inference in Survey Sampling. New York: John Wiley & Sons.

Chaudhuri, A. and Vos, J.W.E. (1988) Unified Theory and Strategies of Survey Sampling. Amsterdam: North Holland.

Cochran, W.G. (1977) Sampling Techniques. New York: John Wiley & Sons.

DuMouchel, W.H. and Duncan, G.J. (1983) Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples. JASA 78: 535 - 543.

Rao, J.N.K. (1966) Alternative Estimators in PPS Sampling for Multiple Characteristics. Sankhya, A 28: 47 - 60.

Rizzo, L. (1990) Predictive Regression Estimators Using Functions of the Probability of Selection. Unpublished paper, University of Iowa, Department of Statistics and Actuarial Science.

Särndal, C.E. (1980) On π-inverse Weighting versus Best Linear Unbiased Weighting in Probability Sampling. Biometrika, 67,3: 639 -650.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K.A. and Tourangeau, R. (August, 1990) National Education Longitudinal Study of 1988, Base Year Sample Design Report. Washington, D.C.: U.S. Department of Education.

Stokes, L. (1990) A Comparison of Truncation and Shrinkage of Sample Weights. Unpublished paper, University of Texas at Austin.