Optimized Whole Sample Procedures versus Traditional Draw-by-Draw Procedures

Sun-Woong Kim[1], Steven G. Herringa[2], Sung-Joon Hong[3], Peter W. Solenberger[4]
[1]Department of Statistics, Dongguk University, 26, 3Ga, Pil-Dong, Jung-Gu, Seoul, South Korea 100-715
[2]Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan 48106
[3]Department of Statistics, Dongguk University, 26, 3Ga, Pil-Dong, Jung-Gu, Seoul, South Korea 100-715
[4]Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan 48106

Abstract

Kim, Heeringa, and Solenberger (2006) suggested two approaches for model-based sampling designs of reducing the variance of the Horvitz-Thompson (1952)'s estimator. Their methods are whole sample procedures based on optimization theory under a superpopulation model. With respect to the sample selection probabilities, we theoretically present the differences between those procedures and popular draw-by-draw procedures such as sampling methods of Mizuno (1952), Brewer (1963a), and Murthy (1957), when the sample size is two, which is a common situation in nationwide samples with many strata. We also compare the efficiencies between them for natural populations in the published literature. It appears that the methods of Kim, Heeringa, and Solenberger (2006) may be preferable to those draw-by-draw procedures in many situations.

Key Words: average variance, πPS sampling, regression superpopulation model, optimization problem

1. Introduction

In recent decades there have been a number of developments in inclusion probabilities proportional to size (πPS) sampling methods that employ an auxiliary variable as a measure of size of each unit in a finite population. These sample selection methods are applicable to a generalized regression (GREG) estimator based on an underlying superpopulation model, as well as the Horvitz and Thompson (H-T) (1952) estimator. The majority of these are traditional draw-by-draw procedures (TDDP) and some of them, described in the next section, can be easily run using software programs available in SAS or SPSS.

With respect to the efficiency of the GREG estimator, sampling methods for minimizing the anticipated variance (ANV) of the estimator have been developed (Fuller and Isaki, 1982). These methods depend on the variance pattern of the error terms in the superpopulation model. One such selection procedure is a model-based stratified simple random sampling (SSRS), proposed by Wright (1983).

The GREG estimator is a design-consistent estimator when a large sample is selected, but it may be appreciably biased for a small sample size. For this reason, many sampling statisticians still view the H-T estimator as an attractive alternative for estimation of population parameters.

For samplers who prefer the H-T estimator and are eager to increase its efficiency there have been few options to select samples, except for the traditional πPS sampling methods, shown in the next section.

One alternative to those πPS sampling methods, proposed by Kim, Heeringa, and Solenberger (2006), is model-based πPS sampling methods based on the optimization problem for minimizing the model expectation of the variance, or simply, the average variance (AV) of the H-T estimator under the superpopulation model. Note that the AV of the H-T estimator is the reduced form for the ANV of the GREG due to the unbiasedness of the H-T estimator. Their methods are whole sample procedures termed optimized whole sample procedures (OWSP). OWSP are optimized πPS sampling methods that differ from traditional πPS sampling. The following is the background for the development of the OWSP.

Considering a finite population of N units, one often assumes the population is drawn from an infinite superpopulation with the regression model ξ , as given by

$$y_i = b\,x_i + e_i\,,\ i = 1,\cdots,N\,, \tag{1.1}$$

where $E_x\left(e_i \middle| x_i\right) = 0$, $V_x\left(e_i \middle| x_i\right) = a x_i^g$ ($a > 0$, $g \geq 0$), and $E_x\left(e_i e_j \middle| x_i, x_j\right) = 0$, $i \neq j$.

E_x and V_x in (1.1) denote the expected value and variance under the model x. It is further assumed that e_i' s in (1.1) are normally distributed.

As shown by Rao and Bayless (1969) and Kim, Heeringa, and Solenberger (2006), when assuming (1.1) with zero intercept, the AV of the H-T estimator is the same for any pPS sampling methods.

However, if one assumes the regression superpopulation model includes an intercept, this more practical choice of a model in many survey populations may be expressed as:

$$y_i = a + b\,x_i + e_i\,,\ i = 1,\cdots,N\,. \tag{1.2}$$

Under this model, the AVs of the H-T estimator differ for the various pPS sampling methods (Kim, Heeringa, and Solenberger, 2006). Accordingly, the OWSP for minimizing the AV of the H-T estimator can be considered, although the AV is a theoretical value that depends on the assumed superpopulation model.

This study focuses on only the case where the sample size is two (as in two per stratum designs), which requires a simple sampling procedure and is a common situation in nationwide samples with many strata. In this paper, we first describe TDDP and OWSP. Second, we present the differences of the structures of sample selection probabilities between OWSP and TDDP procedures such as the methods of Mizuno (1952), Brewer (1963a), and Murthy (1957). Third, we empirically compare the variances between TDDP and OSWP for natural populations given in Rao and Bayless (1969). In order to estimate the parameters in the model (1.2) for OWSP, we use the algorithm of Harvey (1976).

2. TDDP and OWSP

The classification of TDDP and OWSP for sampling without replacement procedures follows the manner of Brewer and Hanif (1983) based on that of Carroll and Hartley (1964). As noted by Brewer and Hanif (1983), for TDDP, one unit is selected at each successive draw and the probability of selection is defined for each draw.

A number of TDDP have been developed and many books on survey sampling basically refer to TDDP such as the methods of Mizuno (1952) and Brewer (1963a), which are pPS sampling methods. Also, Rao and Bayless (1969) showed the superiority of Murthy (1957)' s method, which is not a pPS sampling method. For his method, Murthy' s estimator, which is an unbiased estimator, can be used instead of the H-T estimator.

"Optimized" in the OWSP indicates "variance minimization." One of the oldest methods for variance minimization is that of Raj (1956), which is a pPS sampling method, followed by Jessen (1969) and Kim, Heeringa, and Solenberger (2003, 2005). The OSWP developed by Kim, Heeringa, and Solenberger (2006) is an alternative to those. For OWSP, the units are not drawn individually but the selection probability, $p(s)$, of each possible sample of size n is specified by optimization approach. Hence one selection using these probabilities simultaneously selects the n sample units.

3. Structures of Sample Selection Probabilities in TDDP and OWSP

A sample of n units from the population of size N is selected with a sampling method. Let $p_i = x_i / X$ be the relative size of the unit, where $X = \sum_{i=1}^{N} x_i$ and x_i is an auxiliary variable correlated with the variable of the interest, y_i. Let $p(s)$ denote the selection probability of a sample s. Then the probability that unit i will be included in a sample, denoted p_i and called the first-order inclusion probability, is expressed as

$$p_i = \sum_{i \in s} p(s) \tag{3.1}$$

Also, the probability that both of the units i and j will be included in a sample, p_{ij}, called the second-order inclusion probability, is given by

$$p_{ij} = \sum_{i,\, j \in s} p(s) \tag{3.2}$$

Note that $p_{ij} = p(s)$ when $n = 2$.

To differentiate between the sample selection probabilities $p(s)$ for TDDP and OWSP, we will use $p_d(s)$ for TDDP and $p_x(s)$ for OWSP. The $p_d(s)$ indicate the design-based sample selection probabilities and the $p_x(s)$ denote the model-based sample selection probabilities.

3.1 Sample Selection Probabilities in TDDP

The method of Mizuno (1952) uses the following selection procedure:

i. Select the first unit with unequal probabilities, p_i.
ii. Select the remaining $n-1$ units according to a simple random sampling without replacement.

Then the sample selection probabilities for $n = 2$ are as follows:

$$p_d(s) = \frac{1}{N-1} p_i + \frac{1}{N-1} p_j \tag{3.3}$$

(3.3) may be expressed in another form by using $f(x_i, x_j) = x_i + x_j$. This gives

$$p_d(s) = \frac{1}{(N-1)X} f(x_i, x_j) \tag{3.4}$$

It can be re-expressed as

$$p_d(s) = a + b \sum_{i \in s} x_i \, , \tag{3.5}$$

where $a = 0$ and $b = \dfrac{1}{(N-1)X}$.

(3.5) is similar to the form for the sample selection probabilities that Mukhopadhyay and Vijayan (1996, page 776) used for a special method called controlled sampling.

The method of Brewer (1963a) may be used for situations where the sample size is just two (or two per stratum). The selection procedure is as follows:

i. Select the first unit with probability $\dfrac{p_i(1-p_i)}{1-2p_i}$.

ii. Select the second unit with probability $\dfrac{p_i}{1-p_j}$, where j is the unit drawn first.

The sample selection probabilities are the form of

$$p_d(s) = \frac{1}{DX} g(x_i, x_j) \, , \tag{3.6}$$

where $D=\frac{1}{2}\left(1+\sum_{i=1}^{N}\frac{x_i}{X-2x_i}\right)$ and $g(x_i,x_j)=x_ix_j\left(\frac{1}{X-2x_i}+\frac{1}{X-2x_j}\right)$.

Note that although (3.6) looks like (3.4), (3.6) can not be re-expressed as the form of (3.5) because of the difference between $f(x_i,x_j)$ and $g(x_i,x_j)$.

Murthy (1957)'s method follows the most natural sequence of steps, as denoted by Cochran (1977). That is, The successive units are selected with probabilities p_i, $p_j/(1-p_i)$, $p_k/(1-p_i-p_j)$, and so on.

For $n=2$, Murthy's method has the sample selection probabilities given by

$$p_d(s)=\frac{1}{X}h(x_i,x_j), \tag{3.7}$$

where $h(x_i,x_j)=x_ix_j\left(\frac{1}{X-x_i}+\frac{1}{X-x_j}\right)$.

Note that (3.7) is similar to (3.6) for Brewer's method. The methods of Brewer and Murthy can be easily run in software such as SAS or SPSS.

3.2 Sample Selection Probabilities in OWSP

The methods of Kim, Heeringa, and Solenberger (2006) are as follows. The H-T (1952) estimator for the population total $Y=\sum_{i=1}^{N}y_i$ is given by

$$\hat{Y}_{HT}=\sum_{i=1}^{n}\frac{y_i}{\pi_i} \tag{3.8}$$

When using the form of the design variance of the H-T estimator given by

$$Var_1\left(\hat{Y}_{HT}\right)=\sum_{i=1}^{N}\frac{y_i^2(1-\pi_i)}{\pi_i}+2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{\pi_{ij}}{\pi_i\pi_j}y_iy_j-2\sum_{i=1}^{N}\sum_{j>i}^{N}y_iy_j, \tag{3.9}$$

the AV of Var_1 under the regression superpopulation model of (1.2), denoted $E_x\left(Var_1\right)$, is

$$E_x\left(Var_1\left(\hat{Y}_{HT}\right)\right)=\frac{X^2}{n}\left[\frac{2\alpha}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{\alpha+\beta(x_i+x_j)}{x_ix_j}\pi_{ij}+\beta^2(n-1)\right]+\sum_{i=1}^{N}\left(X/nx_i-1\right)\left(\alpha^2+(\beta^2+\sigma^2)x_i^2+2\alpha\beta x_i\right)$$
$$-2\sum_{i}^{N}\sum_{j>i}^{N}\left(\alpha^2+\alpha\beta(x_i+x_j)+\beta^2x_ix_j\right) \tag{3.10}$$

In cases of $n=2$, where $\pi_{ij}=p_x(s)$, the optimization problem (OP) for minimizing (3.10) has the form

$$Minimize\quad \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{\alpha+\beta(x_i+x_j)}{x_ix_j}p_x(s) \tag{3.11}$$

subject to the linear inequality constraints

$$c\pi_i\pi_j\le p_x(s)\le\pi_i\pi_j,\quad j>i=1,\cdots,N, \tag{3.12}$$

where c is a real number between 0 and 1, and

$$\sum_{i \in s} p_x(s) = \pi_i \ , \ i = 1, \cdots, N \ , \tag{3.13}$$

where $\pi_i = np_i$.

Let us call this optimization problem OP I. Since the objective function in (3.11) and the constraints in (3.12) and (3.13) are linear forms, a linear programming (LP) algorithm can be used to find a solution for an optimal sampling design $p_x(\cdot)$.

Considering a different form of the design variance of the H-T estimator given by

$$Var_2\left(\hat{Y}_{HT}\right) = \sum_{i=1}^{N} \sum_{j>i}^{N} \left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 = \sum_{i=1}^{N} \sum_{j>i}^{N} \left(p_i p_j - \frac{\pi_{ij}}{n^2}\right)\left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 \tag{3.14}$$

here, the AV for Var_2, denoted $E_x\left(Var_2\right)$, is as follows:

$$E_x\left(Var_2\left(\hat{Y}_{HT}\right)\right) = V_x + 2\alpha\left(\sum_{i=1}^{N} \sum_{j>i}^{N} \left(x_j - x_i\right)\left(\alpha x_i^{-1} + \beta\right)\right) + \frac{2\alpha X^2}{n^2} \sum_{i=1}^{N} \sum_{j>i}^{N} \left(x_j^{-1} - x_i^{-1}\right)\left(\alpha x_i^{-1} + \beta\right)\pi_{ij} \ , \tag{3.15}$$

where $V_x = \frac{\alpha X^{\gamma}}{n} \sum_{i=1}^{N} (1 - np_i) p_i^{\gamma - 1}$ and $\pi_i < n$.

The AV in (3.15) is different from that in (3.10). For example, in order to simplify, let $\alpha = 0$. Then $E_x\left(Var_2\left(\hat{Y}_{HT}\right)\right)$ reduces to V_x, while $E_x\left(Var_1\left(\hat{Y}_{HT}\right)\right)$ amounts to

$$\beta^2\left[\frac{n-1}{n} X^2 + \frac{X}{n}\sum_{i=1}^{N} x_i^{\gamma-1} - \sum_{i=1}^{N} x_i^{\gamma} - 2\sum_{i=1}^{N}\sum_{j>i}^{N} x_i x_j\right] + V_x \tag{3.16}$$

Note that since the first term in (3.16) involving β^2 is not zero, (3.10) is not equal to (3.15).

We may consider the following optimization problem to minimize (3.15), the AV of the H-T estimator for $n = 2$.

$$\textbf{\textit{Minimize}} \ \sum_{i=1}^{N} \sum_{j>i}^{N} (x_j^{-1} - x_i^{-1})\left(\alpha x_i^{-1} + \beta\right) p_x(s) \tag{3.17}$$

subject to the constraints (3.12) and (3.13).

A solution on this problem can be obtained by using LP algorithm like OP I. Here we call this problem OP II. Note that both OP I and OP II depend only on α and β, regardless of the values of σ^2 or γ in the superpopulation model.

3.3 Differences between TDDP and OWSP

As shown in (3.4), (3.6) and (3.7), the sampling design $p_d(\cdot)$ in TDDP is a known function of x_i and x_j. But $p_x(\cdot)$ in OWSP is an unknown function of x_i and x_j.

The sampling design in OWSP depends on the optimization problems denoted by OP I or OP II, which are to minimize the AV of the H-T estimator under the assumed superpopulation model. In other words, $p_x(\cdot)$ is a solution to the optimization problems and the solution can be obtained by using a LP algorithm. A number of software programs are available to specify LP algorithm and solve for the optimum.

4. Estimation of the Superpopulation Model

As mentioned before, we assume that a finite population is drawn from an infinite superpopulation with the regression model x . If we know or can estimate the superpopulation model at the design stage, it is applicable to the theory of OWSP, developed by Kim, Heeringa, and Solenberger (2006). As denoted by Godfrey, Roshwalb and Wright (1984) and Särndal and Wright (1984), Harvey (1976)' s algorithm may be used to calculate the maximum likelihood estimates of a , b , s^2 and g in the superpopulation model such as (1.2). In Harvey' s algorithm the starting values of a and b are the ordinary least squares (OLS) estimates and in each iteration the values of a and b depend on s^2 and g or the reverse. See Harvey (1976, pages 463-464) for the details on the algorithm and see Godfrey, Roshwalb and Wright (1984) and Särndal and Wright (1984) for the real applications of the algorithm.

5. Empirical Comparison of the Efficiencies between TDDP and OWSP

Rao and Bayless (1969) empirically studied the efficiencies of several unequal probability sampling methods for two units per stratum under the superpopulation model (1.1). In this study we consider the model (1.2) with the intercept, which may be more appropriate than (1.1) in practical surveys. For the comparison between TDDP and OWSP we also choose 5 natural populations with the size N ranging from 10 to 20 in Kish (1965), Rao (1963), Hanurav (1967), and Horvitz and Thompson (1952) among the populations Rao and Bayless considered in the paper. They have different scales and different distributions.

Table 1 presents the maximum likelihood estimates of a , b , s^2 and g for those finite populations obtained from Harvey (1976)' s algorithm, assuming that those populations are drawn from infinite superpopulations. Some small values of a , which are away from zero, are negative and some are large. Although the values of g are expected in the interval (0, 2) or (1,2), as discussed by Cochran (1953), Brewer (1963b), and Särndal, Swensson and Wretman (1992), some here have values larger than 2.

Table 1. Maximum Likelihood Estimates

Pop No.	N	No. of Iterations	a	b	s^2	g
1	10	41	-0.5256	0.5058	0.0203	2.2964
2	10	13	-0.8130	0.5951	0.1322	1.4108
3	14	11	25.9264	1.0272	0.1989	1.6099
4	20	38	185718.57	1015.3631	0.5370	3.2817
5	20	14	1.1426	1.0381	0.0038	2.7461

To obtain $p_x(\cdot)$, a solution to the optimization problems OP I or OP II, the "LP procedure" in SAS/OR was used. This procedure adopts the two phase revised simplex method. The solution from "LP procedure" may depend on the order of the units. But any solution can be accepted. We followed the order of units presented in the original sources. See SAS/OR (2004) for the details on "LP procedure."

Table 2 shows the results of the comparison of the variances under the different methods. In all populations Mizuno' s method has greater variance than the other methods. Brewer' s method and OP II do not perform as well as Murthy' s method and OP I. More specifically, for population 1, where $N = 10$, Murthy' s method performs the best and OP I when $c = 0.3$ yields the second lowest variance. For population 2, again with $N = 10$, Murthy' s method is still best and OP I when $c = 0.1$ is again in "second place." In population 3 ($N = 14$) OP I with $c = 0.2$ is the best, beating Murthy' s method. In population 4 ($N = 20$) OP II with $c = 0.1$ is the best and OP I when $c = 0.5$ is the second best, beating Murthy' s method. In population 5, OP I performs best. Based on these empirical tests with a limited number of populations, OP I is preferred to the other sample selection methods.

Table 2. Comparison of Efficiency

Method			$Var(\hat{Y})$				
			Pop1	Pop2	Pop3	Pop 4	Pop 5
Mizuno			2,908	2,183	53,353	1.49E13	6,411
Brewer			622	568	37,211	3.46E12	3,011
Murthy			598	480	36,771	3.44E12	3,031
c	0.1	OP I	795	520	37,536	3.49E12	2,989
	0.2		793	539	32,021	3.47E12	2,875
	0.3		621	NA	38,639	3.49E12	2,831
	0.4		660	NA	37,588	3.57E12	2,859
	0.5		NA	NA	38,696	3.43E12	3,058
c	0.1	OP II	778	612	53,565	3.33E12	3,195
	0.2		725	602	52,609	3.48E12	2,963
	0.3		662	NA	50,039	3.46E12	3,139
	0.4		694	NA	44,700	3.50E12	3,055
	0.5		NA	NA	38,324	3.49E12	3,009

Note. "NA" indicates "not available due to no solution."

We do not compare the values of the AV between OP I and OP II or those with the design variance such as Var_1 or Var_2 because the value of the AV is a measure of a theoretical error computed for the assumed superpopulation and conceptually different from the design variance. For example, for population 5, OP I when $c = 0.4$ performs the best and OP II when $c = 0.4$ yields the second with regard to the AV. Then the values of the AV are 2,255 and 2,495, respectively. These results for the AV are not consistent with those for the OP I and OP II design variance for population 5 in Table 2. Also, note that for some populations the value of the AV (3.15) can be negative due to the second term in (3.15).

6. Concluding Remarks

We have used the algorithm of Harvey (1976) for estimating the superpopulation model in (1.2) and examined the capacity of OWSP developed by Kim, Heeringa, and Solenberger (2006) to yield a smaller design variance and a smaller AV.

Based on our work, it appears that OWSP, especially OP I, may be preferable in many situations to TDDP such as the methods of Mizuno, Brewer and Murthy in terms of the efficiency. We also observe that OWSP shows better results as the population size increases.

Since the objective function in the optimization problem for OP I and OP II has a simple linear form, finding a solution, the sample selection probabilities, is not complicated. It seems that the linear constraints involving the value of c in (3.12) are quite useful to reduce the variance, but it obviously requires a careful choice of the value. Empirical studies for additional natural populations are needed to determine if there are sample designs where the optimal solution is not feasible.

An investigation of the efficiency for $n > 2$ is also planned as well as a study on the sensitivity of these findings to misspecification of superpopulation model. The stability of the variance estimator as well as the efficiency may need to be studied for more sampling methods. A nonlinear superpopulation model might be adopted to develop a new sample selection procedure.

References

Brewer, K. R. W. (1963a). "A model of systematic sampling with unequal probabilities," *Australian Journal of Statistics*, 5, 5-13.

Brewer, K. R. W. (1963b). "Ratio estimation and finite populations: some results deducible from the assumption of underlying stochastic process," *Australian Journal of Statistics*, 5, 93-105.

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

Carroll, J. L. and Hartley, H. O. (1964). "The symmetric method of unequal probability sampling without replacement," Abstract in *Biometrics*, 20, 908-909.

Cochran, W.G. (1953). *Sampling Techniques*, 1st ed. New York: Wiley.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: Wiley.

Godfrey, J., Roshwalb, A., and Wright, R. L. (1984). "Model-based stratification in inventory cost estimation," *Journal of Business and Economic Statistics*, 2, 1-9.

Hanurav, T. V. (1967). "Optimum utilization of auxiliary information: pPS sampling of two units from a stratum," *Journal of the Royal Statistical Association, Series B*, 29, 374-391.

Harvey, A. C. (1976). "Estimating regression models with multiplicative heteroscedasticity," *Econometrica*, 44, 461-465.

Horvitz, D. G. and Thompson, D. J. (1952). "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, 47, 663-685.

Isaki, C. T. and Fuller, W. A. (1982). "Survey design under the regression superpopulation model." *Journal of the American Statistical Association*, 77, 89-96.

Jessen, R. J. (1969). "Some methods of probability nonreplacement sampling." *Journal of the American Statistical Association*, 64, 175-193.

Kim, S. W., Heeringa, S. G., and Solenberger, P. W. (2003). "A probability sampling approach for variance minimization," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, 2168-2173.

Kim, S. W., Heeringa, S. G., and Solenberger, P. W. (2005). "An empirical comparison of efficiency between optimization and non-optimization probability sampling of two units from a stratum," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, 3211-3219.

Kim, S. W., Heeringa, S. G., and Solenberger, P. W. (2006). "Model-based sampling designs for optimum estimation," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, 3245-3252.

Kish, L. (1965). *Survey Sampling*, New York: Wiley.

Mizuno, H. (1952). "On the sampling system with probability proportional to sum of sizes," *Annals of the Institute of Statistical Mathematics*, 3, 99-107.

Mukhopadhyay, P. and Vijayan, K. (1996). "On controlled sampling designs," *Journal of Statistical Planning and Inference*, 52, 375-378.

Raj, D. (1956). "A note on the determination of optimum probabilities in sampling without replacement," *Sankhya*, 17, 197-200.

Rao, J. N. K. (1963). "On three procedures of unequal probability sampling without replacement," *Journal of the American Statistical Association*, 58, 202-215.

Rao, J. N. K. and Bayless, D. L. (1969). "An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum," *Journal of the American Statistical Association*, 64, 540-559.

Särndal, C. E. and Wright, R. L. (1984). "Cosmetic form of estimators in survey sampling," *Scandinavian Journal of Statistics*, 11, 146-156.

Särndal, C. E., Swensson, B, and Wretman, J. (1992). *Model Assisted Survey Sampling*, New York: Springer.

SAS/OR (2004). *Mathematical Programming*, Version 9.1, Cary, NC: SAS Institute Inc.

Wright, R. L. (1983). "Finite population sampling with multivariate auxiliary information." *Journal of the American Statistical Association*, 78, 879-884.