Survey Methodology, December 1992
Vol. 18, No. 2, pp. 253–262
Statistics Canada

A Sample Allocation Method for Two-Phase Survey Designs

J.B. ARMSTRONG and C.F.J. WU[1]

ABSTRACT

Motivated by a business survey design at Statistics Canada, we formulate the problem of sample allocation for a general two-phase survey design as a constrained nonlinear programming problem. By exploiting its mathematical structure, we propose a solution method that consists of iterations between two subproblems that are computationally much simpler. Using an approximate solution as a starting value, the proposed method works very well in an empirical study.

KEY WORDS: Optimal allocation; Convex programming.

1. INTRODUCTION

The purpose of this paper is to propose a method of sample allocation for two-phase survey designs. Suppose it is necessary to stratify a population of size N into L strata according to an auxiliary variable, z, whose information is not known before sampling. Values of a second auxiliary (size) variable, x, that is correlated with the variable of interest, y, are known for all units in the population. At the first phase of sampling, the population is divided into G strata according to x. An initial sample is drawn from size stratum $g (g = 1, 2, \ldots, G)$, using simple random sampling with sampling fraction v_g, and the z-value for each sampled unit is observed. At the second phase, units in the sample from size stratum g with z-value in class $h (h = 1, 2, \ldots, L)$, are subsampled using sampling fraction v_{gh}. The value of y is observed for units in the second-phase sample.

In the case of no size stratification ($G = 1$) Cochran (1977) gives the allocation that minimizes the variance of the estimate $\hat{Y} = \sum_h \sum_{i \in s2 \cap h} y_i / (v \cdot v_h)$ of the population total $Y = \sum_h N_h \cdot \bar{Y}_h$, subject to a fixed survey cost, C, where N_h and $\bar{Y}_h$ are the population size and population mean, respectively, for stratum h and $\sum_{i \in s2 \cap h} y_i$ denotes the sum of y-values for units in the second phase sample, $s2$, with z-value in class h. If survey estimates are used for analytical purposes, the variance of the estimated total for z class h, $\hat{Y}_h = \sum_{i \in s2 \cap h} y_i / (v \cdot v_h)$, is also of interest. Sedransk (1965), Booth and Sedransk (1969), Rao (1973) and Smith (1989) have studied allocation problems involving the minimization of a function of variances of estimated class totals, subject to a cost constraint.

The method described in this paper can be used to solve the allocation problem for general G when there is a constraint on the variance of the estimated total for each z class. The method was motivated by an application in a business survey conducted by Statistics Canada. The survey involves the sampling of tax records for businesses.

Information about the population of taxfilers is made available to Statistics Canada by Revenue Canada. There is a requirement to produce estimates of financial variables for domains defined by a cross-classification of four-digit Standard Industrial Classification (SIC4) and province. Only two digits of SIC are coded by Revenue Canada with sufficient accuracy. In

[1] J.B. Armstrong, Business Survey Methods Division, Statistics Canada, Ottawa, Ontario K1A 0T6 and C.F.J. Wu, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1.

order to standardize the precision of estimates for SIC4 domains within each province, a two-phase sample design was implemented. The first-phase sample of taxfilers is selected at Revenue Canada using strata defined using SIC2 and gross business income (size). Before the second phase sample is selected, an SIC4 code, considered more accurate than codes available from Revenue Canada, is assigned to each sampled unit by Statistics Canada. Strata defined using SIC4 and size are employed during selection of the second-phase sample. The same size boundaries are used for both phases of sampling. A detailed description of the sample design can be found in Choudhry, Lavallée and Hidiroglou (1989b).

First-phase sample selection is done using Bernoulli sampling (also called Poisson sampling). Suppose that taxfiler i falls in first-phase stratum g within a particular province × SIC2 cell. To determine whether taxfiler i is included in the first-phase sample, a pseudo-random number in the interval (0,1), say R_i, is generated using the taxfiler's unique identification number. The taxfiler is included in the first-phase sample if $R_i \in (0,v_g)$. Bernoulli sampling based on a different set of pseudo-random numbers is used to select the second-phase sample. Using Bernoulli sampling, selection and processing can begin before complete information about the taxfiler universe is available. This advantage of Bernoulli sampling is important, since taxfiler universe information is accumulated over a two-year period. Sample sizes obtained using Bernoulli sampling are random. Choudhry, Lavallée and Hidiroglou (1989b) derive the variance of $\hat{Y}_{h-\mathrm{STRAT}} = \sum_g \sum_{i \epsilon s2 \cap g \cap h} y_i / (v_g \cdot v_{gh})$ using simple random sampling as an approximation to Bernoulli sampling as discussed in Sunter (1986). Under the approximation, a simple random sample of fixed size $n'_g = v_g \cdot N_g$ is selected in size stratum g at the first phase. Let n'_{gh} denote the number of units with SIC4 h in the first-phase sample for size stratum g. At the second phase, a simple random sample of size $n_{gh} = v_{gh} \cdot n'_{gh}$ is selected for SIC4 h and size stratum g, with v_{gh} considered fixed. The variance of $\hat{Y}_{h-\mathrm{STRAT}}$ is given by

$$V_h = \sum_g \left(\frac{1}{v_g \cdot v_{gh}} - 1 \right) \cdot A_{gh} + \sum_g \left(\frac{1}{v_g} - 1 \right) \cdot B_{gh},$$

where

$$A_{gh} = N_{gh} \cdot S^2_{gh},$$

$$B_{gh} = \left(\frac{N_g - N_{gh}}{N_g - 1} \right) \cdot \left(\frac{Y^2_{gh}}{N_{gh}} - S^2_{gh} \right),$$

and S^2_{gh} is the population variance in the second-phase SIC4 × size stratum gh.

The plan of the paper is as follows. In Section 2, the optimal allocation problem is formulated in the context of the two-phase tax sample. An iterative solution procedure, called the exact method, is proposed. Section 3 includes a description of an approximation to the optimal allocation that can be used to obtain starting values for the exact method. The results of an empirical study involving comparison of various starting values for the exact method are reported in Section 4. Section 5 concludes the paper.

2. EXACT METHOD

In this section the optimal allocation problem is described and an iterative solution method, called the exact method, is proposed. To formulate the problem in the context of two-phase tax sampling, it is sufficient to consider one SIC2 cell in a particular province containing N

units. The cost of selecting a unit in the first-phase sample is K_1, regardless of the stratum in which the unit falls, while the cost of selecting a unit in the second-phase sample is K_2, regardless of stratum. Under Bernoulli sampling, the cost function is

$$F^* = K_1 \cdot \sum_g n_g' + K_2 \cdot \sum_g \sum_h n_{gh}.$$

Since sample sizes n_g' and n_{gh} are random, we use the expected cost

$$F = K_1 \cdot \sum_g v_g \cdot N_g + K_2 \cdot \sum_g \sum_h v_g \cdot v_{gh} \cdot N_{gh}. \qquad (1)$$

Rao (1973) and Smith (1989) also solve allocation problems for two-phase sample designs using expected values of random cost functions. In the tax sampling context, the total cost for a province is the sum of the costs for all SIC2 cells within the province. The estimated coefficient of variation of the cost of two-phase tax sampling for the province of Quebec, calculated using 1988 data, was about 1.85%. Coefficients of variation for overall (national) costs were smaller.

It is necessary to minimize (1) with respect to v_g, $g = 1, 2 \ldots, G$, and v_{gh}, $g = 1, 2, \ldots, G$, $h = 1, 2, \ldots, H$ under the constraints

$$\sum_g \left(\frac{1}{v_g \cdot v_{gh}} - 1\right) \cdot A_{gh} + \sum_g \left(\frac{1}{v_g} - 1\right) \cdot B_{gh} \leq C_h^2 \cdot Y_h^2, \quad h = 1, 2, \ldots, H, \quad (2)$$

$$0 < v_g \leq 1, \quad g = 1, 2, \ldots, G,$$

$$0 < v_{gh} \leq 1, \quad g = 1, 2, \ldots, G, \quad h = 1, 2, \ldots, H,$$

where C_h denotes the target coefficient of variation for SIC4 domain h.

Attempts at direct solution of this problem using the IMSL (1987) implementation of the successive quadratic programming algorithm of Schittkowski (1985) produced mixed results. The algorithm worked well for problems with small numbers of variables and constraints. However, satisfactory solutions for problems including more than approximately 35 variables or more than approximately 50 constraints could not be obtained.

Some costs obtained using direct application of Schittkowski's algorithm in the tax sampling context are given in Table 1. The algorithm was applied to the allocation problems for some SIC2 cells in the province of Quebec involving large numbers of variables and/or constraints using data for tax year 1988. All first-phase and second-phase sampling fractions were started at one when the direct approach was used. The lowest cost obtained using the method that we call the exact method, which will be described later in this section, is also given. The information in the table indicates that direct use of the IMSL implementation of Schittkowski's algorithm is an inappropriate strategy for SIC2 cells with large numbers of variables and constraints.

The exact method is based on a substantial simplification of the problem defined by (1) and (2) that can be achieved by exploiting its structure. In particular, we divide the problem into two main steps that can be solved iteratively. At the first step, (1) is minimized with respect to v_g, $g = 1, 2, \ldots, G$, conditional on values for all second-phase sampling fractions. This

Table 1

Results for Direct and Exact Methods

SIC 2	No. of variables	No. of constraints	Cost ($) - direct	Cost ($) - exact
30	62	86	5155**	1897
35	37	51	551	512
39	38	50	1667	1450
427*	39	48	27528**	3383

* Three digits of SIC are used for first-phase stratification for construction industries.

** The IMSL routine terminated with an internal error that could not be rectified after consulting published documentation.

step requires the use of nonlinear optimization techniques. The second step involves minimizing (1) with respect to the second-phase sampling fractions, conditional on the values of the first-phase sampling fractions obtained in the first step. No iterations are required for this minimization, since it has a closed form solution. Furthermore, it can be done independently for each $h = 1, 2, \ldots, H$. After completion of the second step, the first step is repeated and the iterative process continued. Convergence is declared when changes in the cost function between consecutive iterations are small.

Let $v_g^{(i)}$ and $v_{gh}^{(i)}$ denote the estimates of the optimal values of v_g and v_{gh} obtained after i iterations (each iteration including one repetition of the two steps described above). At the beginning of iteration $i + 1$, the transformation of variables given by $X_g^{(i+1)} = 1/v_g^{(i+1)} - 1$ is required. This transformation redefines the optimization problem involved in the first step of the iteration as a problem with linear constraints and a convex objective function. Such a convex programming problem is easier to solve.

More precisely, each iteration involves:

(i) Minimization of

$$F = \sum_g \left(N_g + \frac{K_2}{K_1} \sum_h v_{gh}^{(i-1)} \cdot N_{gh}\right) \bigg/ (X_g^{(i)} + 1)$$

with respect to $X_g^{(i)}$, $g = 1, 2, \ldots, G$, subject to the constraints

$$C_h^2 \cdot Y_h^2 - \sum_g \left(\frac{X_g^{(i)} + 1}{v_{gh}^{(i-1)}} - 1\right) \cdot A_{gh} - \sum_g X_g^{(i)} \cdot B_{gh} \geq 0, \quad h = 1, 2, \ldots, H$$

$$X_g^{(i)} \geq 0, \quad g = 1, 2, \ldots, G.$$

(ii) Calculation of $v_g^{(i)} = 1/(X_g^{(i)} + 1), g = 1, 2, \ldots, G$. Minimization, independently for each $h = 1, 2, \ldots, H$, of

$$F_h = \sum_g v_g^{(i)} \cdot v_{gh}^{(i)} \cdot N_{gh}$$

with respect to $v_{gh}^{(i)}$, $g = 1, 2, \ldots, G$, subject to the constraints

$$C_h^2 \cdot \hat{Y}_h^2 - \sum_g \left(\frac{1}{v_g^{(i)} \cdot v_{gh}^{(i)}} - 1 \right) \cdot A_{gh} - \sum_g \left(\frac{1}{v_g^{(i)}} - 1 \right) \cdot B_{gh} \geq 0,$$

$$0 < v_{gh}^{(i)} \leq 1, \quad g = 1, 2, \ldots, G,$$

where h is considered fixed.

It will be shown in Section 3 that solution of step (ii) does not require use of numerical methods. Therefore, the exact method only requires the solution of a series of convex programming problems, each involving only G variables. A convex programming problem is much easier to solve than a general nonlinear programming problem. A local solution of a convex programming problem is also a global solution.

Let $F^{(i)}$ denote the value of the cost function, (1), obtained using $v_g^{(i)}$ and $v_{gh}^{(i)}$. The $F^{(i)}$ values form a monotonically decreasing sequence and therefore converge to a limit. Whether this limit value and the corresponding sampling fractions give the global minimum depends on the starting value. This problem is caused by the geometry of the constraints in (2). In practice one should try several starting values to get the best solution. One starting value is given by the approximate method, which is described in the next section and does not require iterations.

3. APPROXIMATE METHOD

In this section, an allocation method that gives an approximation to the optimal allocation is described. The method was first suggested by Choudhry, Lavallée and Hidiroglou (1989a). Assuming that all the second-phase sampling fractions are equal to one, an approximation to the optimal allocation of the first-phase sample is calculated. Then the second-phase sample is allocated, conditional on the first-phase sampling fractions. Since the cost of sampling a unit in both phases of sampling does not depend on the stratum in which the unit falls, minimizing cost is equivalent to minimizing sample size at each step of this method.

At the first step of the method, an approximate solution to the optimal allocation problem for a one-phase sample design is calculated. This step involves finding the minimum, independently for each h, of

$$F^{(h)} = \sum_g v_{g|h} \cdot N_g \tag{3}$$

with respect to $v_{g|h}$, $g = 1, 2, \ldots, G$. The notation $v_{g|h}$ is used to denote the fact that a sampling fraction for size stratum g is determined subject to only one precision constraint, namely the constraint for SIC4 domain h, where h is fixed. In particular, the minimization must be done subject to the constraints

$$\sum_g \left(\frac{1}{v_{g|h}} - 1 \right) \cdot (A_{gh} + B_{gh}) \leq C_h^2 \cdot Y_h^2, \tag{4}$$

$$0 < v_{g|h} \leq 1, \quad g = 1, 2, \ldots, G. \tag{5}$$

One can show that the minimum of (3) is obtained when (4) holds with equality, so that the problem defined by (3), (4), and (5) is equivalent to finding the critical point of the lagrangian

$$L = \sum_g v_{g|h} N_g + \lambda \cdot \left[C_h^2 \cdot Y_h^2 - \sum_g \left(\frac{1}{v_{g|h}} - 1 \right) \cdot (A_{gh} + B_{gh}) \right].$$

Setting the derivatives with respect to $v_{g|h}$ equal to zero yields

$$v_{g|h} = ((A_{gh} + B_{gh})/N_g)^{1/2} \cdot (-\lambda)^{1/2}, \quad g = 1, 2, \ldots, G. \tag{6}$$

Setting $\partial L / \partial \lambda = 0$ we obtain

$$(-\lambda)^{1/2} = \sum_g ((A_{gh} + B_{gh}) \cdot N_g)^{1/2} \bigg/ \left(C_h^2 \cdot Y_h^2 + \sum_g (A_{gh} + B_{gh}) \right). \tag{7}$$

After substitution of (7) into (6), we obtain the optimal sampling fraction for size stratum g given only one precision constraint, for SIC4 domain h,

$$v_{g|h}^* = ((A_{gh} + B_{gh})/N_g)^{1/2} \cdot$$

$$\sum_g ((A_{gh} + B_{gh}) \cdot N_g)^{1/2} \bigg/ \left(C_h^2 \cdot Y_h^2 + \sum_g (A_{gh} + B_{gh}) \right). \tag{8}$$

If one or more of the sampling fractions given by (8) are greater than one, one can set them equal to one and solve a modified allocation problem with a reduced number of strata. This approach corresponds to the overallocation procedure discussed by Cochran (1977). It is necessary to calculate (8) for $h = 1, 2, \ldots, H$. The approximate first-phase sampling fraction for size stratum g, v_g^*, is set equal to the largest value in the set $\{v_{g|h}^*, h = 1, 2, \ldots, H\}$ for $g = 1, 2, \ldots, G$, an approach that ensures that the precision constraint for each SIC4 domain will be satisified.

Given first-phase sampling fractions, optimal second-phase sampling fractions can be easily determined. Assume that, for the SIC2 × province cell h, the size strata included in the allocation problem correspond to a set of integers, Γ. We set the second-phase sampling fractions equal to one for those size strata that are not included in the allocation problem. Normally, one would have $\Gamma = \{1, 2, \ldots, G\}$ but because of overallocation during allocation of the second-phase sample, for example, Γ may not include all integers between 1 and G. The problem of allocating the second-phase sample is equivalent to the problem of finding the minimum of

$$F_h = \sum_{g \epsilon \Gamma} v_{gh} \cdot v_g^* \cdot N_{gh} \tag{9}$$

with respect to v_{gh}, $g \epsilon \Gamma$, subject to the constraints

$$\sum_{g \epsilon \Gamma} \left(\frac{1}{v_{gh}} - 1 \right) \cdot \frac{A_{gh}}{v_g^*} \leq M_h, \tag{10}$$

$$0 < v_{gh} \leq 1, \quad g \epsilon \Gamma, \tag{11}$$

where

$$M_h = C_h^2 \cdot Y_h^2 - \sum_g \left(\frac{1}{v_g^*} - 1\right) \cdot (A_{gh} + B_{gh}).$$

Note that the expected number of units with SIC4 h in the second-phase sample for size stratum g, $v_g^* \cdot N_{gh}$, is employed in (9). It is easy to show that (9) attains a minimum when the constraint (10) holds with equality. Consequently, the minimization problem is equivalent to finding the critical point of the lagrangian

$$L_h = \sum_{g \in \Gamma} v_{gh} \cdot v_g^* \cdot N_{gh} + \lambda \cdot \left(M_h - \sum_{g \in \Gamma} \cdot \left(\frac{1}{v_{gh}} - 1\right) \cdot \frac{A_{gh}}{v_g^*}\right),$$

with respect to and v_{gh}, $g \in \Gamma$, and λ, subject to the constraints

$$0 < v_{gh} \leq 1, \quad g \in \Gamma.$$

Setting the first derivatives of L_h equal to zero and simplifying, one obtains

$$v_{gh} = (-\lambda \cdot A_{gh}/N_{gh})^{1/2} \cdot (1/v_g^*), \quad g \in \Gamma, \tag{12}$$

$$(-\lambda)^{1/2} = \sum_g (N_{gh} \cdot A_{gh})^{1/2} / D_{\Gamma h}, \tag{13}$$

where

$$D_{\Gamma h} = C_h^2 \cdot Y_h^2 \sum_{g \in \Gamma} \left(\frac{1}{v_g^*}\right) \cdot A_{gh} - \sum_g \left(\frac{1}{v_g^*} - 1\right) \cdot (A_{gh} + B_{gh}).$$

Note that there is no solution to the allocation problem unless $D_{\Gamma h}$ is positive. Substituting (13) into (12) yields

$$v_{gh}^* = (A_{gh}/N_{gh})^{1/2} \cdot (1/v_g^*) \cdot \sum_{g \in \Gamma} (N_{gh} \cdot A_{gh})^{1/2} / D_{\Gamma h}. \tag{14}$$

If v_{gh}^* is greater than one for certain gh, the overallocation procedure described above can obviously be employed. Note that (14) also provides the solution for step (ii) of each exact method iteration.

4. EMPIRICAL STUDY

The approximate method serves two purposes. First, it provides a good starting value for the exact method. Second, it may be easier to implement in practice. In this section, we report the results of an empirical comparison using data from the province of Quebec for tax year 1988. Results obtained using the exact method with various starting points, as well as the approximate method, are reported. Since the quantities N_{gh}, Y_h and S_{gh}^2 required by both methods were unknown, estimates based on the data were used.

The size stratification used by the survey, including four take-some strata and one take-all stratum, was employed. Allocations were computed for 64 SIC2 cells (all of the Quebec data excluding a few small SIC2s). The number of sampling fractions determined in these allocations ranged from 8 to 92 with a median of 24. The number of constraints ranged from 9 to 115 with a median of 31. There were 20 SIC2 cells involving more than 35 variables and 18 of these cells also involved more than 50 constraints. A total of 1850 second-phase strata including about 230,000 population units were involved.

The first-phase sampling cost, corresponding to the cost of microfilming or photocopying a tax return at Revenue Canada, sending the information to Statistics Canada and determining an SIC4 code, was set at $1.40 per unit. The second-phase sampling cost, corresponding to the cost of transcribing values for financial variables, was set at $7.00. These costs are comparable to those incurred during operation of the actual survey.

Allocations were computed using the exact method with three starting values: I - solution of the approximate method; II - all first-phase sampling fractions set to one with the corresponding conditionally optimal second-phase fractions; and III - a randomly chosen set of feasible first-phase sampling fractions, with the corresponding conditionally optimal second-phase fractions. In addition, the exact method was started at a perturbation of each of these starting values. The perturbed value for the first-phase sampling fraction for size stratum g for starting value I was $v_g^{(0)} = 0.1 + 0.9 \cdot v_g^*$, where v_g^* is the solution of the approximate method. Second-phase sampling fractions were started at values that are optimal, conditional on the perturbed first-phase fractions. Starting value III was perturbed analogously. The perturbed value corresponding to starting value II was $v_{gh}^{(0)} = 0.1 + 0.9 \cdot v_{gh}^{**}$, where v_{gh}^{**} is optimal, conditional on a census at the first phase of sampling. For each starting value, the best result obtained using either the value itself or the corresponding perturbed value was retained. Convergence was declared if the absolute relative change in the cost function between consecutive iterations was less than 10^{-4}. The IMSL implementation of Schittkowski's successive quadratic programming algorithm was used to solve nonlinear programming problems.

Results are reported in Table 2. Total costs for four alternatives are given. In addition, the number of SIC2 cells for which each starting value for the exact method produced better results than alternative starting values is shown. Computing costs are not reported, since they were small enough to be inconsequential.

The results indicate that the approximate solution provided the best starting values for the exact method. Although starting value II produced better results than starting value I for 17 SIC2 cells, the total cost associated with starting value II was higher than the total cost for the approximate method. The exact method performed poorly when starting values were determined by random selection of a feasible set of first-phase sampling fractions.

Table 2

Results for Exact and Approximate Methods

Method	Exact - Starting value			Approximate
	I	II	III	
Total cost ($)	122779	139347	200998	130228
No. cells with best result*	48	17	1	

* For two cells starting values I and II produced the same result, which had lower cost than the result obtained using starting value III. Consequently, the numbers reported in this row of the table add to 66 rather than 64.

Although the total cost using the exact method with starting value I was only 5.7% lower than the cost of the approximate method, it should be noted that the exact method with starting value I can do no worse than the approximate method. The exact method with starting value I produced better results than the approximate method for 42 cells.

5. CONCLUSION

A sample allocation problem for two-phase survey designs is formulated as a constrained optimization problem in Sections 1 and 2. If the numbers of variables and constraints involved in the problem are small, the solution can be obtained through direct application of numerical methods. However, the direct approach does not work well for large numbers of variables and constraints.

By exploiting the mathematical structure of the problem, it can be divided into two subproblems: the first is a convex programming problem with linear constraints that involves a much smaller number of variables, and the second can be solved without the use of numerical methods. The algorithm proposed in Section 2 consists of iterations between the two subproblems. It is computationally simpler and more effective in practice than the direct approach for problems involving large numbers of variables and constraints. An approximate solution to the sample allocation problem that does not require use of numerical methods is proposed in Section 3. The empirical study in Section 4 shows that it works especially well as a starting value for the algorithm proposed in Section 2.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the work of Pierre Lavallée, who was the first to derive the expressions for conditionally optimal second-phase sampling fractions involved in the approximate method. Thanks are due to a referee and an associate editor for useful comments. C.F.J. Wu is supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

ARMSTRONG, J.B., BLOCK, C., and SRINATH, K.P. (1991). Two-phase sampling of tax records for business surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 228-233.

BOOTH, G., and SEDRANSK, J. (1969). Planning some two-factor comparative surveys. *Journal of the American Statistical Association*, 64, 560-573.

CHOUDHRY, G.H., LAVALLÉE, P., and HIDIROGLOU, M. (1989a). Two-phase sample design for tax data. Unpublished document, Business Survey Methods Division, Statistics Canada.

CHOUDHRY, G.H., LAVALLÉE, P., and HIDIROGLOU, M. (1989b). Two-phase sample design for tax data. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 646-651.

COCHRAN, W.G. (1977). *Sampling Techniques*, (3rd Ed.). New York: John Wiley.

IMSL (1987). Math/Library FORTRAN Subroutines for Mathematical Applications. Houston: IMSL Inc.

RAO, J.N.K. (1973). On double sampling for stratification and analytical surveys. *Biometrika*, 60, 125-133.

SCHITTKOWSKI, K. (1985). NLPQL: A FORTRAN subroutine solving constrained nonlinear programming problems. *Annals of Operations Research*, 5, 485-500.

SEDRANSK, J. (1965). A double sampling scheme for analytical surveys. *Journal of the American Statistical Association*, 60, 985-1004.

SMITH, P.J. (1989). Is two-phase sampling really better for estimating age composition? *Journal of the American Statistical Association*, 84, 916-921.

SUNTER, A.B. (1986). Implicit longitudinal sampling from administrative files: a useful technique. *Journal of Official Statistics*, 2, 161-168.