GENERALIZED LEAST SQUARES ESTIMATION UNDER POSTSTRATIFICATION

Michael D. Bankier, Statistics Canada
Social Survey Methods Division, Ottawa, Ontario K1A 0T6, Canada

KEY WORDS: Raking ratio estimator, multiple regression estimator

1. Introduction

In the Canadian Census, in addition to the basic demographic questions asked of all persons, there are a number of supplementary questions dealing with cultural and socio-economic characteristics. Except in remote areas of the country, these supplementary questions are asked of those persons who fall in a 1 in 5 sample of households selected systematically from each Enumeration Area (EA). In both Canada and the U.S.A., published Census sample data have utilized univariate Raking Ratio Estimators (RREs). RREs generally have smaller mean square errors (MSEs) than Horvitz-Thompson (H-T) estimators (see Brackstone and Rao (1979)). For the basic demographic characteristics, RREs also reduce or eliminate inconsistencies between known population counts and the corresponding sample based estimates.

For the 1991 Canadian Census, alternatives to RREs such as multivariate Generalized Least Squares Estimators (GLSEs) are being examined. GLSEs were first described by Deming and Stephan (1940). Renewed interest in GLSEs has been generated by Bethlehem and Keller (1987) and Luery (1986). Articles by authors such as Wright (1983) and Sarndal and Hidiroglou (1989) are also relevant because the GLSE is a multiple regression estimator that is approximately design unbiased.

In this paper, the GLSE theory as described by Bethlehem and Keller (1987) is generalized to allow for poststratification. The multivariate GLSE can then be compared more directly to the poststratified univariate RREs traditionally used in Censuses. The results of a numerical example are given which indicate that the estimated MSEs of Census estimators can be significantly reduced by using the multivariate GLSEs.

2. Multivariate GLSE and RRE

Assume a sample of fixed size n is selected randomly, without replacement, by some method from a population of N units. It will be assumed that each unit in the population has a non-zero probability of being selected in the sample and that the joint probability of selection for any two units is non-zero. The population and sample are partitioned into H poststrata or weighting adjustment classes. An initial unbiased weight $W_{hk}^{(0)}$ is assigned to the kth sampled unit (or subunit, as appropriate) that belongs to the hth poststratum or weighting adjustment class. It is desired to determine final weights $W_{hk} = c_h W_{hk}^{(0)}$ such that the loss function

$$L_{GM} = (\mathbf{c}-\mathbf{1})'\mathbf{V}(\mathbf{c}-\mathbf{1}) \qquad (2.1)$$

is minimized subject to the constraints

$$\hat{X}_{.i} = X_i \qquad i = 1 \text{ to } I \qquad (2.2)$$

where $\mathbf{c} = [c_h] = [\hat{X}_{h.}/\hat{X}_{h.}^{(0)}]$ and $\mathbf{1} = [1]$ are Hx1 matrices while $\mathbf{V}$ is a HxH symmetric positive definite matrix which is not a function of $\mathbf{c}$. In addition, $\hat{X}_{hi} = \sum_k \hat{x}_{hik}$ while $\hat{x}_{hik} = W_{hk}x_{hik}$ and x_{hik} is the value for the ith type of auxiliary information for the kth unit in the sample in the hth poststratum. $\hat{X}_{hi}^{(0)}$ has the same definition as $\hat{X}_{hi}$ except that the weight W_{hk} is replaced by the weight $W_{hk}^{(0)}$. For $\hat{X}_{hi}$ and $\hat{X}_{hi}^{(0)}$, summation over a subscript is indicated by replacing it with a dot. Finally, $X_i = \sum_k X_{ik}$ where X_{ik} is the value for the ith type of auxiliary information for the kth unit in the population. It will be assumed that the summation in the loss function is restricted to those poststrata where $\hat{X}_{h.}^{(0)}$ does not equal zero.

With the multivariate GLSE, in contrast to the univariate RRE, there does not have to be any relationship between the variables used to define the poststrata or weighting adjustment classes and the variables used to define the constraints. The term weighting adjustment classes is used in this paper to indicate that the H categories are defined using sample information. For example, the weighting adjustment classes can be defined to each contain a single sampled household as was done by Bethlehem and Keller (1987). It can be shown using Lagrange multipliers that the vector $\mathbf{c}$ which minimizes equation (2.1) subject to the constraints of equation (2.2) is

$$\mathbf{c} = \mathbf{1} + \mathbf{V}^{-1}\hat{\mathbf{X}}^{(0)}\lambda \qquad (2.3)$$

with $\lambda = (\hat{X}^{(0)\prime}V^{-1}\hat{X}^{(0)})^{-1}(X'1_N - \hat{X}^{(0)\prime}1)$ where $\hat{X}^{(0)} = [\hat{X}_{hi}^{(0)}]$ is a HxI matrix, X is a N x I matrix with $X' = [X_{ik}]$, λ is a I x 1 matrix and 1_N is a N x 1 matrix of ones. It can be demonstrated that $\hat{X}^{(0)\prime}V^{-1}\hat{X}^{(0)}$ is invertible and hence there is a unique solution c if $\hat{X}^{(0)}$ is of rank I (i.e. the columns of $\hat{X}^{(0)}$ which correspond to the constraints defined in equation (2.2) are linearly independent). It will be assumed in this report that the columns of $\hat{X}^{(0)}$ are linearly independent. This is a reasonable assumption to make since columns of $\hat{X}^{(0)}$ that are linearly dependent are usually caused by constraints that are redundant. In practice, if some of the columns of $\hat{X}^{(0)}$ are found to be linearly dependent, they can be eliminated.

It is possible to determine c by using the Gauss-Seidel method (Pizer (1975)) to iteratively solve for λ. This may be more efficient computationally than the approximately $I^3/3$ multiplications required to perform Gaussian elimination to determine λ. There will only be a gain in computational efficiency, however, if the iterative procedure converges reasonably quickly. It can be shown that the iterative procedure will converge if $\hat{X}^{(0)}$ is of rank I and V is a positive definite matrix.

Having determined the adjusted weights W_{hk}, they are applied to produce estimates for characteristics available only on a sample basis. Let $\hat{Y}^{(0)} = [\hat{Y}_h^{(0)}]$ and $\hat{Y} = [\hat{Y}_h]$ be Hx1 vectors where $\hat{Y}_h^{(0)} = \sum_k W_{hk}^{(0)} y_{hk}$, $\hat{Y}_h = \sum_k W_{hk} y_{hk}$ and y_{hk} is the value for the characteristic of interest for the kth unit in the sample in the hth post-stratum. Then applying equation (2.3)

$$\hat{Y} = \sum_h \hat{Y}_h = \hat{Y}'1 = \hat{Y}^{(0)\prime}c$$
$$= \hat{Y}^{(0)\prime}1 + \hat{\beta}'(X'1_N - \hat{X}^{(0)\prime}1) \quad (2.4)$$

where $\hat{\beta} = (\hat{X}^{(0)\prime}V^{-1}\hat{X}^{(0)})^{-1}\hat{X}^{(0)\prime}V^{-1}\hat{Y}^{(0)} = [\hat{\beta}_i]$ is an Ix1 vector. It can be shown that $\hat{\beta}$ minimizes the loss function

$$L_{GM}^* = (\hat{Y}^{(0)} - \hat{X}^{(0)}\hat{\beta})'V^{-1}(\hat{Y}^{(0)} - \hat{X}^{(0)}\hat{\beta}) \quad (2.5$$

From equations (2.4) and (2.5), it can be seen that the generalized least squares estimator $\hat{Y}$ is a multiple regression estimator that is approximately designed unbiased. It can be shown using a Taylor Series expansion of degree one that $E(\hat{Y}) \simeq Y$ and

$$MSE(\hat{Y}) \simeq V(\hat{Y}) \simeq V(\hat{Y}^{(0)} - \sum_i \beta_i \hat{X}_{.i}^{(0)})$$
$$= V(\sum_h \sum_k^{n_h} W_{hk}^{(0)} u_{hk})$$
$$= V(\hat{U}^{(0)}) \quad (2.6)$$

where $u_{hk} = y_{hk} - \sum_i \beta_i x_{hik}$ and $\beta_i = E(\hat{\beta}_i)$. Thus an approximation of $MSE(\hat{Y})$ can be determined by numerically calculating u_{hk} and then substituting it into the H-T estimator variance formula. An estimator of $MSE(\hat{Y})$ can be determined by replacing the β_i with $\hat{\beta}_i$ when calculating u_{hk} and then substituting the u_{hk} into an estimator of the H-T estimator variance formula.

The definition of the multivariate RRE is identical to the definition of the multivariate GLSE in terms of assumed sample design, poststratification and constraints. The only difference is the loss function

$$L_{RM} = \sum_h \hat{X}_{h.}^{(0)} c_h \ln c_h \quad (2.7)$$

The loss function L_{RM} can be minimized subject to the constraints of equation (2.2) by applying Lagrange multipliers. This generates a system of non-linear equations to be solved. For this reason, Darroch and Ratcliff (1972) developed an iterative solution. They proved that if there is at least one positive solution to the constraints, the iterative process will converge to the unique positive solution which minimizes the L_{RM} loss function.

3. Comparison of GLSE to RRE

In this section, the multivariate GLSE and RRE are compared. Assume that the V matrix of L_{GM} is a diagonal matrix with $\hat{X}_{h.}^{(0)}$ h = 1 to H on the diagonal. In this situation, L_{RM} is approximately equal to L_{GM} times a constant. This can be derived from a Taylor series expansion around $\hat{X}_{h.}^{(0)}$ of degree 2 for L_{RM}. Thus, in this case, the multivariate RRE and GLSE should give similar results if the values of c_h are close to 1. For this V, equation (2.3) shows that

$$c_h = 1 + \sum_i \lambda_i \hat{X}_{hi}^{(0)}/\hat{X}_{h.}^{(0)} \quad (3.1)$$

for the multivariate GLSE. For the multivariate RRE, c_h is replaced by $\ln(c_h)$ in equation (3.1). Thus with the RRE, a log linear model is being used. From these two equations, it is obvious that negative weights are possible with the GLSE but not with the RRE. With either procedure, weights less than one can be generated. This may indicate that either the sample is unrepresentative of the population or that too stringent constraints are being applied to the estimates. If either situation holds, it should be investigated and modifications made regardless of whether the RRE or the GLSE are being used.

One advantage of the GLSE is that it allows greater flexibility than the RRE in terms of the loss function used. The derivation of a solution for the multivariate RRE becomes very complex if a loss function more general than that given in equation (2.7) is used. Assume that poststrata have been defined and that V is a diagonal matrix with $(\hat{X}_{h.}^{(0)})^{2q}$ running down the diagonal and $0 \leq q \leq 1$. It will be also assumed that $(\hat{X}_{h.}^{(0)})^{2q}$ takes on a large range of values, varying from small to large for the various poststrata h. In the discussion which follows, post-strata with small values of $(\hat{X}_{h.}^{(0)})^{2q}$ will be called small poststrata while those with large values of $(\hat{X}_{h.}^{(0)})^{2q}$ will be called large poststrata. Because poststrata are weighted by $(\hat{X}_{h.}^{(0)})^{2q}$ in loss function L_{GM}, large poststrata will tend to have values of c_h close to 1. Based on this $\hat{Y}_h \approx \hat{Y}_h^{(0)}$ for large poststrata, i.e. approximately design based estimators will be used. Choosing values of q closer to 1 rather than 0 will tend to accentuate this pattern. If one has faith that the model assumed under the multiple regression estimator is approximately correct for all poststrata, values of q closer to 0 might be chosen since more reliable estimators, particularly for the smaller poststrata may result. The use of these weights $(\hat{X}_{h.}^{(0)})^{2q}$ is similar to that suggested by Bankier (1988). Oh and Scheuren (1987) have suggested using a combination of separate and raking ratio estimators in this situation instead.

In the Canadian Census, the iterative solution of the univariate RRE person weights can take many iterations to converge. Up to 160 adjustments to the weights are done. It was found that residual discrepancies still existed between the population values and the final estimates for a few constraints. Some experiments carried out with multivariate RRE suggest that it may be even slower to converge. For these reasons, the exact solution to the GLSE may, in certain situations, be computationally more efficient than the iterative solution to the RRE. The exact solution to the GLSE will also guarantee the elimination of any discrepancies between the population values and the final estimates for the constraints. The exact solution to the GLSE also offers the possiblity of simpler variance calculations than have been possible under the RRE (see Bankier (1986) and Binder and Theberge (1988) for a description of RRE variances).

4. Numerical Example

In this section, multivariate GLSEs are calculated using 1986 Canadian Census data. The efficiency of these estimators is compared to that for estimators similar to those used in the 1986 Census.

The 1986 Census estimators will be briefly described. A 1 in 5 sample of households was selected systematically from each EA in non-remote areas. Adjacent EAs were formed into groups called weighting areas (WAs). A WA generally contained 2000-7000 persons. Calculation of weights was carried out independently within each of the 5,347 WAs in Canada. Households and persons were poststratified separately at the WA level by defining two dimensional cross-classification (c-c) matrices. The household c-c matrix had 25 rows and 4 columns. The first row contained two person family households with a male head aged 15-24. The first column contained households in single detached owned dwellings. The person c-c matrix had 26 rows and 27 columns. The first row contained never married males aged 0-4. The first column contained husbands without children whose mother tongue was English. Other rows and columns were defined in a similar fashion. Rows and columns were combined where necessary to ensure that the sample and population counts were large enough for the weighting algorithm to work well. Then univariate RREs were determined such that the estimated number of persons or households which fell in each collapsed row or column agreed with the known population count. Household weights were used to produce the household estimates while person weights were used to produce the person estimates.

For the numerical example, univariate GLSEs (see, for example, Friedlander (1961)) based on the above collapsed person c-c matrix were calculated. The univariate GLSE was used instead of the univariate RRE because it was computationally more convenient. The estimated CVs for the two types of univariate estimators were found to be almost identical for a number of estimates that were compared.

The multivariate GLSEs were designed to do better, according to two criteria, than the 1986 Census estimators for the majority of estimates produced. It was desired that there be lower coefficients of variation (CVs) and smaller discrepancies between estimates and known population counts, particularly at the EA level. In the 1986 Census, there was no control on discrepancies at the EA level, which seemed undesirable given the increased interest in small area data. The weighting methodology also had to be highly automated and require a minimum amount of manual intervention given the large amount of data being processed.

A WA from a rural area of the province of Alberta was selected for the numerical example. The characteristics of the persons and housing in that WA were typical of its county. The WA contained 1797 private households with 4914 persons and 355 of these households were sampled. There were ten EAs in the WA. Four were small with between 16 and 26 households. The other six EAs had between 245 and 328 households.

For the numerical example, it seems reasonable that a single weight associated with each household be used to produce all person, household and family estimates. Besides being computationally efficient, a single household weight reflects the fact that the units sampled are households. Also, if the sample is unrepresentative for certain household characteristics, person estimates should be adjusted to reflect this.

It was decided to apply the multivariate GLSE twice. The first step of the estimation procedure reduced the discrepancies between certain estimates and the corresponding population counts at the EA level. The second step eliminated the discrepancies between these estimates and the corresponding population counts at the WA level and for a few other characteristics at the EA level. The constraints applied at the EA level in the first step were the same constraints applied at the WA level in the second step. It was hoped that this two step procedure would eliminate discrepancies at the WA level while generally reducing them at the EA level. Weighting adjustment classes consisting of single sampled households were used rather than poststrata. One step GLSEs were tried where each EA was poststratified based on household characteristics. It was found, however, that this was not effective in reducing EA level discrepancies and was also difficult to automate.

Sixty constraints were defined for the numerical example. It was required that the estimated number of persons equal the known number at the WA level for various values of age, sex, marital status, mother tongue and family status. For households, the constraints at the WA level applied to the age, sex and marital status of the head of household. It was also required that there be consistency at the WA level for household size, tenure and dwelling type. In addition, the estimated numbers of census and economic families were required to equal the known population counts at the WA level. At the EA level, the estimated number of households was to equal the known number. Also, the estimated and known number of persons was to agree for the eight EAs with the largest samples. In this numerical example, constraints were based on single characteristics such as age rather than two or more crossed characteristics such as age by sex. In the 1986 Census, extensive collapsing of rows and columns of the c-c matrices caused discrepancies to remain for both single and crossed characteristics.

More details are provided now about the two step GLSE procedure. In the first step, initial weights were set equal to the inverse of the achieved EA household sampling fraction. Four subsets of the 42 WA level constraints were applied separately to each of the six largest EAs. The four GLSE weights generated for each household were averaged. In the second step, these averaged weights were used as the initial weights when the GLSEs were calculated at the WA level using the sixty WA and EA level constraints.

Before creating the four subsets of constraints in the first step, any constraints that only applied to a small number of sampled households in an EA were dropped. It was felt that very large or very small weights might otherwise result. In addition, any constraints that were linearly dependent, based on the $\hat{X}^{(0)}$ matrix for a particular EA were dropped. This was done to ensure a unique solution. The number of constraints dropped to between 21 and 27 after these checks. The constraints were then subdivided into four overlapping subsets for each of the six largest EAs. Each of the twelve constraints with the largest discrepancies between the initial estimate and known population count at the EA level appeared in two of the subsets. There were six of these constraints with large discrepancies in each of the subsets. Each of the other constraints appeared in only one of the subsets. Thus there were between two and four of the other constraints in each of the subsets. This approach was taken because it was felt that the number of sampled households at the EA level was insufficient to allow all of the linearly independent constraints to be applied simultaneously. Also, reduction rather than elimination of the discrepancies for these constraints was required at the EA level. The constraints were also checked at the WA level in the second step for linear dependence and for small samples

The L_{GM} loss function applied in both steps used a diagonal V matrix with $\hat{x}_{h.}^{(0)}$ h = 1 to H on the diagonal where H equals the number of households in an EA or the WA as appropriate. If a household had a weight assigned by the GLSE that was negative or greater than 25, then the size measure $\hat{x}_{h.}^{(0)}$ for that household in the V matrix was doubled and the weights were recalculated. This process was repeated up to ten times. In step 1, two negative weights remained for two of the subsets after this was done. When averaged, however, none of the weights were negative. In step 2, there were four negative weights initially. The V matrix had to be adjusted eight times to achieve positive weights. One household had its size factor in the V matrix become 16 $\hat{X}_{h.}^{(0)}$ in order to have its weight become nonnegative.

For the 42 constraints applied at the WA level in the second step , it was desired to examine the level of discrepancies at the EA level using the two step GLSE weights. The percentage increase or decrease of the absolute value of the discrepancies compared to the H-T estimator was calculated. For the two step GLSEs, 55% of the discrepancies decreased by 10% or more, while 35% increased by 10% or more for the six largest EAs. One step GLSEs were also calculated where the first step to reduce discrepancies at the EA level was eliminated. In this case, it was found for the six largest EAs that 40% of the discrepancies decreased by 10% or more, while 49% increased by 10% or more. Thus the two step GLSE does a much better job reducing discrepancies at the EA level.

Estimated CVs of two step multivariate GLSEs were calculated assuming that the systematic sample selected in each EA was equivalent to a simple random sample. This was done for 96 estimators of characteristics such as number of persons classified by age and highest degree or number of persons classified by census family status and total income. Estimated CVs of one step univariate GLSEs based on the 1986 collapsed person c-c matrix were also calculated. The estimated CVs for two step multivariate GLSEs when compared to the one step univariate GLSEs were lower for approximately 85% of the estimators. They were lower by at least 10% for 65% of the estimators.

Estimates of highest degree and income cross-classified by EA were calculated with the four smallest EAs grouped together. Percentage reductions in the estimated CVs were the same at the EA level as at the WA level.

5. MSE of Two Step GLSE

In this section, the approximate MSE of the two step GLSE is derived. Assume that there are G EAs within the WA. The number of households in the gth EA will be represented by N_g while the number sampled will be represented by n_g. The initial weight for the gth EA is $W_g^{(0)} = N_g/n_g$ for g = 1 to G. The gth EA is split into H_g weighting adjustment classes which each contain a single sampled household. Thus $H_g = n_g$.

The calculation of the two step GLSE weights is described first. Let I_g represent the number of constraints that remain in the first step for the gth EA after constraints are eliminated for linear dependence or small samples. Let T_{gr}, r = 1 to 4, represent the four overlapping subsets of constraints for the gth EA. The number of constraints in T_{gr} will be represented by I_{gr}. For each subset T_{gr}, weights $W_{gh}^{(Rr)} = c_{gh}^{(Rr)} W_g^{(0)}$ are determined in the first step such that the loss function

$$L_g^{(Rr)} = \sum_h \hat{X}_{gh.}^{(0)} (c_{gh}^{(Rr)} - 1)^2 \quad (5.1)$$

is minimized subject to the constraints

$$\hat{X}_{g.i}^{(Rr)} = X_{gi},\ i = 1 \text{ to } I_{gr} \quad (5.2)$$

where $\hat{X}_{ghi}^{(Rr)} = c_{gh}^{(Rr)} \hat{X}_{ghi}^{(0)}$, $\hat{X}_{ghi}^{(0)} = W_g^{(0)} x_{ghi}$, x_{ghi} = the value for the ith type of auxiliary information for the hth sampled household from the gth EA, $X_{gi} = \sum_k X_{gik}$ and X_{gik} = the value for the ith type of auxiliary information for the kth unit in the population of the gth EA. In the above expressions, summation over a subscript is indicated by replacing it with a dot. Next, the average weight

$$W_{gh}^{(A)} = (1/4) W_g^{(0)} \sum_r c_{gh}^{(Rr)} \quad (5.3)$$

is calculated. For small EAs which do not have the first step applied $W_{gh}^{(A)} = W_g^{(0)}$.

In the second step, GLSE weights $W_{gh} = c_{gh} W_{gh}^{(A)}$ are determined such that the loss function

$$L_{GM} = \sum_g \sum_h \hat{X}_{gh.}^{(A)} (c_{gh} - 1)^2 \quad (5.4)$$

is minimized subject to the constraints

$$\hat{X}_{..i} = X_{.i},\ \text{for } i = 1 \text{ to } I \quad (5.5)$$

where $\hat{X}_{ghi} = c_{gh} \hat{X}_{ghi}^{(A)}$, $\hat{X}_{ghi}^{(A)} = W_{gh}^{(A)} x_{ghi}$ and $X_{.i} = \sum_g X_{gi}$.

Let $\hat{Y} = \sum_g \sum_h W_{gh} y_{gh}$ represent the two step GLSE for the sample characteristic of interest. It can be shown using a Taylor series expansion of degree one that

$$\hat{Y} = \hat{Y}^{(A)} + \sum_i \hat{\beta}_i (X_{.i} - \hat{X}_{..i}^{(A)}) \simeq$$
$$\hat{Y}^{(A)} + \sum_i \beta_i (X_{.i} - \hat{X}_{..i}^{(A)}) \quad (5.6)$$

where $\hat{\beta} = [\hat{\beta}_i] = (\hat{\mathbf{X}}^{(A)\prime} \mathbf{V}^{-1} \hat{\mathbf{X}}^{(A)})^{-1} \hat{\mathbf{X}}^{(A)\prime} \mathbf{V}^{-1} \hat{\mathbf{Y}}^{(A)}$ and $\beta_i = E(\hat{\beta}_i)$. It can then be shown that $E(\hat{Y}) \simeq Y$ and

$$MSE(\hat{Y}) \simeq V(\ \hat{Y}^{(A)} - \sum_i \beta_i \hat{X}_{..i}^{(A)})$$

$$= V(\sum_g \sum_h^{n_g} W_{gh}^{(A)} u_{gh})$$
$$= V(\hat{U}^{(A)}) \qquad (5.7)$$

where $u_{gh} = y_{gh} - \sum_i \beta_i x_{ghi}$.

Next, it should be noted that $\hat{U}^{(A)} = (1/4)\sum_g \sum_r \hat{U}_g^{(Rr)}$. Using a Taylor series expansion of degree one

$$\hat{U}_g^{(Rr)} =$$
$$\hat{U}_g^{(0)} + \sum_i \hat{\beta}_{gi}^{(Rr)}(X_{gi} - \hat{X}_{g.i}^{(0)}) \simeq$$
$$\hat{U}_g^{(0)} + \sum_i \beta_{gi}^{(Rr)}(X_{gi} - \hat{X}_{g.i}^{(0)}) \qquad (5.8)$$

where $\hat{\beta}_g^{(Rr)} = [\hat{\beta}_{gi}^{(Rr)}] = (\hat{X}_{gr}^{(0)\prime} V_g^{-1} \hat{X}_{gr}^{(0)})^{-1} \hat{X}_{gr}^{(0)\prime} V_g^{-1} \hat{U}_g^{(0)}$ is a I_{gr} x 1 vector, the n_g x I_{gr} matrix $\hat{X}_{gr}^{(0)}$ contains columns for those constraints in subset T_{gr}, the n_g x n_g diagonal matrix V_g has $\hat{X}_{gh.}^{(0)}$ h = 1 to n_g running down the diagonal and $\beta_{gi}^{(Rr)} = E(\hat{\beta}_{gi}^{(Rr)})$. Then

$$V(\hat{U}^{(A)}) \simeq$$
$$= V(\sum_g(\hat{U}_g^{(0)} - \sum_i \beta_{gi}^{(A)} \hat{X}_{g.i}^{(0)}))$$
$$= V(\sum_g \sum_h^{n_g} W_{gh}^{(0)} z_{gh})$$
$$= V(\hat{Z}^{(0)}) \qquad (5.9)$$

where $\beta_{gi}^{(A)} = (1/4) \sum_r \beta_{gi}^{(Rr)}$ and $z_{gh} = u_{gh} - \sum_i \beta_{gi}^{(A)} x_{ghi}$. When $\beta_{gi}^{(A)}$ is calculated, it is assumed that $\beta_{gi}^{(Rr)} = 0$ if constraint i is not in subset T_{gr}.

To calculate a sample based estimate of $V(\hat{Z}^{(0)})$, $\hat{\beta}_i$ is substituted for β_i in the expression for u_{gh} while $\hat{\beta}_{gi}^{(A)}$ is substituted for $\beta_{gi}^{(A)}$ in the expression for z_{gh}. The values of z_{gh} are then substituted into the unbiased estimator of the H-T variance formula for stratified simple random sampling without replacement.

6. Concluding Remarks

Based on the numerical example, the multivariate GLSE procedure shows great potential, when compared to the univariate RRE procedure, for generating Census estimates with lower CVs and smaller discrepancies between estimates and known population counts. Further study is required, however, before a decision can be made on the 1991 Canadian Census weighting procedure. The procedure implemented must be computationally efficient while providing lower CVs and smaller discrepancies. It must maintain this performance over a large number of WAs with varying characteristics. Modified versions of the multivariate GLSEs will be applied to a number of other WAs. The performance of the multivariate RREs will also be investigated. Preliminary work, however, indicates that the multivariate RREs are slow to converge.

References

Bankier, M. D. (1986), "Estimators Based on Several Stratified Samples With Applications to Multiple Frame Surveys", Journal of the American Statistical Association, 81, 1074-1079.

Bankier, M. D. (1988), "Power Allocations: Determining Sample Sizes For Sub-National Areas", American Statistician, 42, 174-177.

Bethlehem, Jelke G. and Keller, Wouter J. (1987), "Linear Weighting of Sample Survey Data", Journal of Official Statistics, 3, 141-153.

Binder, D.A. and Theberge, A. (1988), "Estimating the Variance of Raking Ratio Estimators", Canadian Journal of Statistics, Vol. 16 Supplement, 47-56.

Brackstone, G.J., and Rao, J.N.K. (1979), "An Investigation of Raking Ratio Estimators", Sankhya, Ser. C, 41, 97-114.

Darroch, J.N., and Ratcliff, D. (1972), "Generalized Iterative Scaling For Log-Linear Models", Annals of Mathematical Statistics, 43, 1470-1480.

Deming, W.E. and Stephan, F.F. (1940), "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known", Annals of Mathematical Statistics, 34, 911-934.

Friedlander, D. (1961), "A Technique for Estimating a Contingency Table Given the Marginal Totals and Some Supplementary Data", Journal of the Royal Statistical Society, Ser. A, 124, 412-420.

Luery, Donald M. (1986), "Weighting Sample Survey Data Under Linear Constraints on the Weights", in Proceedings of the Social Statistics Section, American Statistical Association, 325-330.

Oh, H. Lock and Scheuren, F. (1987), "Modified Raking Ratio Estimation", Survey Methodology, Vol. 13, No. 2, 209-219.

Pizer, Stephen M. (1975), Numerical Computing and Methematical Analysis, Science Research Associates, Inc.

Sarndal, Carl -Erik, and Hidiroglou, Michael A. (1989), "Small Domain Estimation: A Conditional Analysis", to appear in Journal of the American Statistical Association.

Wright, Roger L. (1983), "Finite Population Sampling with Multivariate Auxiliary Information", Journal of the American Statistical Association, 78, 879-884.