TWO STEP GENERALIZED LEAST SQUARES ESTIMATION IN THE 1991 CANADIAN CENSUS

Michael D. Bankier, Stephen Rathwell and Mark Majkowski, Statistics Canada
Stephen Rathwell, 11th Floor, R.H. Coats Bldg, Tunney's Pasture, Ottawa, Canada, K1A 0T6

KEY WORDS: Regression Estimation, Raking Ratio Estimation

1. INTRODUCTION

In the 1991 Canadian Census, a 1 in 5 systematic sample of private households was selected from each of 40,072 enumeration areas (EAs). Sampled EAs contained on average 249 households. Besides the basic demographic questions asked of all households, sampled households were required to answer additional questions.

In the 1986 Census, raking ratio (RR) estimation generated sample weights that ensured agreement between certain sample estimates and known population counts at the weighting area (WA) level. In both 1986 and 1991, WAs contained on average 7 EAs. RR estimators generally have smaller variances than estimators based on weights equal to the inverse of the probability of selection (see, for example, Brackstone and Rao 1979). However, residual differences remained between some sample estimates and population counts because the RR iterative solution (as proposed by Deming and Stephan 1940) had not completely converged after 40 cycles. Also, the agreement between sample estimates and population counts with RR estimates was usually no better at the EA level than for estimates calculated using the initial weights. Finally, because different weights were used to produce household and person estimates, this caused discrepancies between these estimates in certain cases.

For the 1991 Canadian Census, two step generalized least squares (GLS) estimation was used. GLS estimation is a form of regression estimation. The Census weights were adjusted in two steps because this made it possible to achieve reasonable consistency between sample estimates and population counts at the EA level. At the same time, the variance of the two step GLS estimator was significantly lower than that of the 1986 Census estimator at the EA level and somewhat lower at the WA level. This was important because EAs are the basic building blocks for tabulations of larger geographical areas. GLS estimation was used because its methodology is well known and well accepted. In addition, GLS estimation had a non-iterative solution so there are no problems with lack of convergence.

Besides illustrating a major application of GLS estimation, this article describes an effective method for eliminating GLS weights less than 1 by discarding constraints. A new method of identifying nearly collinear constraints so they can be discarded is described. The results of a Monte Carlo study are reported which provide an assessment of the size of the bias of the two step GLS estimators as well as of the bias for different estimators of the variance. The performance of this method is then evaluated by applying it to 79 WAs from the 1986 Census.

2. THE ONE STEP GLS ESTIMATION TECHNIQUE

Sample weights are calculated separately for each WA. In a particular WA, assume that there are G sampled EAs. In order to simplify the variance formulae, it will be assumed that a simple random sample of households is selected without replacement from each EA. (Estimated variances under the assumption of systematic sampling are discussed later in the section on the Monte Carlo study.) Let n_g and N_g represent the number of households in the sample and population respectively for the g^{th} EA in the WA. The initial household weight is $W_g^{(0)}=N_g/n_g$. Horvitz-Thompson estimators based on this weight are unbiased.

The basic characteristics for which agreement between sample estimates and population counts is desired are called "constraints". Examples of characteristics for which agreement is required at the WA level are number of persons, number of males, number of persons of age 25 to 29, number of census families, number of households and number of owned dwellings. In addition, agreement is required at the EA level for number of persons and number of households. Characteristics that are used as constraints appear in published Census tabulations. Inconsistencies between the sample estimates and population counts for these characteristics cause concern to users of Census data.

The constraints can be represented by the n x I matrix $\underset{\sim}{x}=[x_{ghi}]$ where n equals the number of sampled households in the WA, I equals the total number of constraints and x_{ghi} represents the value of the i^{th} constraint for the h^{th} sampled household in the g^{th} EA. For example, if the i^{th} constraint is number of males, then $x_{ghi}=3$ indicates that there are 3 males in the h^{th} sampled household of the g^{th} EA. Also, let

$\hat{\underset{\sim}{X}}^{(0)}=diag(\underset{\sim}{W}^{(0)})\,\underset{\sim}{x}=[W_g^{(0)}x_{ghi}]$ where $diag(\underset{\sim}{W}^{(0)})$ is a n x n matrix with $\underset{\sim}{W}^{(0)}$ running down the diagonal with zeros elsewhere. Here $\underset{\sim}{W}^{(0)}$ is a n x 1 vector with $W_g^{(0)}$ the vector entry for each sampled household of the g^{th} EA.

The one step GLS estimator is derived by determining the adjusted weights $W_{gh}=c_{gh}W_g^{(0)}$ such that the distance function

$$D=(\underset{\sim}{c}-\underset{\sim}{1}_n)'\underset{\sim}{V}(\underset{\sim}{c}-\underset{\sim}{1}_n)$$

is minimized subject to the constraints

$$\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{c}=\underset{\sim}{X}$$

where $\underset{\sim}{c}=[c_{gh}]$ is a n x 1 vector of weighting adjustment factors, $\underset{\sim}{1}_n$ is a n x 1 vector of 1's, $\underset{\sim}{X}=[X_{..i}]$ is an I x 1 matrix and $X_{..i}=\sum_{g=1}^{G}\sum_{h=1}^{N_g}x_{ghi}$ is the known population value for the i^{th} constraint. $\underset{\sim}{V}$ has to be positive definite to ensure that the distance measure D is non-negative. In the Canadian Census, $\underset{\sim}{V}=diag(\hat{\underset{\sim}{X}}^{(0)}\underset{\sim}{1}_I)$, where $\underset{\sim}{1}_I$ is an I x 1 vector of 1's. This is consistent with the recommendation of Särndal, Swensson and Wretman (1989) that $\underset{\sim}{V}=diag(\hat{\underset{\sim}{X}}^{(0)}\underset{\sim}{\gamma})$ where $\underset{\sim}{\gamma}$ is an I x 1 vector which does not result in any of the elements of $\hat{\underset{\sim}{X}}^{(0)}\underset{\sim}{\gamma}$ becoming zero. The solution to this problem is

$$\underset{\sim}{c}=\underset{\sim}{1}_n+\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}(\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)})^{-1}(\underset{\sim}{X}-\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{1}_n)$$

It can be seen that the GLS estimator $\hat{Y}=\sum_{g=1}^{G}\sum_{h=1}^{n_g}W_{gh}y_{gh}$ (where $W_{gh}=c_{gh}W_g^{(0)}$) is a regression estimator by noting that $\hat{Y}=\hat{\underset{\sim}{Y}}^{(0)\prime}\underset{\sim}{1}_n+\hat{\underset{\sim}{\beta}}'(\underset{\sim}{X}-\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{1}_n)$ where y_{gh} is the value of the sample characteristic of interest for the h^{th} sampled household in the g^{th} EA, $\hat{\underset{\sim}{Y}}^{(0)}=[W_g^{(0)}y_{gh}]$ is a n x 1 vector and $\hat{\underset{\sim}{\beta}}=(\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)})^{-1}\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{Y}}^{(0)}$ is an I x 1 vector.

It can easily be shown, using a Taylor series approximation, that $E(\hat{Y})\approx Y$ where Y is the true population value for the sample characteristic of interest. It can also be shown, using a Taylor series approximation, that

$MSE(\hat{Y})\approx V(\hat{Y})\approx V(\hat{Z}^{(0)})$ where $\hat{Z}^{(0)}=\sum_{g=1}^{G}\sum_{h=1}^{n_g}W_{gh}^{(0)}z_{gh}$, $\underset{\sim}{z}=\underset{\sim}{y}-\underset{\sim}{x}\underset{\sim}{\beta}=[z_{gh}]$, $\underset{\sim}{y}=[y_{gh}]$ and $\underset{\sim}{\beta}=E(\hat{\underset{\sim}{\beta}})$. An estimator of $MSE(\hat{Y})$ can be determined by replacing the $\underset{\sim}{\beta}$ with $\hat{\underset{\sim}{\beta}}$ when calculating $\underset{\sim}{z}$ and then substituting the $\underset{\sim}{z}$ into the standard estimator for the variance of a stratified Horvitz-Thompson estimator. Hidiroglou, Fuller and Hickman (1978, p.37) and Särndal, Swensson and Wretman (1989) suggest that a more accurate estimate of the variance is produced if $\underset{\sim}{z}^*=[c_{gh}z_{gh}]$ is used instead of $\underset{\sim}{z}$.

3. THE TWO STEP GLS ESTIMATION TECHNIQUE

3.1 An Overview of the Technique

One of the objectives of the Census weighting system is to have reasonably small differences between sample estimates and population counts at the EA level for WA level constraints. Because of the relatively small size of the EAs, it is not practical to eliminate the differences entirely at the EA level. The final Census weights take the form $W_{gh}=c_{gh}c_{gh}^{(A)}W_g^{(0)}$. The first step weighting adjustment $c_{gh}^{(A)}$ is done to reduce population/estimate differences at the EA level. The second step weighting adjustment c_{gh} is done to eliminate population/estimate differences for the constraints at the WA level as well as for the two constraints (number of households and number of persons) for each EA.

3.2 First Step Weighting Adjustment

For each EA, the WA level constraints are listed in descending order based on size. The size of a constraint for an EA is defined to be the number of households in the population for which the constraint applies. For person and family constraints, a household is included in the size count if it contains one or more persons or families to which the constraint applies. For example, the size of the constraint number of males equals the number of households in the population that contain at least one male. Next, these constraints are partitioned into two groups. The first group contains the first, third, fifth, etc. constraints from the list ordered by size. The second group contains the remaining constraints. Separate weighting adjustment factors are calculated for each group and then averaged together. This results in the population/estimate differences being generally reduced but not eliminated at the EA level for the WA level constraints. This approach is taken because the sample size is not large enough at the EA level to have all the constraints applied at once. The partitioning was done on the basis of size so that similar partitionings would result for all possible samples in that WA. This is discussed further in Section 4.

More specifically, for each group (where $r=1,2$ represents the first and second group respectively), adjusted weights $W_{ghr}=c_{ghr}W_g^{(0)}$ (where W_{ghr} equals the adjusted weight for the h^{th} sampled household in the g^{th} EA for the r^{th} group of constraints) are determined where

$$\underset{\sim}{c}_{gr}=\underset{\sim}{1}_g+\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)}(\hat{\underset{\sim}{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)})^{-1}(\underset{\sim}{X}_{gr}-\hat{\underset{\sim}{X}}_{gr}^{(0)\prime}\underset{\sim}{1}_g)$$

$$=[c_{ghr}]$$

$\underset{\sim}{1}_g$ is a n_g x 1 vector of 1's, $\underset{\sim}{V}_{gr}=diag(\hat{\underset{\sim}{X}}_{gr}^{(0)}\underset{\sim}{1}_{I_{gr}})$ is a n_g x n_g matrix, $\underset{\sim}{1}_{I_{gr}}$ is an I_{gr} x 1 vector of 1's and I_{gr} is the number of constraints in the r^{th} group of the g^{th} EA. Also, $\hat{\underset{\sim}{X}}_{gr}^{(0)}=[W_g^{(0)}x_{ghri}]$ is a n_g x I_{gr} matrix, x_{ghri} represents the value of the i^{th} constraint for the r^{th} group and the h^{th} sampled household in the g^{th} EA, $\underset{\sim}{X}_{gr}=[X_{g.ri}]$ is an I_{gr} x 1 vector and $X_{g.ri}=\sum_{h=1}^{N_g}x_{ghri}$ is the known population value for the i^{th} constraint in the r^{th} group in the g^{th} EA.

The weighting adjustment factors c_{ghr} based on these two groups of constraints are then averaged together to produce $W_{gh}^{(A)}=c_{gh}^{(A)}W_g^{(0)}$ where $\underset{\sim}{c}_g^{(A)}=[c_{gh}^{(A)}]=[(c_{gh1}+c_{gh2})/2]$ is a n_g x 1 vector. The averaged weighting adjustment factors $\underset{\sim}{c}_g^{(A)}$ usually reduce but do not eliminate the population/estimate differences for the g^{th} EA.

3.3 Second Step Weighting Adjustment

The final Census weights $W_{gh}=c_{gh}W_{gh}^{(A)}$ are determined by calculating

$$\underset{\sim}{c}=\underset{\sim}{1}_n+\underset{\sim}{V}_A^{-1}\hat{\underset{\sim}{X}}^{(A)}(\hat{\underset{\sim}{X}}^{(A)\prime}\underset{\sim}{V}_A^{-1}\hat{\underset{\sim}{X}}^{(A)})^{-1}(\underset{\sim}{X}-\hat{\underset{\sim}{X}}^{(A)\prime}\underset{\sim}{1}_n)$$

where $\hat{\underset{\sim}{X}}^{(A)}=diag(\underset{\sim}{W}^{(A)})\underset{\sim}{x}$, $\underset{\sim}{W}^{(A)}=[W_{gh}^{(A)}]$ is a n x 1 vector and $\underset{\sim}{V}_A=diag(\hat{\underset{\sim}{X}}^{(A)}\underset{\sim}{1}_I)$.

It can be shown, using two successive Taylor series approximations, that

$MSE(\hat{Y})\approx V(\hat{Y})\approx V(\hat{Z}^{(0)})$ where $\hat{Y}=\sum_{g=1}^{G}\sum_{h=1}^{n_g}W_{gh}y_{gh}$,

$$\hat{Z}^{(0)}=\sum_{g=1}^{G}\sum_{h=1}^{n_g} W_g^{(0)} z_{gh} \text{ and}$$

$\underset{\sim}{z}_g=[z_{gh}]= \underset{\sim}{u}_g-\frac{1}{2}(\underset{\sim}{x}_{g1}\underset{\sim}{\beta}^*_{g1}+\underset{\sim}{x}_{g2}\underset{\sim}{\beta}^*_{g2})$ is a n_g x 1 vector. Also, $\underset{\sim}{u}=\underset{\sim}{y}-\underset{\sim}{x}\underset{\sim}{\beta}=[u_{gh}]$ is a n x 1 vector, $\underset{\sim}{u}_g=[u_{gh}]$ is a n_g x 1 vector containing the elements of $\underset{\sim}{u}$ for the g[th] EA, $\underset{\sim}{\beta}=E(\hat{\underset{\sim}{\beta}})$,

$\hat{\underset{\sim}{\beta}}=(\hat{\underset{\sim}{X}}^{(A)\prime}\underset{\sim}{V}_A^{-1}\hat{\underset{\sim}{X}}^{(A)})^{-1}\hat{\underset{\sim}{X}}^{(A)\prime}\underset{\sim}{V}_A^{-1}\hat{\underset{\sim}{Y}}^{(A)}$ and

$\hat{\underset{\sim}{Y}}^{(A)}=[W_{gh}^{(A)}y_{gh}]$ is a n x 1 vector.

Finally, $\underset{\sim}{\beta}^*_{gr}=E(\hat{\underset{\sim}{\beta}}^*_{gr})$,

$\hat{\underset{\sim}{\beta}}^*_{gr}=(\hat{\underset{\sim}{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)})^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{U}}_g^{(0)}$ and

$\hat{\underset{\sim}{U}}_g^{(0)}=[W_g^{(0)}u_{gh}]$ is a n_g x 1 vector. An estimator of $MSE(\hat{Y})$ can be determined by replacing the $\underset{\sim}{\beta}$ with $\hat{\underset{\sim}{\beta}}$ when calculating $\underset{\sim}{u}$ and replacing the $\underset{\sim}{\beta}^*_{gr}$ with $\hat{\underset{\sim}{\beta}}^*_{gr}$ when calculating $\underset{\sim}{z}_g$. Then $\underset{\sim}{z}'=[\underset{\sim}{z}_1'\underset{\sim}{z}_2'\cdots\underset{\sim}{z}_G']$ can be substituted into the standard estimator for the variance of a stratified Horvitz-Thompson estimator. Alternatively, following the approach of Särndal, Swensson and Wretman (1989), $\underset{\sim}{z}^*=[c_{gh}c_{gh}^{(A)}z_{gh}]$ instead of $\underset{\sim}{z}$ can be used.

4. DISCARDING CONSTRAINTS

4.1 An Overview of the Technique

When calculating the weighting adjustment factors $\underset{\sim}{c}_{gr}$ and $\underset{\sim}{c}$, the matrices $\hat{\underset{\sim}{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)}$ and $\hat{\underset{\sim}{X}}^{(A)\prime}\underset{\sim}{V}_A^{-1}\hat{\underset{\sim}{X}}^{(A)}$ are inverted. Linearly dependent constraints will cause these matrices to be singular. Thus, the smallest constraint (with size defined as the number of households in the population to which it applies) in each set of linearly dependent constraints is dropped.

Next, the condition number of the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ is checked. The condition number is defined as the absolute value of the ratio of the largest eigenvalue to the smallest eigenvalue. Large condition numbers are of concern because small variations in the sample can cause large variations in the weighting adjustment factors. These large variations, in turn, tend to increase the variance of the estimators based on the adjusted weights. Large condition numbers are usually the result (see Pizer 1975, p. 92) of some columns of the matrices being inverted representing hyperplanes that are nearly parallel or, equivalently, the columns are nearly linearly dependent. One technique for identifying groups of nearly linearly dependent columns is described in Chapter 8 of Montgomery and Peck (1982). Another method, described in Subsection 4.2, was found to be more effective at reducing the condition number of the matrix to be inverted without eliminating a large number of constraints.

Having discarded constraints for being nearly linearly dependent, the weighting adjustment factors are calculated. If they result in the adjusted weights falling outside the range [1, 25] (these will be called outlier weights), additional constraints are discarded as described in Subsection 4.2.

Before discarding constraints for being linearly dependent, nearly linearly dependent or causing outlier weights, some constraints are discarded because their size (as defined earlier) is less than 60. This is done to save computational resources since these small constraints are frequently discarded later in processing for one of the other three reasons. In addition, discarding constraints on the basis of size ensures that the same constraints will be discarded for every sample. This is an advantage because the estimator of the variance of the GLS estimator does not take into account the variability introduced by somewhat different constraints being dropped for different samples. This can cause a downward bias in the estimator of the variance as shown in the Monte Carlo study of Section 5. For similar reasons, the two groups of constraints used in the first step weighting adjustment are defined based on size so that similar partitionings will result for all possible samples in that WA.

Because constraints of size less than 60 are discarded, it was decided to combine any EAs with a population of less than 60 households with the smallest EA having a population of 60 or more and treat them as a single EA when calculating the first step weighting adjustment factors.

4.2 Details of Methods Used to Discard Constraints

First, all constraints of size less than 60 are immediately discarded. Next, the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ is calculated. Then this matrix is assessed for linearly dependent constraints. (It can be shown that if a set of columns for the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ are linearly dependent that the corresponding columns of the matrix $\hat{\underset{\sim}{X}}^{(0)}$ are linearly dependent.) The smallest constraint in a set is dropped.

Next, constraints are discarded in order to lower the condition number of the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ To do this, the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ is recalculated based on only the two largest constraints (number of households and number of persons at the WA level), resulting in a 2 x 2 matrix. If the condition number of the $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ matrix exceeds 1,000, the constraint number of persons is discarded. Otherwise, both constraints are retained. Then the next largest constraint is added, the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ is recalculated and its condition number is determined. If the condition number increases by more than 1,000, the constraint just added is discarded. Otherwise, it is retained. This process continues until all constraints have been checked in this fashion. The number 1,000 was selected because it was found to retain a large number of constraints while at the same time significantly reducing the size of the final condition number. If, after dropping these nearly linearly dependent constraints, the condition number of the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ exceeds 10,000 (which rarely happens with Census data), additional constraints are dropped. Constraints are dropped in descending order of the amount by which they increased the condition number when they were initially included in the matrix. The condition number of the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ is recalculated each time a constraint is dropped. When the condition number drops below 10,000, no more constraints are dropped. Any constraints dropped up to this point are not used in the weighting calculations which follow.

Before calculating the first step weighting adjustment factors $\underset{\sim}{c}_{gr}$ for the g[th] EA, the remaining constraints are dropped as necessary because they are small for the g[th] EA. The constraints which remain are partitioned into two groups, as described in Subsection 3.2. Then $\hat{\underset{\sim}{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)}$ is calculated and linearly dependent constraints are identified and dropped (constraints which are linearly dependent at the EA level may not be linearly dependent at the WA level). Based on the remaining constraints, the first step weighting adjustment factors $\underset{\sim}{c}_{gr}$ are calculated. If any of the first step adjusted weights $\tilde{W}_{ghr}$ fall outside the range [1,25], additional constraints are dropped. A method similar to that used to discard nearly linearly dependent constraints is applied here except that a constraint is discarded if it causes outlier weights.

Next, the second step weighting adjustment factors $\underset{\sim}{c}$ are calculated based on those constraints that were not discarded for being small, linearly dependent or nearly linearly dependent based on the initial analysis of the matrix $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$. If any of the second step adjusted weights $W_{gh}=c_{gh}W_{gh}^{(A)}$ fall outside the range [1,25], then additional constraints are dropped from the matrix $\hat{\underset{\sim}{X}}^{(A)\prime}\underset{\sim}{V}_A^{-1}\hat{\underset{\sim}{X}}^{(A)}$ using the method outlined for the first step weighting adjustment.

5. A MONTE CARLO STUDY

A Monte Carlo study was done to assess the size of the bias of the two step GLS estimator as well as the bias of different estimators of the variance. In addition, it was used to finalize the criteria for discarding constraints for being small or nearly linearly dependent. The majority of the constraints are discarded based on properties of the sample rather than the population. Consequently, different constraints can be discarded with different samples. This might cause an increase in the variance of the two step GLS estimator which would not be accounted for in the estimates of variance (and hence downward bias them). Thus the criteria for discarding constraints were chosen to maximize the consistency of the constraints discarded from sample to sample while at the same time retaining as many constraints as possible.

A WA was created for this study from five similar 1986 Census WAs. It contained only sampled households. A random sample of 250 systematic samples for use in the Monte Carlo study was selected without replacement from the WA.

For each selected sample, the two step GLS weights were calculated and applied to produce estimates for 31 EA level and 39 WA level person and household characteristics known only on a sample basis in the Census. All characteristics applied to 60 or more households in the population. For each characteristic, its estimated relative bias (the difference between the average estimate and the population count, expressed as a percentage of the population count) was calculated. The absolute value of the estimated relative bias for the characteristics ranged from almost zero to as high as 4.7%, although it was less than 2% for most characteristics. It was less than 1% for the majority of the characteristics with a population count greater than the median value for the characteristics considered. The bias is similar for EA and WA level characteristics. The estimated standard errors for the estimates of the relative bias were all below 0.2.

For each sample and each characteristic, the estimated variance of the two step GLS estimator was calculated in four ways. First, $\underset{\sim}{z}$ and $\underset{\sim}{z}^*$ were calculated as described in Subsection 3.3. They were then substituted into the standard estimator for the variance of a stratified Horvitz-Thompson estimator. The estimates of the variance which resulted will be labelled $v_h(\underset{\sim}{z})$ and $v_h(\underset{\sim}{z}^*)$ respectively. These two estimators of the variance were calculated under the assumption that a simple random sample was selected from each EA while in reality, a systematic sample was selected from each EA. Wolter (1985, p.250) suggests regarding the systematic sample as a stratified random sample with two households selected from each successive stratum of ten households. $\underset{\sim}{z}$ and $\underset{\sim}{z}^*$ can be substituted into the variance formula which results from making this assumption. The estimates of the variance generated in this way will be labelled $v_s(\underset{\sim}{z})$ and $v_s(\underset{\sim}{z}^*)$ respectively. The estimated relative bias of each of these four estimators of the variance was calculated as the difference between the average value of the estimated variance and an unbiased estimate of the mean square error of the two step GLS estimator, expressed as a percentage of the estimate of the mean square error.

Tables 1a and 1b provide the 10[th], 25[th], 50[th], 75[th], and 90[th] percentiles of the distribution of the estimated relative bias of the variance estimators for WA and EA level characteristics separately. It can be seen that the relative bias is negative for the majority of the WA level characteristics. This is particularly pronounced for $v_s(\underset{\sim}{z})$ and $v_s(\underset{\sim}{z}^*)$. The relative biases of $v_h(\underset{\sim}{z})$ and $v_h(\underset{\sim}{z}^*)$ are both relatively small with the bias of $v_h(\underset{\sim}{z}^*)$ being generally slightly less negative (or more positive) than $v_h(\underset{\sim}{z})$. The biases for the characteristics at the EA level are evenly distributed between positive and negative for $v_h(\underset{\sim}{z})$ and $v_h(\underset{\sim}{z}^*)$, while they are mostly negative for $v_s(\underset{\sim}{z})$ and $v_s(\underset{\sim}{z}^*)$. $v_h(\underset{\sim}{z})$ will be used in the numerical example of Section 6 since it is so similar to $v_h(\underset{\sim}{z}^*)$, and is much less downward biased than $v_s(\underset{\sim}{z})$ and $v_s(\underset{\sim}{z}^*)$. The estimated standard errors of the estimated relative bias of all four variance estimators ranged from 2.3 to 8.6 at the EA level, and from 3.1 to 11.0 at the WA level.

Note that since the estimates of variance at the WA level are just sums of EA level estimates, the pattern of the bias should be similar for both EA and WA level characteristics. The fact that downward bias was found for more WA level than EA level characteristics is due simply to the fact that different characteristics were studied at the two levels. This was done to maximize the diversity of the characteristics examined.

The negative biases are not unexpected because the variance estimates do not account for the variability introduced by somewhat different constraints being dropped in each sample. Also, Rao (1968) has shown that the estimated variance for a regression estimator can be badly downward biased when the sample is small.

Estimates of the variance were also calculated regarding the systematic sample as a stratified random sample with 4 households selected from each successive stratum of 20 households. This estimator was as badly downward biased as $v_s(\underset{\sim}{z})$.

The above results were achieved by dropping constraints less than 60 in size for smallness and by dropping constraints for near linear dependence if they caused the condition number to increase by 1,000 or more. The Monte Carlo study was repeated using other values for these two parameters. It was found, however, that the values 60 and 1,000 tended to minimize the bias of the variance estimators while retaining a reasonable number of constraints.

A repeat of the Monte Carlo study on a different WA (but only for 25 samples) indicated that the variance estimator for a given characteristic can be downward biased for one WA and upward biased for another. Also, the variance estimator for a given characteristic was often downward biased for one or more EAs and upward biased for one or more other EAs. Consequently, the bias should be smaller than that shown in Tables 1a and 1b for estimates at geographic levels above WA, since some of the bias should cancel out.

6. APPLYING THE TWO STEP GLS ESTIMATOR TO CENSUS DATA

To assess its performance, the two step GLS estimation method was applied to a sample of 79 WAs using 1986 Census data. A total of 62 WA level constraints were applied plus the two EA level constraints for each EA. No EA level constraints were defined for the smallest EA in each WA, however, since they would have been discarded for being linearly dependent with the other EA level constraints. Since there were 7.4 EAs on average for the sampled WAs, an average of 74.8 WA and EA level constraints were initially applied to each WA. In the discussion which follows, all counts of constraints will be taken to be averages.

First, 7.7 constraints were discarded for being small, 6.6 for being linearly dependent and 9.8 for being nearly linearly dependent. As a result, 50.7 of the original 74.8 constraints were retained, of which 40.4 were WA level constraints. The initial average condition number of $\hat{\underset{\sim}{X}}^{(0)'}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ after discarding small and linearly dependent constraints was 2,392,056. The average condition number of $\hat{\underset{\sim}{X}}^{(0)'}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ after discarding nearly linearly dependent constraints was 6,350.

Then, at the EA level, before the first step weighting adjustments were calculated, 22.3 of the 40.4 WA level constraints were discarded for being small. This left 18.1 WA level constraints to be partitioned into two groups of 9.0 each at the EA level. After discarding 0.1 linearly dependent constraints per group as well as 1.0 constraints per group for causing outlier weights, the number of WA level constraints in each of the two groups was 8.0. The average condition number of the $\hat{\underset{\sim}{X}}_{gr}^{(0)'}\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)}$ matrix was 379 after discarding the constraints which caused outlier weights.

At the second step weighting adjustment, 7.4 of the 50.7 WA and EA level constraints were dropped for causing outlier weights. This left 43.3 constraints that were used to determine the second step weighting adjustments. The average condition number of the $\hat{\underset{\sim}{X}}^{(A)'}\underset{\sim}{V}_{A}^{-1}\hat{\underset{\sim}{X}}^{(A)}$ matrix was 4,855 after discarding the constraints which caused outlier weights.

The differences between known population counts and the corresponding sample estimates for 68 selected characteristics appearing in Census publications were calculated at both the EA and WA levels, for all 79 WAs. The absolute values of the relative population/estimate differences are summarized in Tables 2a to 2d for two step GLS estimates, one step GLS estimates using the approach outlined in Section 2, raking ratio estimates based on the 1986 Census weights (see Brackstone and Rao 1979, for a description of the raking ratio weighting methodology), and Horvitz-Thompson estimates using the initial weights $W_g^{(0)}$. The 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of the differences are given. Differences for each characteristic were only included for EAs and WAs in which the characteristic applied to at least 60 households. All relative population/estimate differences are in percentage terms. The tables show that the two step GLS estimator in general produced much smaller population/estimate differences than the one step GLS estimator at the EA level, while producing differences of similar size at the WA level. Compared to the 1986 raking ratio and Horvitz-Thompson estimators, the two step GLS estimator in general produced much smaller differences at both the EA and WA levels. The constraints used with the two step GLS estimator more closely represent those characteristics which appear in Census publications than the constraints used with the 1986 raking ratio estimator. This contributed to the smaller differences at both the EA and WA levels, while the first step of the weight calculations also contributed to the smaller EA level differences. Note that, in general, the differences are actually larger at the EA level for the 1986 raking ratio estimator than for the Horvitz-Thompson estimator.

For the 79 WAs, estimated coefficients of variation (CVs) were calculated for estimates of 507 EA level and 642 WA level characteristics (all of which applied to at least an estimated 60 households in the population) known only on a sample basis. Selected percentiles of the distribution of the estimated CVs of two step GLS estimators are compared in Tables 3a to 3d to the percentiles of estimated CVs of corresponding 1986 raking ratio estimators and Horvitz-Thompson estimators. All CVs are in percentage terms. The tables show that the two step GLS estimator generally had smaller CVs than both the 1986 raking ratio and Horvitz-Thompson estimators, especially at the EA level.

The two step GLS weighting procedure and the associated methodology for discarding constraints also worked well for the 5,730 WAs in the 1991 Canadian Census. Compared to the 1986 Census, population/estimate differences in the 1991 Census were dramatically reduced at the EA level for most characteristics. At the Census Division level, population/estimate differences were reduced for two thirds of the characteristics examined.

7. CONCLUDING REMARKS

The two step GLS estimator worked well, with no manual intervention, on all 5,730 WAs in the 1991 Canadian Census. Adjusting the initial weights in two steps substantially reduced the population/estimate discrepancies and CVs for small areas compared to methods used in the 1986 Census. Discarding constraints to eliminate adjusted weights less than 1 and to lower the condition numbers of matrices being inverted also proved effective. The computational costs of determining these adjusted weights and estimating the variances of the resulting estimators were very reasonable.

REFERENCES

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

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.

Hidiroglou, M.A., Fuller, W.A. and Hickman, R.D. (1978), SUPERCARP, Statistical Laboratory, Iowa State University, Ames, Iowa.

Montgomery, D.C. and Peck, E.A. (1982), Introduction to Linear Regression Analysis, New York: John Wiley and Sons.

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

Rao, J.N.K. (1968), "Some Small Sample Results in Ratio and Regression Estimation", Journal of the Indian Statistical Association, 6, 160-168.

Särndal, C.E., Swensson, B. and Wretman, J.H. (1989), "The Weighted Residual Technique for Estimating the Variance of the General Regression Estimator of the Finite Population Total", Biometrika, 76, 527-537.

Wolter, Kirk M. (1985), Introduction to Variance Estimation, New York: Springer-Verlag.

Table 1a. Relative Bias of Variance Estimators - WA Level Characteristics

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
$V_h(\underset{\sim}{Z})$	-51	-43	-21	2	12
$V_h(\underset{\sim}{Z}^*)$	-50	-37	-18	10	21
$V_s(\underset{\sim}{Z})$	-68	-63	-53	-38	-26
$V_s(\underset{\sim}{Z}^*)$	-67	-59	-47	-35	-20

Table 1b. Relative Bias of Variance Estimators - EA Level Characteristics

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
$V_h(\underset{\sim}{Z})$	-64	-44	3	69	131
$V_h(\underset{\sim}{Z}^*)$	-62	-41	16	85	138
$V_s(\underset{\sim}{Z})$	-79	-61	-40	11	75
$V_s(\underset{\sim}{Z}^*)$	-76	-58	-35	10	94

Table 2a. Relative Population/Estimate Differences - WA Level Estimates > Median Value

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Two Step GLS	0.0	0.0	0.0	0.2	1.7
One Step GLS	0.0	0.0	0.0	0.2	1.7
Raking Ratio	0.0	0.0	0.1	0.8	2.6
Horvitz-Thom.	0.3	1.0	2.6	5.6	10.2

Table 2b. Relative Population/Estimate Differences - WA Level Estimates ≤ Median Value

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Two Step GLS	0.0	0.0	0.0	6.8	15.8
One Step GLS	0.0	0.0	0.0	6.8	15.7
Raking Ratio	0.0	0.5	3.5	9.2	16.1
Horvitz-Thom.	1.5	4.0	8.7	15.7	23.9

Table 2c. Relative Population/Estimate Differences - EA Level Estimates > Median Value

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Two Step GLS	0.0	0.8	2.2	4.7	11.5
One Step GLS	0.1	1.3	3.8	7.7	12.9
Raking Ratio	0.8	2.2	5.1	9.4	14.8
Horvitz-Thom.	0.4	1.7	4.5	8.9	14.7

Table 2d. Relative Population/Estimate Differences - EA Level Estimates ≤ Median Value

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Two Step GLS	1.2	3.4	7.4	14.2	22.8
One Step GLS	1.7	4.7	10.6	19.6	29.2
Raking Ratio	1.9	5.0	11.0	19.8	30.1
Horvitz-Thom.	1.5	4.6	10.9	19.8	31.0

Table 3a. Estimated CVs - WA Level Estimates > Median Value

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Two Step GLS	2.4	3.4	4.8	6.7	8.3
Raking Ratio	2.7	3.9	5.6	7.6	9.4
Horvitz-Thom.	3.8	5.1	7.2	9.2	11.0

Table 3b. Estimated CVs - WA Level Estimates ≤ Median Value

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Two Step GLS	9.1	10.7	13.3	17.8	20.3
Raking Ratio	10.6	12.3	15.2	19.7	23.2
Horvitz-Thom.	12.9	14.7	17.6	21.7	24.3

Table 3c. Estimated CVs - EA Level Estimates > Median Value

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Two Step GLS	5.6	7.6	9.7	12.0	14.3
Raking Ratio	7.6	10.0	12.6	14.6	16.7
Horvitz-Thom.	7.8	10.4	13.1	15.3	17.8

Table 3d. Estimated CVs - EA Level Estimates ≤ Median Value

Estimator	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
Two Step GLS	10.4	12.4	14.9	17.9	20.2
Raking Ratio	13.4	15.5	17.6	20.6	22.9
Horvitz-Thom.	13.8	16.2	18.4	21.3	23.2