Statistics Canada | Statistique Canada

Methodology Branch

Social Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes sociales

Canada

[redacted]

C 2

WORKING PAPER
METHODOLOGY BRANCH

Two Step
Generalized Least Squares Estimation
in the 1991 Canadian Census

Michael D. Bankier, Stephen Rathwell and Mark Majkowski

SSMD 92-007E

Census Operations Section
Social Survey Methods Division, Statistics Canada

August 1992

TABLE OF CONTENTS

ABSTRACT

This article describes the method used to produce estimates for the one in five sample of households selected in the 1991 Canadian Census. A primary objective of the estimation methodology is to generate household weights such that differences between known population counts and the corresponding estimates are reduced for small areas (called enumeration areas (EAs)) while at the same time eliminating these population/estimate differences for larger areas (called weighting areas (WAs)). The characteristics for which these differences are to be reduced or eliminated will be called constraints on the weights. To achieve this objective, two adjustments are made to the initial household weights which equal the inverse of the probability of selection. The first adjustment is calculated separately for each EA. The constraints are partitioned into two groups. Weighting adjustment factors are calculated for each group of constraints using generalized least squares (also called regression) estimation. The two resulting sets of weighting adjustment factors are then combined to form an average weighting adjustment factor. When applied to the initial weights, the resulting average adjusted weights reduce but do not eliminate population/estimate differences at the EA level. Generalized least squares (GLS) estimation is then applied a second time at the WA level using all the constraints. The weighting adjustment factors which result are applied to the average adjusted weights. The final adjusted weights eliminate the population/estimate differences at the WA level. Methods are described for discarding constraints which are linearly dependent, nearly linearly dependent or small. In addition, it is shown how to discard constraints to ensure that all the adjusted weights are within the desired range [1,25]. A computationally efficient method of estimating the variance of the two step GLS estimator is described. A Monte Carlo study assesses the bias of both the two step GLS estimator and various estimators of the variance of the two step GLS estimator. The estimation method is then applied to 79 WAs from the 1986 Census. A brief discussion is given of the successful application of this estimation method to the 5,730 WAs in the 1991 Census. Finally, alternative estimators such as the raking ratio estimator and the logit estimator are discussed.

KEY WORDS: Regression Estimation; Raking Ratio Estimation; Logit Estimation.

RÉSUMÉ

Cet article décrit la méthode employée pour produire des estimations pour l'échantillon de "un ménage sur cinq" sélectionné au recensement du Canada de 1991. L'objectif principal de la méthodologie d'estimation consiste à générer des poids de ménage de manière à réduire les écarts entre les chiffres connus de population et les estimations correspondantes pour certains petits domaines (appelés secteurs de dénombrement (SD)), tout en éliminant les écarts population/estimation pour des zones plus grandes, appelées zones de pondération (ZP). Les caractéristiques pour lesquelles ces écarts doivent être réduits ou éliminés seront appelées contraintes sur les poids. Afin de réaliser cet objectif, on effectue deux ajustements aux poids de ménage initiaux, lesquels sont égaux à l'inverse de la probabilité de sélection. Le premier ajustement est calculé pour chaque SD séparément. Les contraintes sont divisées en deux groupes. Pour chaque groupe de contraintes, les facteurs d'ajustement de poids sont calculés selon la méthode d'estimation des moindres carrés généralisée (aussi appelée estimation par la régression généralisée). Les deux ensembles de facteurs d'ajustement résultants sont alors combinés pour former un facteur d'ajustement de poids moyen. L'application des poids moyens ajustés résultants réduit, mais n'élimine pas les écarts population/estimation au niveau des SD. On applique alors la méthode d'estimation par les moindres carrés généralisée (MCG) une seconde fois au niveau des ZP en employant toutes les contraintes. Les facteurs d'ajustement de poids qui en résultent sont appliqués aux poids moyens ajustés. Les poids ajustés finaux éliminent les écarts population/estimation au niveau des ZP. On décrit les méthodes en vue d'exclure les contraintes qui sont linéairement dépendantes, quasi linéairement dépendantes ou peu importantes. De plus, on montre comment exclure les contraintes afin de s'assurer que tous les poids ajustés se situent dans l'étendue désirée [1,25]. Une méthode de calcul efficace pour estimer la variance de l'estimateur MCG à deux étapes est décrite. A l'aide d'une étude de Monte-Carlo, on évalue le biais de l'estimateur MCG à deux étapes ainsi que le biais de divers estimateurs de la variance de ce dernier. La méthode d'estimation est alors appliquée à 79 ZP provenant du recensement de 1986. Une discussion brève sur la performance de cette méthode appliquée à 5 730 ZP du recensement de 1991 est décrite. Enfin, d'autres estimateurs tels que l'estimateur itératif par le quotient et l'estimateur logit sont également discutés.

MOTS CLÉS: Estimation par régression; estimation itérative par le quotient; estimation logit.

Two Step Generalized Least Squares Estimation in the 1991 Canadian Census

Michael D. Bankier, Stephen Rathwell and Mark Majkowski*

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) out of a total of 45,995 EAs. Sampled EAs contained on average 249 dwellings. The other 5,923 EAs (which were remote EAs, Indian reserves or EAs containing exclusively collective dwellings) were sampled 100%. 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 whose households were sampled 20%. Characteristics for which consistency is required between the sample estimates and the population counts are called constraints on the weights. Although 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), certain problems have been documented with these weighting procedures. 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, while there was close or exact agreement between the sample estimates and population counts at the WA level, this was frequently not achieved at the EA level. In fact, usually the agreement with RR estimates was no better than for estimates calculated using the initial weights. Finally, because

---

*Michael D. Bankier and Stephen Rathwell are Senior Methodologists, and Mark Majkowski is Methodologist, Social Survey Methods Division, Statistics Canada, 15th Floor, Coats Building, Ottawa, Ontario K1A 0T6, Canada. The authors would like to thank Gordon Brackstone, Michael Hidiroglou, Carl Särndal and Jocelyn Tourigny for their helpful comments.

different weights were used to produce household and persons 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. Two recent papers on this subject are Zieschang (1990) and Deville and Särndal (1992). Earlier papers include Stephan (1942), Friedlander (1961), Cassel, Särndal and Wretman (1976), Huang and Fuller (1978), Isaki and Fuller (1982), Wright (1983), and Särndal and Hidiroglou (1989). Renewed interest in GLS estimation was generated among practitioners by Luery (1986) and Bethlehem and Keller (1987). Examples of papers applying the GLS estimator are Alexander (1987), Copeland, Peitzmeier and Hoy (1987) and Lemaitre and Dufour (1987).

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. Rather than GLS estimation, a generalized form of RR estimation (Darroch and Ratcliff 1972) or logit estimation (Deville and Sarndal 1992) could have been used. Both the two step weighting adjustment and the calculation of a single household weight can be done with all three estimators. Also, the Newton-Raphson iterative solution as proposed by Deville and Sarndal (1992) generally converges much more quickly than the iterative solution of Deming and Stephan (1940). GLS estimation was used, rather than raking or logit estimation, 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. Also, it is not certain that the variance of the logit and the generalized RR estimators can be estimated accurately in a computationally efficient fashion. More study of this is required. The raking and logit estimators are discussed further in Section 7.

Besides illustrating a major application of GLS estimation, this article describes an

effective method for eliminating GLS weights less than 1 by discarding constraints. This approach contrasts with that of Deville and Sarndal (1992) who attempt to restrict weights to a certain range by picking distance measures such that the optimum weighting adjustment factors fall within a certain range. If the set of solutions that satisfy the constraints does not include a solution where all the adjustment factors fall within the desired range, then some of the constraints must be dropped as described in this paper. A new method of identifying nearly collinear constraints so they can be discarded is described. This article also outlines a computationally efficient method of estimating the GLS estimator variance. 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. Finally, a brief discussion is given of the successful application of this estimation method to the 5,730 WAs in the 1991 Census.

2. A DESCRIPTION OF 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 $\underset{\sim}{\hat{X}}^{(0)}=diag(\underset{\sim}{W}^{(0)})\underset{\sim}{x}=[W_g^{(0)}x_{ghi}]$ 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

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

is minimized subject to the constraints

$$\underset{\sim}{\hat{X}}^{(0)\prime}\underset{\sim}{c}=\underset{\sim}{X} \tag{2}$$

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

$$\underset{\sim}{C}=\underset{\sim}{1}_n+\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}(\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)})^{-1}(\underset{\sim}{X}-\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{1}_n) \tag{3}$$

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. A DESCRIPTION OF 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. The methods for calculating these weighting adjustment factors are described in Subsections 3.2 and 3.3. In Subsection 3.4, an efficient method of estimating the two step GLS MSE is described.

The weighting adjustment factors can be close to 0 or negative. Also, the matrix to be inverted to determine the adjustment factors can be singular or have a high condition number. Methods of dropping constraints to deal with these problems are described in Section 4.

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

$$\underset{\sim}{c}_{gr} = \underset{\sim}{1}_g + \underset{\sim}{V}_{gr}^{-1} \hat{\underset{\sim}{X}}_{gr}^{(0)} \left(\hat{\underset{\sim}{X}}_{gr}^{(0)'} \underset{\sim}{V}_{gr}^{-1} \hat{\underset{\sim}{X}}_{gr}^{(0)} \right)^{-1} \left(\underset{\sim}{X}_{gr} - \hat{\underset{\sim}{X}}_{gr}^{(0)'} \underset{\sim}{1}_g \right) = [c_{ghr}] \qquad (4)$$

$\underset{\sim}{1}_g$ is a $n_g \times 1$ vector of 1's, $\underset{\sim}{V}_{gr} = diag(\hat{\underset{\sim}{X}}_{gr}^{(0)} \underset{\sim}{1}_{I_{gr}})$ is a $n_g \times n_g$ 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 $n_g \times I_{gr}$ 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

$$\underset{\sim}{c} = \underset{\sim}{1}_n + \underset{\sim}{V}_A^{-1} \underset{\sim}{\hat{X}}^{(A)} (\underset{\sim}{\hat{X}}^{(A)\prime} \underset{\sim}{V}_A^{-1} \underset{\sim}{\hat{X}}^{(A)})^{-1} (\underset{\sim}{X} - \underset{\sim}{\hat{X}}^{(A)\prime} \underset{\sim}{1}_n) \qquad (5)$$

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

It can be shown, using two successive Taylor series approximations, that $MSE(\hat{Y}) \approx V(\hat{Y}) \approx V(\hat{Z}^{(0)})$ where $\hat{Y} = \sum_{g=1}^{G} \sum_{h=1}^{n_g} W_{gh} y_{gh}$, $\hat{Z}^{(0)} = \sum_{g=1}^{G} \sum_{h=1}^{n_g} W_g^{(0)} z_{gh}$ 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, $\underset{\sim}{\beta} = E(\underset{\sim}{\hat{\beta}})$, $\underset{\sim}{\hat{\beta}} = (\underset{\sim}{\hat{X}}^{(A)\prime} \underset{\sim}{V}_A^{-1} \underset{\sim}{\hat{X}}^{(A)})^{-1} \underset{\sim}{\hat{X}}^{(A)\prime} \underset{\sim}{V}_A^{-1} \underset{\sim}{\hat{Y}}^{(A)}$ and $\underset{\sim}{\hat{Y}}^{(A)} = [W_{gh}^{(A)} y_{gh}]$ is a n x 1 vector. Finally, $\underset{\sim}{\beta}_{gr}^* = E(\underset{\sim}{\hat{\beta}}_{gr}^*)$, $\underset{\sim}{\hat{\beta}}_{gr}^* = (\underset{\sim}{\hat{X}}_{gr}^{(0)\prime} \underset{\sim}{V}_{gr}^{-1} \underset{\sim}{\hat{X}}_{gr}^{(0)})^{-1} \underset{\sim}{\hat{X}}_{gr}^{(0)\prime} \underset{\sim}{V}_{gr}^{-1} \underset{\sim}{\hat{U}}_g^{(0)}$ and $\underset{\sim}{\hat{U}}_g^{(0)} = [W_g^{(0)} u_{gh}]$ is a n_g x 1 vector. An estimator of $MSE(\hat{Y})$ can be determined by replacing the $\underset{\sim}{\beta}$ with $\underset{\sim}{\hat{\beta}}$ when calculating $\underset{\sim}{u}$ and replacing the $\underset{\sim}{\beta}_{gr}^*$ with $\underset{\sim}{\hat{\beta}}_{gr}^*$ 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.

3.4 An Efficient Method of Estimating the Two Step GLS MSE

In the Census, estimates of the MSE are calculated for many characteristics. It is worthwhile, therefore, to calculate $\underset{\sim}{z}$ efficiently. This can be done by noting that $\underset{\sim}{u}=\underset{\sim}{R}_2\underset{\sim}{y}$, $\underset{\sim}{z}=\underset{\sim}{R}_1\underset{\sim}{u}$ and hence $\underset{\sim}{z}=\underset{\sim}{R}\underset{\sim}{y}$ where $\underset{\sim}{R}_2=\underset{\sim}{I}_n-\underset{\sim}{x}(\underset{\sim}{\hat{X}}^{(A)\prime}\underset{\sim}{V}_A^{-1}\underset{\sim}{\hat{X}}^{(A)})^{-1}\underset{\sim}{\hat{X}}^{(A)\prime}\underset{\sim}{V}_A^{-1}diag(\underset{\sim}{W}^{(A)})$ is a n x n matrix and $\underset{\sim}{I}_n$ is a n x n identity matrix. Also, $\underset{\sim}{R}_1$ is a n x n block diagonal matrix with $\underset{\sim}{R}_{1g}$, g = 1 to G running down the diagonal, $\underset{\sim}{R}_{1g}=\underset{\sim}{I}_g-\frac{1}{2}(\sum_{r=1}^{2}\underset{\sim}{x}_{gr}(\underset{\sim}{\hat{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1}\underset{\sim}{\hat{X}}_{gr}^{(0)})^{-1}\underset{\sim}{\hat{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1})\,diag(W_g^{(0)}\underset{\sim}{1}_g)$, $\underset{\sim}{I}_g$ is a n_g x n_g identity matrix and $\underset{\sim}{R}=\underset{\sim}{R}_1\underset{\sim}{R}_2$. $\underset{\sim}{R}$ is not a function of $\underset{\sim}{y}$ and hence only has to be calculated once. The matrix $\underset{\sim}{R}$ can thus be used repeatedly in the calculation of $\underset{\sim}{z}$ for many characteristics.

Frequently, estimated MSEs are required for a table of K cells where a sampled household or a person in a sampled household can fall in only one of the K cells (excluding totals). Assume that it is a table of household characteristics. Let $\underset{\sim}{y}_K = [y_{ghk}]$ be a n x K matrix where y_{ghk} equals the value for the characteristic of interest if the hth sampled household from the gth EA falls in the kth cell and $y_{ghk}=0$ otherwise. Because a sampled household can fall in only one cell, this means that a row of the $\underset{\sim}{y}_K$ matrix can have only one non-zero entry. The matrix

$\underset{\sim}{z}_K=\underset{\sim}{R}\underset{\sim}{y}_K=\sum_{g=1}^{G}\sum_{h=1}^{n_g}\underset{\sim}{y}_{gh}\otimes\underset{\sim}{R}_{gh}$ contains the z values required to estimate the MSEs for the K cells of the table where $\underset{\sim}{y}_{gh}$ represents the ghth row of $\underset{\sim}{y}_K$, $\underset{\sim}{R}_{gh}$ represents the ghth column of the $\underset{\sim}{R}$ matrix and $\otimes$ represents the Kronecker product. Because there is only one non-zero entry for $\underset{\sim}{y}_{gh}$, computer algorithms can be easily written such that only n^2 multiplications and additions are required to determine $\underset{\sim}{z}_K$. This compares to Kn^2 multiplications and additions that would be required to determine $\underset{\sim}{z}_K$ if the sparse nature of $\underset{\sim}{y}_K$ was not accounted for. For a table of person characteristics, a maximum of mn multiplications and additions would be required to calculate $\underset{\sim}{z}_K$ where m equals the number of persons in sampled households for that WA.

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 $\underset{\sim}{\hat{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1}\underset{\sim}{\hat{X}}_{gr}^{(0)}$ and $\underset{\sim}{\hat{X}}^{(A)\prime}\underset{\sim}{V}_A^{-1}\underset{\sim}{\hat{X}}^{(A)}$ 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 $\underset{\sim}{\hat{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\underset{\sim}{\hat{X}}^{(0)}$ 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 $\underset{\sim}{\hat{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\underset{\sim}{\hat{X}}^{(0)}$ is calculated (with the small constraints dropped). Then this matrix is assessed for linearly dependent constraints. (It can be shown that if a set of

columns for the matrix $\underset{\sim}{X}^{(0)'}\underset{\sim}{V}^{-1}\underset{\sim}{X}^{(0)}$ are linearly dependent that the corresponding columns of the matrix $\underset{\sim}{X}^{(0)}$ are linearly dependent.) In some cases, sets of constraints which are always linearly dependent can be identified before processing begins. In other cases, the SAS/IML HERMITE function can be used to identify sets of constraints which are linearly dependent. It does this by using elementary row operations to reduce the matrices being inverted to Hermite normal form (see Graybill 1969, p. 120). The smallest constraint in a set is dropped.

Next, constraints are discarded in order to lower the condition number of the matrix $\underset{\sim}{X}^{(0)'}\underset{\sim}{V}^{-1}\underset{\sim}{X}^{(0)}$. To do this, the matrix $\underset{\sim}{X}^{(0)'}\underset{\sim}{V}^{-1}\underset{\sim}{X}^{(0)}$ 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 $\underset{\sim}{X}^{(0)'}\underset{\sim}{V}^{-1}\underset{\sim}{X}^{(0)}$ 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 $\underset{\sim}{X}^{(0)'}\underset{\sim}{V}^{-1}\underset{\sim}{X}^{(0)}$ 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 $\underset{\sim}{X}^{(0)'}\underset{\sim}{V}^{-1}\underset{\sim}{X}^{(0)}$ 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 $\underset{\sim}{X}^{(0)'}\underset{\sim}{V}^{-1}\underset{\sim}{X}^{(0)}$ 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 $\hat{\underset{\sim}{X}}_{gr}^{(0)'}\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)}$ 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 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. In the interests of computational efficiency, however, the bisection method (see Pizer 1975, p.187) is used to identify the constraints which should be dropped.

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 $\hat{\underset{\sim}{X}}^{(0)'}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$. 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 $\hat{\underset{\sim}{X}}^{(A)'}\underset{\sim}{V}_A^{-1}\hat{\underset{\sim}{X}}^{(A)}$ 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 with seven EAs was created for this study from five similar 1986 Census WAs.

it contained only sampled households. The households in each EA were partitioned into five systematic samples. The $5^7=78,125$ possible combinations of EA level systematic samples were then formed into 15,625 clusters of five disjoint systematic samples each. Thus, each household in the WA fell in one and only one of the five systematic samples for each cluster. A random sample of 50 clusters (250 samples) for use in the Monte Carlo study was selected without replacement from the 15,625 clusters. Clusters of samples were selected rather than a simple random sample of samples because it was determined that the intracluster correlation coefficient of the estimates of variance was negative for the majority of the characteristics being examined. Thus more precise estimators of the bias of the variance estimators resulted.

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 results are summarized in Figure 1. The plot provides the empirical cumulative distribution function (ECDF) for the estimated relative bias over all characteristics. The absolute value of the

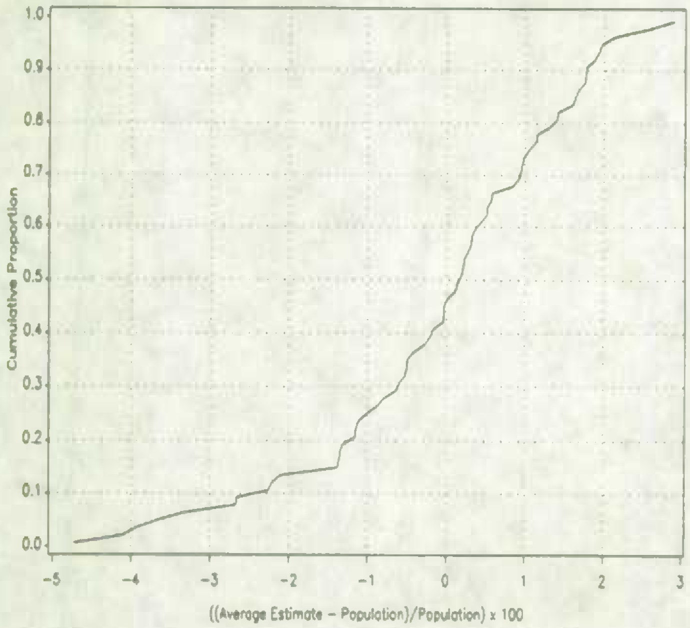

Figure 1. ECDF of Relative Bias of GLS Estimates

estimated relative bias for the characteristics ranges from almost zero to as high as 4.7%, although it is less than 2% for most characteristics. It is 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.

Figure 2 provides ECDFs 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

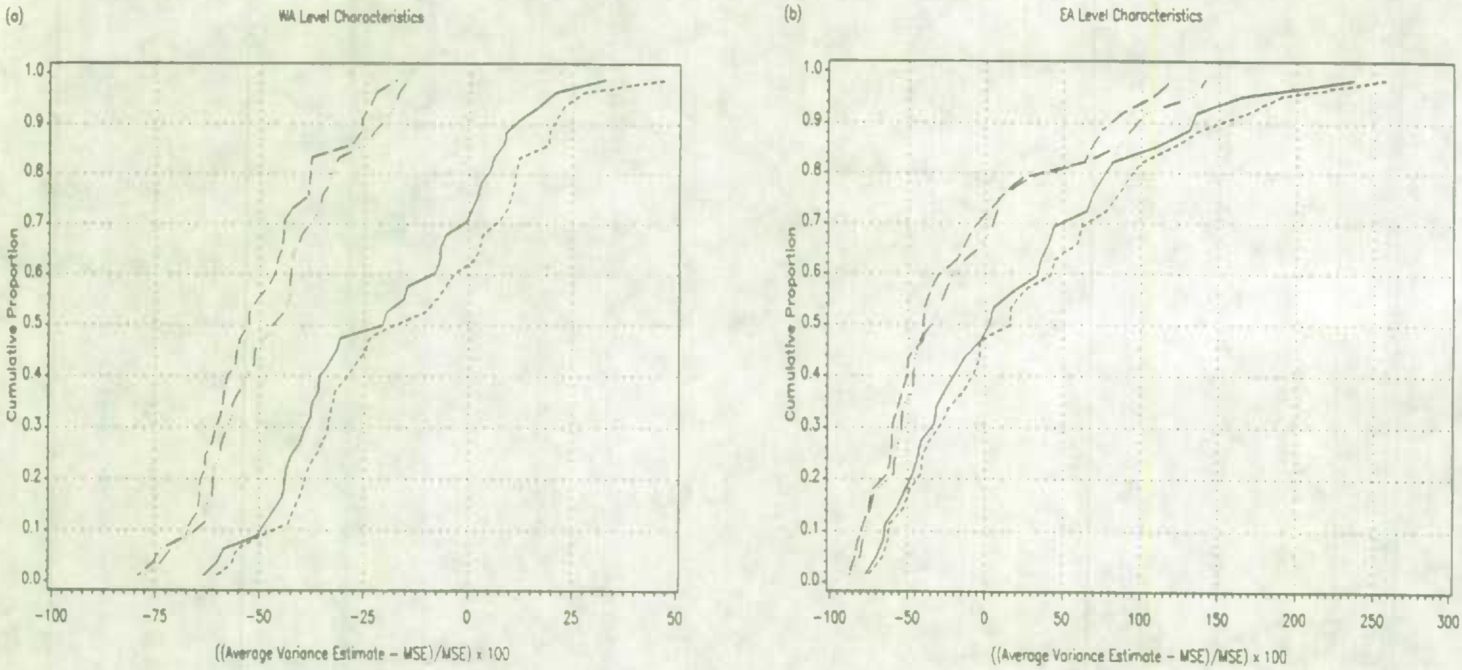

Figure 2. ECDFs of the Estimated Relative Bias of the Variance Estimators.
(——) = $v_h(\underset{\sim}{z})$, (----) = $v_h(\underset{\sim}{z}^*)$, (— — —) = $v_s(\underset{\sim}{z})$, (— ··· —) = $v_s(\underset{\sim}{z}^*)$.

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 $v_g(\underset{\sim}{z})$.

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 Figure 2 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. The sampled WAs were selected from a rural Alberta Census Division (CD) and two urban CDs (Toronto and Montreal). All 24 WAs from the rural CD were selected. The WAs in the urban CDs were partitioned into 27 strata based on household and person characteristics, and 55 WAs were sampled. 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. The constraints and estimation algorithm applied to these 79 WAs were identical to those used to calculate the adjusted weights for the 5,730 WAs in the 1991 Census.

First, 7.7 constraints were discarded for being small. Then, when the $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ matrix was assessed, 6.6 constraints were discarded for being linearly dependent and 9.8 constraints were discarded 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 $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ after discarding small and linearly dependent constraints was 2,392,056. The average condition number of $\hat{\underset{\sim}{X}}^{(0)\prime}\underset{\sim}{V}^{-1}\hat{\underset{\sim}{X}}^{(0)}$ 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 $\hat{\underset{\sim}{X}}_{gr}^{(0)\prime}\underset{\sim}{V}_{gr}^{-1}\hat{\underset{\sim}{X}}_{gr}^{(0)}$ 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 $\hat{\underset{\sim}{X}}^{(A)\prime}\underset{\sim}{V}_{A}^{-1}\hat{\underset{\sim}{X}}^{(A)}$ 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 as ECDFs in Figure 3 for two step GLS estimates, one step GLS estimates using the approach outlined in Section 2, raking ratio

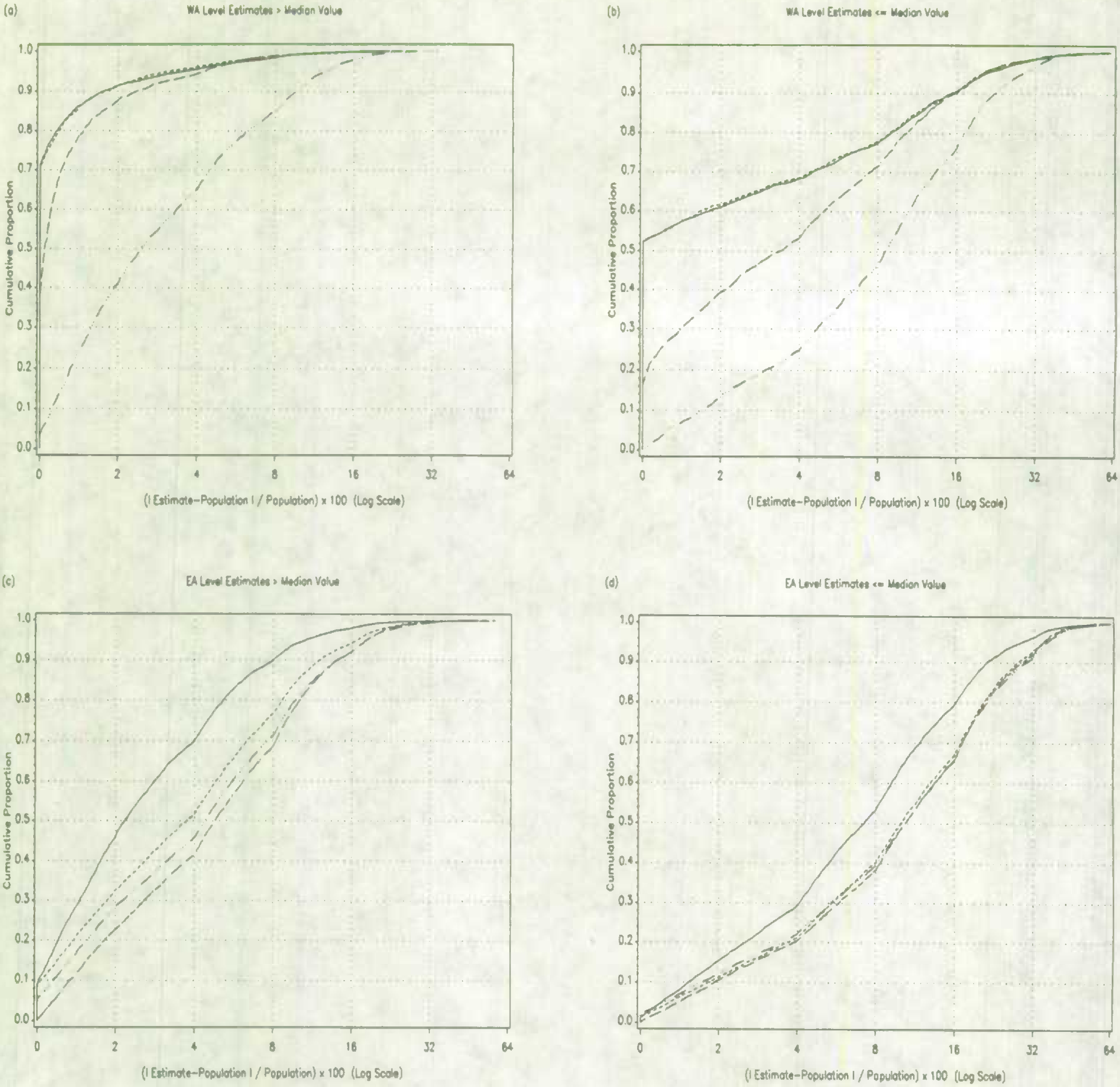

Figure 3. ECDFs of the Absolute Values of the Relative Differences Between the Sample Estimates and the Population Counts. (——) = Two Step GLS, (----) = One Step GLS, (— — —) = 1986 Raking Ratio, (— ··· —) = Horvitz-Thompson. The relative differences for large characteristics (> median value) are plotted separately from the relative differences for small characteristics (< = median value).

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 $W_g^{(0)}$. 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. Figure 3 shows 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. It is worth noting 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. Estimated CVs of two step GLS estimators are compared in Figure 4 to estimated CVs of corresponding 1986 raking ratio estimators and Horvitz-Thompson estimators. All CVs are in percentage terms. Figure 4 shows 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.

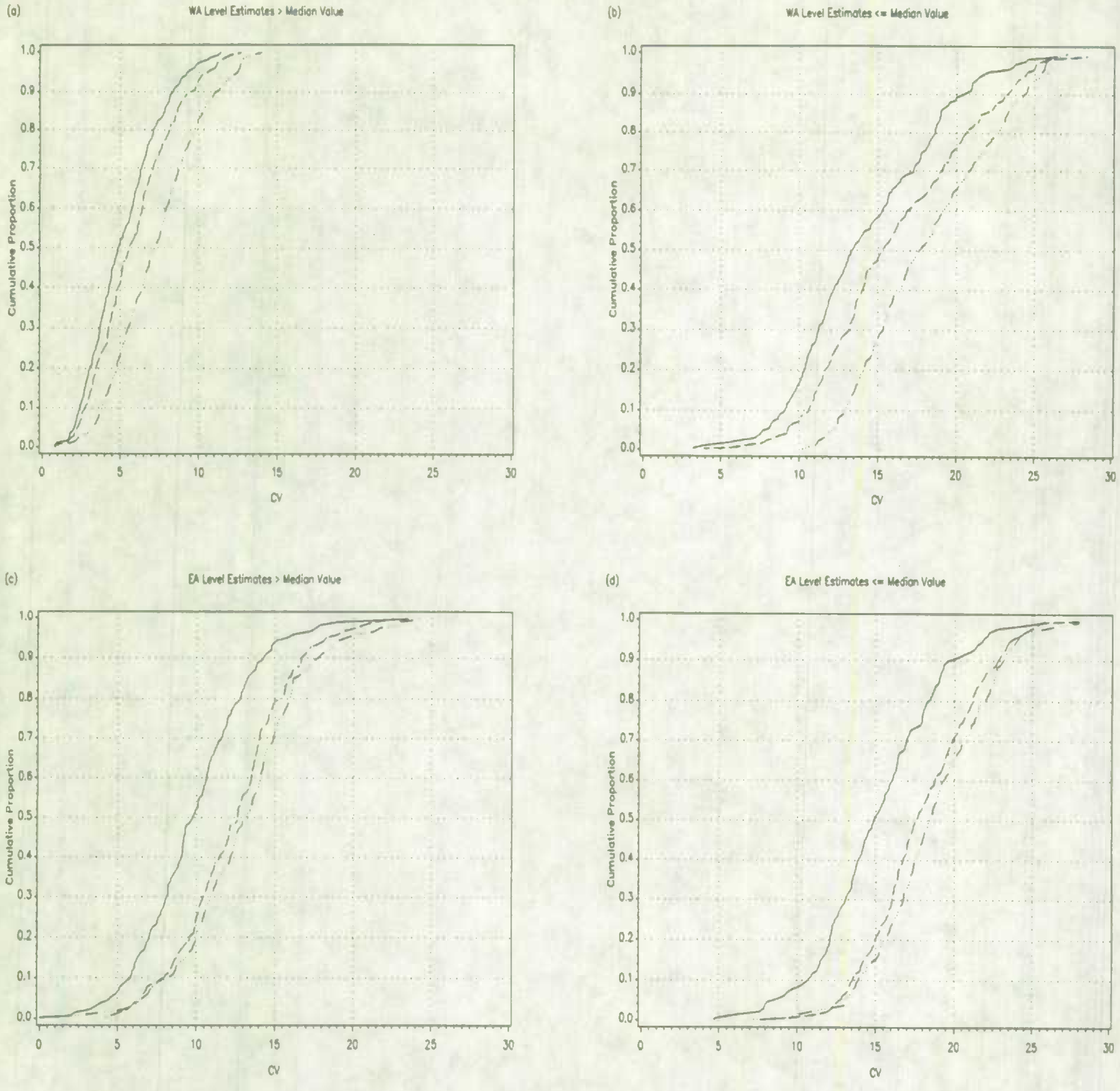

Figure 4. ECDFs of the Estimated CVs of the Estimators. (——) = Two Step GLS, (— — —) = 1986 Raking Ratio, (— ··· —) = Horvitz-Thompson. The CVs for large characteristics (> median value) are plotted separately from the CVs for small characteristics (< = median value).

7. ALTERNATIVE ESTIMATION METHODS

Rather than dropping constraints to eliminate outlier weights, Deville and Särndal (1992) suggest using alternative distance measures of the form $D_T = \sum_{g=1}^{G} \sum_{h=1}^{n_g} \hat{X}^{(0)}_{gh.} D_T(c_{gh})$ (with T equalling R (raking ratio) or L (logit)) where $D_R(c_{gh}) = c_{gh} \log(c_{gh}) - c_{gh} + 1$,

$$D_L(c_{gh}) = \frac{1}{A}\left[(c_{gh}-L) \ln\left(\frac{c_{gh}-L}{1-L}\right) + (U-c_{gh}) \ln\left(\frac{U-c_{gh}}{U-1}\right)\right]$$

, A = (U-L)/(1-L)(U-1) and L < 1 < U. It can be demonstrated that c_{gh} will fall in the range (L,U) for the logit estimator while c_{gh} is always positive for the raking ratio estimator. With both distance measures, c_{gh} can be determined by solving a system of non-linear equations using the Newton-Raphson (N-R) iterative method (see Pizer, 1975, p. 230).

A two step logit estimator was calculated for a sample of 12 of the 79 WAs studied in Section 6. In the Census, it is desired to bound the adjusted weights in the range [1,25]. Thus a lower bound (MINWT) and an upper bound (MAXWT) on the adjusted weights were defined. Then at the first step, L = MINWT/ $W_g^{(0)}$ while U = MAXWT/ $W_g^{(0)}$ were calculated separately for each EA where MINWT = 1.5 and MAXWT = 25. At the second step, L = MINWT/(Minimum value of $W_{gh}^{(A)}$ in the WA) while U = MAXWT/(Maximum value of $W_{gh}^{(A)}$ in the WA) where MINWT = 1 and MAXWT = 37.5. MINWT was set equal to 1.5 in the first step so that L in the second step would be smaller than 1/1.5 = 0.66. Similarly, the values of MAXWT were chosen such that U > 37.5/25 = 1.5 in the second step. Having values of L and U closer to 1 could have resulted in the logit iterative process not converging, or the distribution of the weighting adjustment factors at the second step could have been clustered around L and U. For the 12 WAs, the largest two step logit weight was less than 25 even though MAXWT = 37.5 at the second step.

The two step logit estimation procedure used identical criteria to those used with the

two step GLS estimator for discarding constraints for smallness, linear dependence and near linear dependence. The two step logit estimator procedure discarded additional constraints if the matrix being inverted in the N-R iterative process was singular or had a condition number exceeding 10,000. At the EA level, on average 0.1 constraints were discarded from each of the two groups of constraints for this reason. At the WA level, 3.4 constraints were discarded on average for this reason. With the two step GLS estimator for the same 12 WAs, the corresponding average number of constraints discarded for causing weights outside the range [1,25] were 1.2 and 9.0 respectively. The absolute value of the relative population/estimate differences for the two step logit estimator were compared to those for the two step GLS estimators. For WA level characteristics equal to or less than the median value, it was found that approximately 50% of the two step logit estimators had zero relative population/estimate differences compared to 38% of the two step GLS estimators. In contrast, however, the ECDF for EA level characteristics above the median value for the two step logit estimator was somewhat to the right of the ECDF for the two step GLS estimator. Thus the population/estimate differences were not consistently reduced.

Deville and Särndal (1992) recommend using the estimated variance of the one step GLS estimator as an estimate of the variance of the one step logit estimator since the asymptotic variances of the two estimators are equal. It is reasonable, therefore, to consider using the estimated variance of the two step GLS estimator as an estimate of the variance of the two step logit estimator. In both cases, the same constraints retained with the logit estimator would be used with the GLS estimator. This could, however, result in negative weights being generated at both the first and second steps for the GLS estimator.

The $\underset{\sim}{V}_A$ matrix at the second step would have to be defined as $\underset{\sim}{V}_A = diag(\underset{\sim}{x}\underset{\sim}{1})$. Otherwise, it would not be positive definite if some of the weights $W_{gh}^{(A)}$ were negative. A Monte Carlo study would be required to assess the accuracy of using the estimated variance of a two step GLS estimator with some negative weights as the estimated variance of a two step logit estimator with all weights greater than 1.

8. 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.

There are a number of areas of possible research for the 1996 Canadian Census. The logit estimator should be examined further since it may allow more constraints to be retained. A Monte Carlo study could be done to determine if the estimated variance of the two step GLS estimator provides a good estimate of the variance of the two step logit estimator. Alternative methods of partitioning the constraints in the first step and discarding constraints for near linear dependence or causing weights less than 1 could be considered. Ways of reducing the bias of the estimators of the variance could also be studied.

REFERENCES

Alexander, Charles H. (1987), "A Class of Methods for Using Person Controls in Household Weighting", Survey Methodology, 13, 183-198.

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

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

Cassel, Claes M., Särndal, Carl E. and Wretman, Jan (1976), "Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations", Biometrika, 63, 615-620.

Copeland, K.R., Peitzmeier, F.K. and Hoy, C.E. (1987), "An Alternative Method of Controlling Current Population Survey Estimates to Population Counts", Survey Methodology, 13, 173-181.

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.

Deville, Jean-Claude, and Särndal, Carl-Erik (1992), "Calibration Estimators in Survey Sampling", Journal of the American Statistical Association, 87, 376-382.

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.

Graybill, F.A. (1969), Introduction to Matrices with Applications in Statistics, Belmont, CA: Wadsworth, Inc.

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

Huang, Elizabeth T., and Fuller, Wayne A. (1978), "Nonnegative Regression Estimation For Sample Survey Data", Proceedings of the Social Statistics Section, American Statistical Association, 300-305.

Isaki, Cary T., and Fuller, Wayne A. (1982), "Survey Design Under the Regression Superpopulation Model", Journal of the American Statistical Association, 77, 89-96.

Lemaitre, G. and Dufour, J. (1987), "An Integrated Method for Weighting Persons and Families", Survey Methodology, 13, 199-207.

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

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, Carl-Erik, and Hidiroglou, Michael A. (1989), "Small Domain Estimation: A Conditional Analysis", Journal of the American Statistical Association, 84, 266-275.

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.

Stephan, F.F. (1942), "An Iterative Method of Adjusting Sample Frequency Tables When Expected Marginal Totals are Known", Annals of Mathematical Statistics, 13, 166-178.

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

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

Zieschang, Kimberly D. (1990), "Sampling Weighting Methods and Estimation of Totals in the Consumer Expenditure Survey", Journal of the American Statistical Association, 85, 986-1001.