ALTERNATIVE FAMILY WEIGHTING PROCEDURES FOR THE CURRENT POPULATION SURVEY

Flora Kwok Peitzmeier, Arthur L. Hughes, and C. Easley Hoy
Bureau Of Labor Statistics
Flora Peitzmeier, 441 G. St., NW, Wash. D.C. 20212

KEY WORDS: Principal person weighting, generalized least squares, multivariate raking ratio estimation

1. INTRODUCTION

The Current Population Survey (CPS) produces monthly labor force and demographic estimates for the U.S. working-age civilian noninstitutional population. Monthly, quarterly, and/or annual estimates are also produced for subaggregates of the population (i.e., 50 States and D.C., families, veterans, wage and salary earners, and persons not in the labor force). The CPS is a stratified multi-stage probability cluster sample of approximately 56,000 housing units. Each month a rotating sample of eight panels (called rotation groups) of housing units is interviewed.

The survey estimates for the total population (person level estimates) are generated through the application of weights derived from the probability of selection, adjustment for nonresponse, adjustment to reduce the variance due to the sampling of primary sampling units (first-stage ratio adjustment), and a post-stratification estimation procedure to reduce coverage errors and sampling variability (second-stage ratio adjustment - using univariate raking ratio estimation). A composite estimation procedure, and a seasonal adjustment procedure are used to generate the monthly person estimates (Hanson 1978). Estimates for families and other subaggregates are based on sample weights derived from adjustment procedures built on top of the second-stage person weights.

Estimates produced from person weights in the CPS are adjusted to agree with independent person controls. Because of differential coverage and nonresponse rates among various demographic groups in the CPS, the second-stage procedure can and does assign different weights to persons within a sample household. So if household or family characteristics are tabulated from CPS person weights, some inconsistencies will surface.

This paper investigates the application of alternative weighting procedures -- the Generalized Least Squares (GLS) and Multivariate Raking Ratio Estimation (MRRE) in the CPS for computing weights at the family level. The goal is to produce family weights that agree with independent demographic person controls and are as "close" as possible to the initial weights assigned to the persons in the families. Such procedures should: 1. Reduce the large number (as many as 5) and possibly greatly different weights assigned to each sample record under the current CPS weighting system. 2. Provide a single weight for a family unit. 3. Provide consistency between family estimates and estimates for the full population.

2. THE MARCH FAMILY WEIGHTING

In March of every year, the CPS Income Supplement includes questions on individual income, work experience, migration, tenure, and expanded questions on relationship and martial status. The March Supplement produces detailed data on household and family characteristics which are used to study family: income, employment status, marital status, and relationship. A family is defined as all individuals within a household related to one another by blood, marriage, or adoption.

The March principal person weighting was originally used only in the March supplement but is now also applied to CPS data each month to procuce controls for the Survey of Income Program Participation (SIPP). The objective of the March family weighting is to equalize the sum of the husbands' weights and the sum of the wifes' weights so that either person's weight can be used to represent the entire family in tabulating family data. The scheme is a series of adjustment steps applied to the male's final person weights.

1. Designate the wife in each married-spouse-present (MSP) household as the principal person and use her final person (second stage) weight as the husband's final person weight to tabulate estimates for the entire household (This is done because it is believed that in general females have better coverage than males). It is possible for MSP males to receive large changes (usually reductions) in final person weights due to large differences in the final person weights of the husbands and wives.

2. The current March family weighting assumes that the coverage rate for other male heads (OMH) was similar to that of the MSP males (although there is no evidence demonstrating that this was indeed true). The March weighting ratio-adjusts OMH weights' in the same direction as the average adjustment of weights of males in MSP households. This is done in the following way: In MSP households, compute a factor RO for each age(the same 16 age groups used in the second-stage estimation)-race(White/NonWhite) category,

$$RO = \frac{\Sigma \text{ (weights of female MSP)}}{\Sigma \text{ (weights of male MSP)}}$$

The family weight for OMH is the product of the appropriate factor RO and his final person weight. Lower OMH estimates usually resulted in the March Supplement when compared to basic CPS.

3. A final adjustment is performed so that the March Supplement estimates of the Hispanic population (in MSP, OMH, and all other males (AOM) households) agree with a set of independent population controls for males. The adjustment is carried out in two steps, and generally results in higher AOM estimates in the March Supplement than in the basic CPS.

a. For each male age-race (as defined in step 2 above) category, compute a factor RA.

$$RA = \frac{\text{Control} - \Sigma a - \Sigma b}{\text{Control} - \Sigma c - \Sigma d}$$

where a, b, c, and d are the adjusted male MSP, adjusted OMH, unadjusted male MSP, and unadjusted OMH weights respectively.

b. For each male age (the same 4 Hispanic age groups used in the second-stage estimation procedure) -race (White/NonWhite)-ethnicity(Hispanic/NonHispanic) category, compute a factor RE.

$$RE = \frac{\Sigma e - \Sigma f - \Sigma g}{\Sigma \text{ adjusted AOM}}$$

where e, f, and g are the unadjusted males, adjusted male MSP, and adjusted OMH weights respectively.

The family weight for AOM is the product of the appropriate factors RA, RE, and his final person weight. The changes of all non-MSP males weights are a function of the adjustments in the MSP males.

The March family weighting scheme is performed on the entire March Supplement sample rather than on each rotation group separately.

3. THE MONTHLY FAMILY WEIGHTING

The CPS Monthly family weighting procedure is a weighting rule, rather than an estimation procedure. A principal person is chosen from each family and his or her second-stage weight is used to tabulate family estimates. The principal person is defined to be the reference person (the first person listed as owner or renter of the sample housing unit), unless the reference person is a married male whose wife is also living in the housing unit, in which case the principal person is the wife. No other adjustments were made in this procedure. The CPS Monthly family weighting procedure is currently used only for generating quarterly family estimates of labor force and earnings characteristics at the Bureau of Labor Statistics.

4. THE GENERALIZED LEAST SQUARES

The generalized least squares (GLS) procedure adjusts the sample weights from prior stages of weighting by minimizing the weighted squared adjustments subject to a set of linear 'control' constraints the adjusted weights must satisfy. These linear constraints require that certain weighted estimates be equal to known control totals. GLS was first proposed for use in CPS weighting research by Luery (1980).

4.1 Properties of the GLS Procedure

1. When all the cells of the contingency table are nonempty, the GLS procedure produces best asymptotically normal (BAN) estimates (Neyman, 1949).
2. 'Convergence' is guaranteed under the GLS procedure. The GLS procedure controls each dimension of the weighting so that the sum equals the respective control totals.
3. The procedure is designed to minimize the statistic

$$M_A = \sum_{i=1}^{n} (W_i - \Omega_i)^2/\Omega_i \quad \text{(Luery, 1980)}$$

where Ω_i = the weight for the i^{th} record from previous stages of estimation (GLS uses the assumption that the set of Ω's produces unbiased estimates) (Luery 1986)

W_i = the final GLS weight for the i^{th} record

4.2. Application of GLS in the CPS

Each dimension that defines a set of controls in the post-stratification estimation in the CPS will define a set of linear constraints for the GLS procedure. The GLS procedure investigated in this paper will simultaneously adjust weights subsequent to the "first-stage" adjustment in the CPS to satisfy two criteria : 1) controlling the sample estimates to meet 3 sets of Census population controls: State, age/sex/ethnicity, and age/sex/race controls and 2) restraining each person's weight within a family to be equal. An exact solution is guaranteed under GLS with constraints on the sample marginal totals. This GLS procedure was performed independently for each rotation group. The goal is to minimize

$$f(W) = (W - \Omega)' \Omega_O^{-1} (W - \Omega)$$

subject to the constraint $X'W = N$ (Luery 1986) where

W = (nx1) vector of final weights derived for each of the n sample families.

Ω = (nx1) vector of family first-stage weights for each of the n sample families. This is defined as the first-stage weight from the sample record belonging to the 'lowest' race/sex/age classification within the family (If exist in the family, the 'lowest' race/sex/age classification is a White male between 16-17 years old. The next 'lowest' race/sex/age classification is a White male between 18-19 years old, etc..) Approximately 1 percent of the sample families have different first-stage weights within the family (This occurs when members within a family were of different races).

Ω_O = (nxn) diagonal matrix with the W_i's on the diagonal.

X = (nxk) design matrix (in CPS, k is 132) whose rows correspond to sample families, and whose columns correspond to control cells.

N = (kx1) vector of independent person population controls. The k is the number of controls specified where each control corresponds to a column of X. The columns of X were required to be linearly independent so that an inverse of the matrix $(X' W_O X)$ is achievable. Therefore, in setting up matrix X and vector N, the 137 control cells used in the CPS post-stratification estimation were reduced to a set of K=132 linearly independent cells.

The size of the matrices involved for CPS were quite large, with the number of rows for W, W_O , and X being around 7,700 for each rotation group. The unique solution to $X'W = N$ that minimizes (W_i is "closest" to W_i in the least squares sense) the specified f(W) was found to be (Luery, 1986):

$$W = W + \Omega_O X (X'\Omega_O X)^{-1} (N - X'\Omega)$$

This GLS procedure achieves agreement between basic person weight and family weight. However, it is possible for some W_i's to be negative under this f(W). In this application of the GLS, some elements of W were found to be negative.

4.3 Adjustment for Negative Weights

Methodologies for providing nonnegative weights under linear constraints on the weights are discussed, among others, in Huang and Fuller (1978), Zieschang (1986), and Luery (1986) (Luery's methodology was to reduce the possibility of achieving negative weights.). Due to complexity and time constraints, the methodology proposed by Huang and Fuller was not explored and implemented in this paper.

One of Zieschang's proposed methodologies was to recode the adjusted weights when they fell outside a tolerance interval containing the unadjusted weight. This method for providing nonnegative weights was not implemented in this investigation because it would provide an upward bias with unknown magnitude in estimating the number of families in the CPS and it would produce sample estimates that do not meet all of the independent population controls.

Luery's methodology was implemented in this investigation because it provided an alternative distance measure that may reduce the occurrence of negative weights and at the same time satisfy all of the constraints.

In this application of the GLS procedure to the CPS family weighting, a total of 38 (9 families) of the 114,117 sample person (53,923 families) records were found to have negative weights in four rotation groups (rotation groups 5-8) in four States (Texas, South Carolina, New Mexico, and California).

There did not appear to be a pattern to the occurrence of negative weights by demographic characteristics. However, the magnitude of the adjustments applied to the family first-stage weights are related to the coverage rates and family size of the family being adjusted for. The average family size (for population aged 16 and over) for the families that had negative weights was 4.3 which was approximately twice the national average family size (for population age 16 and over). These negative values seemed to indicate that some sets of subaggregate estimates in certain rotation groups possessed a high sampling variability due to differential coverages by the frame and rotation group. The effect of using biased estimates (due to CPS coverage problems) on the GLS procedure, which assumes that unbiased pre-adjustment weights were used, could also have contributed to the occurrence of negative adjusted weights. This paper deals with the problem that a sample weight has to be assigned to a family when only person level population controls are known. Some households and persons within the sample households have been left out of the sampling frame due to undercoverage. Adjustments for coverage losses are then crucial before using the GLS procedures for assigning weights.

The GLS distance measure D_1 can be written as

$$D_1(W,\Omega) = \sum_{i=1}^{n} W_i([W_i/\Omega_i]-1)^2$$

Luery (1986) indicated that an alternate distance measure that may reduce a systematic displacement of W_i/W_i from 1 (in this paper, the alternate distance measure will be used to provide nonnegative weights), might be

$$D_2(W,\Omega) = \sum_{i=1}^{n} W_i([W_i/\Omega_i]-a)^2$$

where 'a' is an estimate of the average adjustment factor, and can be estimated by:

$$a = \frac{\sum_{r=1}^{c} N_r}{\sum_{r=1}^{c}\sum_{i=1}^{n} X_{ir} W_i}$$

In this application, 'a' is the ratio of independent population controls to the estimates produced based on the family first-stage weights. The weights resulting from minimizing D_2 are

$$W = a\Omega + \Omega_0 X(X'\Omega_0 X)^{-1}(N - X'a\Omega)$$

It can be shown (Luery 1986) that these weights are invariant under scale change in Ω.

The 'a' values were calculated by rotation groups (RG). Since the W's in rotation groups 1-4 were all positive, the corresponding 'a' values were set equal to 1. The 'a' values for RG 5, 6, 7, and 8 are defined as: the ratio of State population (combined Oklahoma and Texas) controls and the corresponding State estimates based on family first-stage weights from the records in the two States in RG 5, the ratio of South Carolina's population control and its sample weight estimates in RG 6, and the ratio of Iowa and New Mexico's population controls and their sample weight estimates in RG 7. In RG 8, the initial 'a' value used was the ratio of California's population control and its sample weight estimate (a=1.08). The use of this 'a' value did not produce positive weights. After further investigation, it was decided to use the 'a' value of the lower 95 percent confidence limit on the average (over 8 RG) 'a' factor from California.

The 'a' values were multiplied to the family first-stage weights of the sample records in the States that resulted in negative weights. Following the application of Luery's alternative methodology, the GLS procedure was carried out as described in Section 4.2 and the elements of W based on the alternative distance measure D_2 were all found to be positive.

5. MULTIVARIATE RAKING RATIO ESTIMATION

Ireland and Scheuren (1975) developed a multivariate extension to the univariate raking ratio estimation (RRE) procedure created by Deming and Stephan (1940). This procedure is called Multivariate Raking Ratio Estimation (MRRE) and will iteratively adjust data from cells in a contingency table to agree with known marginal (control) totals. In RRE, each cell contains a scalar while in MRRE, each cell contains a vector. Consider the following, let $\underline{n}_{ij} = (n1_{ij}, n2_{ij})'$ be a 2-component vector of weighted sample counts corresponding to the cell positioned in the i^{th} row and j^{th} column of a 2-way contingency table.

It is desired to produce adjusted vectors $\underline{\boldsymbol{n}}_{ij} = (\boldsymbol{n}1_{ij}, \boldsymbol{n}2_{ij})'$ such that:

(i) $\sum_j \boldsymbol{n}1_{ij} = m1_{i.}$, $\sum_i \boldsymbol{n}2_{ij} = m2_{.j}$

(ii) Each component in a particular vector receive the same adjustment.

The MRRE procedure will proportionately adjust the weighted sample counts in each cell in the 2-way table the following way:

1. Row Rake - ratio adjust the vector components in the i^{th} row by $(m1_{i.}/n1_{i.})$ to produce:

$$\underline{\boldsymbol{n}}_{ij}(1) = \{m1_{i.}/n1_{i.}\}\underline{n}_{ij} = t_i(1)\,\underline{n}_{ij}$$

2. Column Rake - ratio adjust the vector components in the j^{th} column by $(m2_{.j}/\boldsymbol{n}2_{.j}(1))$ to produce:

$$\begin{aligned}\underline{\boldsymbol{n}}_{ij}(2) &= \{m2_{.j}/\boldsymbol{n}2_{.j}(1)\}\underline{\boldsymbol{n}}_{ij}(1)\\ &= \underline{\boldsymbol{n}}_{ij}\, u_j(1)\\ &= t_i(1)u_j(1)\,\underline{n}_{ij}\\ &= f_{ij}(1)\,\underline{n}_{ij}\end{aligned}$$

where $\boldsymbol{n}2_{.j}(1) = \sum_i t_i^{(1)} n2_{ij}$

The values $\underline{\boldsymbol{n}}_{ij}(1)$ and $\underline{\boldsymbol{n}}_{ij}(2)$ represent the adjusted vectors after the row (1^{st} rake) and column (2^{nd} rake) adjustment respectively. At the end of each rake, all components of the vector $\underline{n}_{ij}$ will receive the same adjustment factor ($t_i(1)$ for the row rake and $u_j(1)$ for the column rake). The completion of the row and column rake constitutes one iteration of MRRE.

Using the adjusted vectors generated from the previous iteration, steps (1) and (2) are repeated until either a specified number of iterations are completed or a specified level of 'closeness' between each sample total and control total is met.

After the g^{th} iteration, the quantity:

$$F_{ij}(g) = \prod_{h=1}^{g} t_i^{(h)} u_j^{(h)}$$

becomes the overall adjustment factor that is multiplied to the original weighted sample counts ($n1_{ij}$ and $n2_{ij}$) in vector $\underline{n}_{ij}$ in order to obtain the final adjusted vector $\underline{\boldsymbol{n}}_{ij}(g)$.

5.1 Properties of the MRRE Procedure

1. All components within the same vector receive the same adjustment factor.
2. Among all procedures satisfying conditions (i) and (ii) in section 5 (e.g., GLS family weighting), the MRRE procedure minimizes the statistic (Oh and Scheuren 1978):

$$M_B = \sum_{v,i,j} n_{vij}^{(g)} \ln \frac{n_{vij}^{(g)}}{n_{vij}}$$

where the summation is over each vector component, row, and column (indexed by v, i, and j respectively) in the contingency table.
3. Only nonnegative weights are produced.
4. Expressed informally, some of the necessary conditions for the convergence of MRRE are (Oh and Scheuren 1978):
a. The equivalent univariate raking procedure performed on the MRRE table of vector components will converge.
b. When setting up the contingency table, a 'sufficient' number of sample units should contribute to each marginal sample total.
c. There should not be 'substantial' differences between the expected marginal and the known marginal values.

5.2 Application of the MRRE Procedure

A MRRE weight was calculated for each sample family using the following procedure:
1. The first-stage weight was equalized for each record within a family by using the first-stage weight from the sample record belonging to the 'lowest' race/sex/age cell as the family first-stage weight. In order to reduce the number of iterations necessary for MRRE to converge, this weight was normalized (by rotation group) such that the sum of all family first-stage weights equal the overall control total.
2. Using this normalized first-stage family weight, a data set with one observation per unique family composition was created consisting of the sum of weighted counts of family members belonging to each of the 137 demographic cells. For example, one observation represented all families living in Maryland that are composed of Black males age 30-34 and Black females age 25-29. Families with the same composition but with different family sizes were represented in the same observation.
3. An 'order' of raking was determined by using a divergence measure developed by Kullback (1959) and used by Ireland and Scheuren (1975) and Oh and Scheuren (1978). This measure was defined as:

$$D_k = 1/n_k \sum_{h=1}^{n_k} [C_{k,h} - S_{k,h}^{(g)}] \ln [C_{k,h}/S_{k,h}^{(g)}]$$

where n_k = 51 for the average State measure, and 1 for the other 86 measures.

$C_{k,h}$ = Control total for the h^{th} level of the k^{th} demographic group. There are 51 levels for the State group and 1 level for each of the other 86 groups.

$S_{k,h}^{(g)}$ = Adjusted sample total for the h^{th} level of the k^{th} demographic group after the g^{th} rake.

D_k was computed for the 87 demographic groups (k=1,...,87). The first rake was performed by using the factor(s):

$$F_{k,h}^{(1)} = C_{k,h}/S_{k,h}^{(1)}$$

corresponding to the demographic group with the highest D_k value. The sample totals were then recomputed and the factor(s) corresponding to the demographic group with the highest value of the resultant divergence measure was used in the next rake.
4. The MRRE procedure was done independently for each rotation group so that a sampling variance, using a random group estimator, can be easily produced for analysis purposes.

6. DATA PRODUCED

SAS programs were written to calculate the March, Monthly, GLS, and MRRE weights for each sample family on the CPS 1984 microdata file.

6.1 Macro-level Estimates

Estimates were tabulated, using the weights derived from the Monthly, March, GLS, and MRRE family weighting procedures for the following characteristics: race x age x male household status (married-spouse-present, other male heads, and all other males); ethnicity x age; and tenure (owner-occupied and renter-occupied) x race x sex x family size; estimates of unemployment in families by type of family and race; and mean weekly earnings by type of family x race/ethnicity. Labor force estimates (employed, unemployed, unemployment rate, and not in labor force) by sex x race x ethnicity were tabulated based on the RRE, GLS, and MRRE procedures.

Standard errors were produced using a random group estimator, with the rotation groups used as random groups. The variance estimator is of the form (Wolter 1985)

$$\mathrm{Var}(Y_{est}) = \sum_{k=1}^{8} (8Y_k - Y_{est})^2/56$$

where Y_k = sum of the weights of sample records from the k^{th} rotation group with the characteristic Y and

$$Y_{est} = \sum_{k=1}^{8} Y_k$$

Because of a lack of design information on the CPS public use data file, this random group variance estimator was used even though it has certain weaknesses. These estimates were only used to measure relative variability between estimates based on weights derived by the different estimation procedures.

It should be noted that all family estimates produced in this investigation were based only on data from primary families (excluding related and unrelated subfamilies).

6.2 Micro-level Estimates

Sample records in the same control category for one marginal (State, age/sex/ethnicity, or age/sex/race) but in different control categories for another marginal most likely will not receive the same adjustment factor from GLS and MRRE. Therefore, to compare the GLS and MRRE procedures relative to the application of the adjustment factors, a summary of the sample distribution by coverage rate E/C (where E is the sample estimate weighted by the first-stage weight and C is the control) categories and by the adjustment factors derived by the GLS and MRRE procedures are tabulated.

The distribution of percent change in final weights after adjustment by the four procedures and a tabulation of two measures of closeness for evaluating estimation procedures (Fagan and Greenberg, 1985) are also tabulated.

The GLS measure is

$$M_A = \sum_{i=1}^{n} (W_i - \Omega_i)^2/\Omega_i$$

The raking measure is

$$M_B = \sum_{i=1}^{n} W_i \ln(W_i/\Omega_i)$$

where W_i = weight for sample record i prior to adjustment
and W_i = weight for sample record i following adjustment.

The measure M_A is a sum of weighted least squares and is minimized by applying the GLS procedure. The measure M_B has been shown to be minimized by MRRE when convergence of the marginal constraints is obtained. The measures were calculated by race, sex, and ethnicity.

7. RESULTS

(Space limitations do not permit inclusion of the tabulated data but are available upon request.)

7.1 Estimates

There were noticeable differences between the March family weighting (MAR) and GLS; MAR and MRRE; and MAR and Monthly family weighting (MTH) procedures for estimates of males subaggregated by household status x race x age.

Population estimates did not show any noticeable differences based on the four procedures when subaggregated to the ethnicity x age level. The absolute relative difference between the four procedures were all less than 1% (well below the estimated relative standard errors of the estimates).

Weighted family estimates by tenure, race, sex, and family size did not show any noticeable differences except for 4 and 5+ person owner-occupied families headed by nonwhite males. Here, both the GLS and MRRE produced higher estimates than the MAR and the MTH procedures.

For national labor force estimates by sex x race/ethnicity , absolute relative differences between the RRE and GLS, and RRE and MRRE were, for the majority of the characteristics, less than 1%. The absolute relative differences were largest for the unemployment levels across demographic characteristics. The largest absolute relative difference between the three procedures were detected for the estimated number of unemployed: Blacks (1.27%), Hispanics (2.67%), and female Hispanics (3.88%).

Unemployed family estimates by type and race based on the MTH and the MAR weighting did not show any noticeable differences. Larger absolute relative differences were obtained between the MAR and GLS, and MAR and MRRE; as high as 8.5%.

Estimates of mean weekly earnings were virtually the same across the four procedures. The most difference between weighting procedures occurred in Hispanic families ($329 for GLS and MRRE versus $318 for MTH and MAR).

7.2 Standard Error of Estimates

Standard errors for the GLS and MRRE procedures were generally much lower than MTH and MAR for estimates of males by household status x race x age and total population by ethnicity x age. The MAR and MTH produced similar standard errors while the GLS and MRRE were similar.

Estimated standard errors for the four procedures were basically the same for family estimates at the tenure x race x sex x family size level except for some noticeable differences in owner-occupied families headed by White males.

The absolute relative difference between the GLS and RRE estimates of standard error for national estimates of: total population labor force estimates were all less than 3.4%; White labor force estimates were all less than 13%; Black labor force estimates were all higher (26% for total Black UE); Hispanic labor force estimates by sex were also high (9%).

There were noticeable differences in standard errors between the current and alternative weighting procedures; particularly for estimates of total families with children under 18 years for total (all races) and Blacks.

After further investigation, it was found that among the 40,500 sample family records, about 90 of them have family weights over 5000 and contributed the most to this occurrence. Of the 90 records, 10 Black MSP records showed the greatest difference between weights based on the current procedures and weights based on the alternative procedures. In some cases, the GLS and MRRE weigthts were twice as large as the one's for MTH and MAR. This may be due to the coverage differences between females (whose second-stage weights were used in the MTH and MAR for the MSP males) and Black males (whose GLS and MRRE weights reflect all Black males in the same demographic control cell).

Except for Black and Hispanic families, there were no noticeable differences between the four weighting procedures for estimates of mean weekly family earnings.

Some of the differences in standard errors, especially in the person estimates, may be due to the fact that the GLS and MRRE weights were generated independently by rotation group (used as the random groups) while the MAR weights for males were produced based on the entire sample. The similarity in the MTH and MAR estimates and standard errors may be due to the fact that they both use the same weighting procedure for over 80% of the sample (the part consisting of all females and MSP-males).

7.3 Distribution of Weight Adjustments

Our data indicate that sample persons that were over or undercovered in the CPS (coverage rate not close to 1) seemed to be adjusted differently by GLS and MRRE.

The distribution of percent change between the first-stage weights and final weights were generally the same. The only exception was that GLS and MRRE showed a higher proportion of the sample receiving more than a 20% change in the weights.

The distribution of percent changes between the pre- and post- adjustment weights for MTH and MAR were similar.

7.4 Measures of Closeness

When compared to the MRRE, the GLS yielded larger values for both measures for the total sample, even Measure A which GLS is designed to minimize. The GLS did produce smaller values for some subaggregates of minority population in the GLS measure. The GLS also produced the smallest value for Measure B for females. It should be noted that the MRRE investigated in this paper did not converge to all the control cells. The extent of the nonconvergence was very small; less than 1.89% in each rotation group.

The value of Measure B (the raking measure) for the total sample of primary families based upon the MRRE weights was less than the corresponding value based upon the GLS weights. Comparing to the GLS, the MRRE produced smaller values of Measure B for subaggregates except for females.

One of the possible reasons that GLS produced the largest values for Measure A (the GLS measure) among the four estimation procedures was that the MTH, MAR, and MRRE investigated in this paper did not produce sample totals that agreed with the 137 population controls (as did the GLS). Therefore, they may produce smaller values for Measure A. Another possible reason was that only primary families (approximately 18,000 subfamilies and nonfamily households were excluded) or persons were considered in the calculation of the two measures.

When an adjustment procedure produced the smallest measure of closeness value for the total sample, it would be a desirable property if it also produced smaller values for subaggregates of the sample. When compared to the GLS, the MRRE produced smaller measures of closeness values for the total sample, but yielded larger values for a few subaggregates (Blacks, Hispanics, and females in measure A and females in Measure B).

In general, the MRRE yielded smaller values than the GLS procedure for both measures (24 out of 28 values for primary families and 13 out of 14 values for persons).

8. SUMMARY AND CONCLUSIONS

In the two alternative family weighting procedures investigated in this paper (the GLS and MRRE) the additional information of the composition of the family (i.e. the number of persons in the family with different demographic characteristics) controls how the adjusted weights are assigned . It appears that the GLS and MRRE tend to give similar results, and also that the March and Monthly weighting tend to be similar to one another. Based on the goals and results of this investigation, the authors recommend the MRRE procedure for family weighting.

8.1 Principal Person Weighting

The authors do not recommend the use of principal person weighting for the CPS family weighting. The two principal person weighting procedures evolve from a basic coverage assumption -- that the females generally have better coverage than the males. While principal person weighting addresses the issue of undercoverage, the implementation of the principal person weighting procedure in CPS is somewhat ad hoc and arbitrary. The post-stratified estimation cells are the result of a sequential change in requirements rather than an overall review of all the requirements. In addition, it is difficult to fit principal person weighting into a statistical model with its estimators having desirable statistical properties.

The authors feel that the principal person method investigated in this paper may be underestimating the number of MSP and OMH families and overestimating the number of AOM's. Also the March family weighting, unlike the GLS and MRRE, does not produce family estimates that agree with all of the 137 demographic person controls in the CPS second-stage weighting procedure (they do agree with a subset of the 137 demographic person controls). The Monthly family weighting does not control the male sample estimates to meet any Census population controls. Finally, the two procedures do not restrain each person's weight within a family to be equal.

8.2 GLS

The possibility of negative weights curtails the recommendation of implementing the GLS procedure in the CPS.

The GLS procedure assumes that the pre-adjustment weights produce unbiased population estimates (i.e., assume perfect survey coverage on the average). Separate weighting adjustments prior to applying the GLS are needed to correct for differential undercoverage of households or individuals within households. Without the adjustment for systematic undercoverage, the GLS will probably not perform optimally.

8.3 Multivariate Raking Ratio Estimation

The authors recommend that the MRRE be implemented as the CPS family weighting procedure. The MRRE is statistically defensible with its estimators having desirable statistical properties such as the minimization of a distance measure and the guarantee of positive adjusted weights. Based on this investigation, the MRRE produced smaller standard errors and closeness measure values than GLS and produced labor force and other person estimates that were similar to those based on the second-stage weights.

The issue of convergence and the possibility of extreme positive weights seem to be the only concerns, and while they are not trivial, the current CPS second-stage procedure has similar problems. Due to differential coverage of persons in the CPS, adjustments such as the normalization of first-stage weights are recommended prior to running MRRE. As part of the implementation strategy, the authors suggest that more efficient computer programs be developed for executing MRRE.

The MRRE methodology investigated in this paper could be extended to include a set of independent family control counts (if and when they become available). Although problems may arise with consistency between the independent person and household counts, this extension may improve the accuracy of the family estimates while at the same time, not seriously distort the person estimates.

Acknowledgement

The authors would like to express appreciation to Ken Copeland and Joe DiGennaro for their contributions in the preparations of the input data file to the generalized least squares estimation procedure.

REFERENCES

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, 11, 427-444.

Fagan, J., and Greenberg, B. (1984). "Making Tables Additive in the Presence of Zeros." Proceedings of the American Statistical Association. Section of Survey Research Methods.

Hanson, R.H. (1978). The Current Population Survey Design and Methodology, Technical Paper 40, U.S. Bureau of the Census.

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

Ireland, C.T. and Scheuren, F.J. "The rake's progress", The George Washington University, Computer Programs For Contingency Table Analysis, 155-216, 1975

Luery, D. (1980). "An Alternative to Principal Person Weighting. Draft Memorandum. U.S. Bureau of the Census.

Luery, D. "Weighting Sample Survey Data Under Linear Constraints on the Weights." Presented at the Conference on CE Weighting Methods, March 18, 1986.

Oh, H. Lock and Scheuren, F. J. (1978). "Multivariate Raking Ratio Estimation in the 1973 Exact Match Study." Proceedings of the Section on Survey Research and Methodology, American Statistical Association Meetings.

Neyman, J. (1949). Contribution to the theory of the X^2 Test. Proceedings 1st Berkeley Symposium, 239-273

Wolter, K.M. (1985). Introduction to Variance Estimation. Springer-Verlag.

Zieschang, K.D. (1986). "A Generalized Least Squares Weighting System for the Consumer Expenditure Survey." Proceedings of the Section on Survey Research Methods, American Statistical Association Meetings, 64-71.