BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: Census/SRD/RR-87/16

SOME ASPECTS OF ESTIMATING VARIANCES BY HALF-SAMPLE REPLICATION IN CPS

by

Lawrence R. Ernst and Todd R. Williams
Statistical Research Division
Bureau of the Census
Washington, D.C. 20233

Recommended by: Kirk M. Wolter

Report completed: August 5, 1987

Report issued: December 23, 1987 (Revised)

1. INTRODUCTION

The focus of this paper is an empirical study of certain aspects of variance estimation using a replication approach for the Current Population Survey (CPS). The CPS is a monthly labor force survey of approximately 60,000 U.S. households drawn from a multistage stratified design, with one primary sampling unit (PSU) per stratum.

There have been several previous studies of variance estimators which used data from complex surveys. For example, in Frankel (1971) and Bean (1975), CPS data and Health Interview Survey data were used respectively. The approach taken in this paper has at least one fundamental difference from the previous studies. In the works cited, the sample from the complex survey was treated as if it were the population of interest. Samples were selected from the full sample and variance estimates computed from the subsamples. In this paper, the full CPS sample is viewed, as it actually is, a sample from a national population. Consequently, the variance estimates computed here are for the full sample.

The two approaches each have advantages and disadvantages. The chief advantage of the first approach described is that since a known population is assumed, such key information as estimates of biases in the variance estimators can be directly computed, while in this paper it cannot. On the other hand, the results in the previous studies only apply directly to the relatively small samples choosen from the artificial populations. It is generally not evident how well the results also apply to variance estimates for the full sample.

The following are some of the principal areas investigated in this study.

A. A comparison of reweighting each replicate as opposed to using the parent sample weights for all replicates.

B. The constants to be used in the collapsed stratum estimator to reduce the bias of this estimator.

C. A comparison of random replication and partially balanced replication.

D. The effect of the number of replicates on the precision of the variance estimates.

The items just listed, along with other aspects to be studied, are described in detail in Section 2. This section also includes a description of the form of the variance estimator considered here. The numerical results are presented and analyzed in Section 3.

2. TOPICS OF STUDY

The general form of the variance estimator studied in this paper is explained in Section 2.1. In the remaining subsections each of the specific aspects to be studied is described.

2.1 The Replicated Variance Estimator

For one PSU per stratum designs like CPS, a collapsed stratum variance estimator is generally employed as explained in Wolter (1985). We begin by reviewing this form of variance estimation, using the notation of Wolter for the most part, and then explain how it is used in this paper in conjunction with a replicate variance estimator.

The first step in using a collapsed stratum estimator is the partitioning or "collapsing" of the set of all strata into groups of two or more strata. Then consider a population total Y that is estimated by a linear estimator of the form

$$\hat{Y} = \sum_{g=1}^{G} \hat{Y}_g = \sum_{g=1}^{G} \sum_{h=1}^{L_g} \hat{Y}_{gh},$$

where G denotes the number of groups of collapsed strata; L_g the number of strata in the g-th group; $\hat{Y}_g$ the estimator of total for the g-th group; and $\hat{Y}_{gh}$ the estimator of total for the h-th stratum in the g-th group. The general form of the collapsed stratum variance estimator is then

$$v_{cs}(\hat{Y}) = \sum_{g=1}^{G} \frac{L_g}{L_g - 1} \sum_{h=1}^{L_g} (\hat{Y}_{gh} - \frac{A_{gh}}{A_g} \hat{Y}_g)^2, \qquad (2.1)$$

where A_{gh} is a known measure associated with the gh-th stratum that tends to be well correlated with $\hat{Y}_{gh}$, and $A_g = \sum_{h=1}^{L_g} A_{gh}$. Commonly used values for A_{gh} include $A_{gh}=1$ for all g,h and $A_{gh} = p_{gh}$ where p_{gh} is the population of the gh-th stratum from the most recent census. The terms $v_{cs}(\hat{Y})$ and A_{gh} will be discussed further in Section 2.3.

In the CPS there are 379 nonself-representing strata, which we partitioned into 188 pairs of strata and one group of three strata. There are also 350 self-representing strata. To take into account the variability arising from sampling from these strata, the sample in each of them is divided into two panels, with the assignment of ultimate sampling units alternating between the panels. In applying (2.1), the two panels corresponding to each self-representing strata are treated as if they constituted a pair of nonself-representing strata collapsed together. Thus, G = 539 for the entire sample, with $L_g=3$ for one group and $L_g=2$ for all other groups.

Returning now to (2.1), it can be shown that this is algebraically equivalent to

$$v_{cs}(\hat{Y}) = d \sum_{g=1}^{G} \frac{1}{L_g} \sum_{h=1}^{L_g} [(1 + \frac{L_g}{(L_g - 1)^{1/2} d^{1/2}} (1 - \frac{A_{gh}}{A_g})) \hat{Y}_{gh}$$

$$+ \sum_{\substack{t=1 \\ t \neq h}}^{L_g} (1 - \frac{L_g}{(L_g - 1)^{1/2} d^{1/2}} \frac{A_{gh}}{A_g}) \hat{Y}_{gt} - \hat{Y}_g]^2, \qquad (2.2)$$

where d is a parameter introduced by Fay, with different notation (see Dippo, Fay and Morganstein (1984)), that leads to a more general form of the replicate variance estimator than the standard form for which d=1. This parameter is discussed in Section 2.4. The form of the replicate variance estimator, $v_k(\hat{Y})$, considered in this paper is

$$v_k(\hat{Y}) = \frac{d}{k} \sum_{\alpha=1}^{k} (\hat{Y}_\alpha^R - \hat{Y})^2, \qquad (2.3)$$

where k is the number of replicates and each replicate estimate $\hat{Y}_\alpha^R$ is obtained as follows. Corresponding to each α and each group g, a stratum gh is selected from the g-th group. Then

$$\hat{Y}_\alpha^R = \sum_{g=1}^{G} [(1+ \frac{L_g}{(L_g-1)^{1/2} d^{1/2}} (1- \frac{A_{gh}}{A_g}))\hat{Y}_{gh} + \sum_{\substack{t=1 \\ t\neq h}}^{L_g} (1- \frac{L_g}{(L_g-1)^{1/2} d^{1/2}} \frac{A_{gh}}{A_g})\hat{Y}_{gt}]. \qquad (2.4)$$

Now provided that for all g, each stratum in the g-th group is selected k/L_g times, (2.3) reduces to (2.2) plus a sum of cross-product terms involving the bracketed portion of (2.4) from pairs of groups. If additionally, each pair of strata gh and g'h' from two groups are selected together $k/L_gL_{g'}$ times then the cross-product terms cancel and (2.4) reduces to (2.2). These assertions are all explained in Borack (1971) and Wolter (1985) for the case when the L_g are the same for all g, but the concept is not limited to only that case. A set of replicates satisfying these conditions is said to be in full orthogonal balance.

For linear estimators, there is no particular advantage to computing variance estimates using (2.3), since (2.1) can be computed directly just as readily. However, as explained in Section 2.2, CPS estimators using the final weights are nonlinear estimators even for estimates of totals. Expressions such as (2.3) are used to estimate variances for nonlinear estimators also. The previous empirical studies cited in the Introduction support the use of this approach as do certain asymptotic results, such as those of Krewski and Rao (1981).

The particular topics to be studied here derive from the many specific forms that (2.3) and (2.4) can take for CPS data. For estimators of total using the final CPS weights, $\hat{Y}_{\alpha}^{R}$ can be computed in several ways, as explained in Section 2.2. The different possible values for A_{gh} and d are discussed in Sections 2.3 and 2.4, respectively. Finally, in Section 2.5, two less expensive alternatives to a fully balanced set of replicates, partially balanced replication and random replication are considered, along with the question of number of replicates to be used.

2.2 Weighting the Replicates

The final weights used in CPS are obtained by beginning with the reciprocal of probability of selection for each sample unit, which we will refer to as the base weight, and then subjecting the set of weights to three successive adjustments: the noninterview adjustment, the first-stage ratio adjustment and the second-stage ratio adjustment. Of these adjustments, the second-stage ratio adjustment generally has the largest impact on both the expected values and the variances of the estimates (Hanson 1978). The adjustment for the population 16 years and older, which is the one of interest here, uses the following procedure (Jones 1984). First the sample weights after the first-stage adjustment are ratio adjusted to obtain estimates that agree with independently derived estimates of the total population for that month in each of the 50 states and the District of Columbia. The resulting weights are then further ratio adjusted to obtain agreement with independently derived national estimates in 16 age/Hispanic ethnicity/sex cells. Finally, these weights are adjusted again to obtain agreement with independent national estimates in 70 age/race/sex cells. Note that each successive adjustment destroys the agreement with the independent estimates controlled to in the previous adjustment. The entire procedure is therefore repeated five more times. This repeated iteration of the procedure, a process known as iterative proportional fitting or "raking," results in a set of final weights which yields estimates in near agreement with all three sets of controls.

For the replication method of estimating variances, each replicate is subject to the same weighting procedures as the parent sample. That is, to obtain a final value for $\hat{Y}^R_\alpha$, first compute (2.4) using the base weights to obtain estimates of strata totals and then perform the same ratio adjustments that are done for the parent sample. As one might expect from the length of the second-stage adjustment just described, this can require extensive computer time. A short cut would be to use the final weights from the parent sample for each replicate; that is, $\hat{Y}^R_\alpha$ would be computed directly from (2.4) using the final weights to obtain the estimates for the strata totals. The effectiveness of this short cut has been studied previously by a number of authors, including Bean (1975), who found it produced little loss in accuracy, and Lemeshow (1979), who found evidence of greater bias and lower precision for variance estimates computed using the parent sample weights.

For this part of the study, variance estimates were computed using three different approaches to account for the weighting. The first two are the Reweighting method and the Parent Sample Weights method that we have been discussing. (Actually to simplify matters for the Reweighting method, only the second-stage weights are replicated; that is, the computation of a replicate estimate begins by computing (2.4) using the first-stage weights from the parent sample). The final method, the Base Weights method, simply uses the base weights in the replicate estimates in order to allow for a comparison of variance estimates using unadjusted weights to those based on the other two procedures.

For the Reweighting method, 6 cycles of raking are used.

Since some cost savings would ensue if fewer cycles were used, variances estimates were also obtained for 1, 2, and 3 cycles for the purpose of determining if the variances estimates would be substantially affected by fewer cycles.

The question of reweighting versus not reweighting replicate estimates is one area where analytic results that provide some insight into the problem can be presented. To achieve this we assume the following simple situation. $Y = \sum_{i=1}^{N} y_i$ is the total for a population characteristic for a group of known size N, which is to be estimated by a sample of variable size n, with $E(n) = n_o$. Furthermore, the sample is self-weighting with sampling fraction n_o/N. Two estimators of Y are then,

$$\hat{Y}_{UN} = \frac{N}{n_o} \sum_{i=1}^{n} y_i \ ,$$

$$\hat{Y}_{RAT} = \frac{N}{n} \sum_{i=1}^{n} y_i \ .$$

$\hat{Y}_{UN}$ is the simple unbiased estimator and $\hat{Y}_{RAT}$ is a ratio estimator that adjusts the sample estimate of the number of people in the group of interest to the control total, N. If this was the only weighting adjustment that was done then we can view $\hat{Y}_{UN}$ and $\hat{Y}_{RAT}$ as analogous to a CPS estimator before and after the second-stage adjustment. We proceed to first show that under certain conditions

$$V(\hat{Y}_{RAT}) < V(\hat{Y}_{UN}), \qquad (2.5)$$

and that $V(\hat{Y}_{UN})$ and the expected value of the replicate variance estimator for the Parent Sample Weights methods are generally approximately equal. Since, under appropriate assumptions, this is also true by asymptotic results such as those of Krewski and Rao (1981) for $V(\hat{Y}_{RAT})$ and the replicate variance estimator for the Reweighting method, the amount by which the right side of inequality (2.5) exceeds the left provides some indication of the bias in using the Parents Sample Weights method to estimate $V(\hat{Y}_{RAT})$.

To obtain expressions for $V(\hat{Y}_{UN})$ and $V(\hat{Y}_{RAT})$, we first abbreviate $\bar{y} = \sum_{i=1}^{n} y_i/n$, $\bar{Y} = Y/N$, and assume that the sample design is such that $V(\bar{y}|n) \doteq \sigma_y^2/n$; that is, the design effect is approximately 1. Then

$$V(\hat{Y}_{UN}) = \frac{N^2}{n_o^2}[V(n\bar{y}|n) + V(n\bar{Y})]$$

$$\doteq \frac{N^2}{n_o^2}[n_o\sigma_y^2 + \bar{Y}^2V(n)]$$

$$= \frac{N^2\sigma_y^2}{n_o} + \frac{N^2\bar{Y}^2V(n)}{n_o^2} .$$

Furthermore, if the distribution of n is such that $E(1/n) \doteq 1/n_o$, then

$$V(\hat{Y}_{RAT}) = N^2V(\bar{y}) \doteq N^2\sigma_y^2\ E(\tfrac{1}{n}) \doteq \frac{N^2\sigma_y^2}{n_o},$$

and hence

$$\frac{V(\hat{Y}_{UN})}{V(\hat{Y}_{RAT})} \doteq 1 + \frac{\bar{Y}^2V(n)}{n_o\sigma_y^2} . \qquad (2.6)$$

With the assumption that the second term on the right side of (2.6) is at least of order 1, then (2.6) establishes (2.5).

In particular, if y_i is a 0-1 variable, as it is for all the characteristics of interest in this paper, and $\bar{Y} = p$, then (2.6) reduces to

$$\frac{V(\hat{Y}_{UN})}{V(\hat{Y}_{RAT})} \doteq 1 + \frac{V(n)p}{n_o(1-p)} .$$

Thus the gains in precision by using $\hat{Y}_{RAT}$ instead of $\hat{Y}_{UN}$ increase with increasing p.

We next establish that the expected value of the replicate variance estimator for the Parent Sample Weights method is approximately $V_{UN}(\hat{Y})$ for the following simple situation. Assume the sampling design is two PSUs per stratum with replacement, with y'_{ij}, j = 1,2, denoting the unweighted sum of the characteristic values for all sample units in the j-th sample PSU in the i-th stratum. Then the expected value of (2.3) with $\hat{Y} = \hat{Y}_{RAT}$, but $\hat{Y}^R_\alpha$ computed with the Parent Sample Weights method, reduces to

$$E[v_k(\hat{Y}_{RAT})] = N^2 \sum_{i=1}^{L} E\left(\frac{y'_{i1} - y'_{i2}}{n}\right)^2 ,$$

where L is the number of strata. Furthermore, although n is not independent of y'_{i1} and y'_{i2}, in general if L is large enough,

$$E\left(\frac{y'_{i1} - y'_{i2}}{n}\right)^2 \doteq E\left(\frac{1}{n^2}\right) E(y'_{i1} - y'_{i2})^2 .$$

If additionally, $E(1/n^2) \doteq 1/n_0^2$, it follows that

$$E[v_k(\hat{Y}_{RAT})] \doteq \frac{N^2}{n_0^2} \sum_{i=1}^{L} E(y'_{i1} - y'_{i2})^2 = V(\hat{Y}_{UN}) .$$

Thus, in this simple situation, whatever gains in precision $\hat{Y}_{RAT}$ has over $\hat{Y}_{UN}$ are generally lost when a replicate estimator together with the Parents Sample Weights method is used to estimate $V(\hat{Y}_{RAT})$.

2.3 Values for A_{gh}

The collapsed stratum variance estimator, like any variance estimator for one PSU per stratum designs, is biased. In Hansen, Hurwitz and Madow (1953), Volume II, Chapter 9, it is established that for a linear estimator $\hat{Y}$ with $v_{cs}(\hat{Y})$ as in (2.1),

$$\text{Bias}\,[v_{cs}(\hat{Y})] = \sum_{g=1}^{G} \Big[\frac{L_g}{L_g - 1} \sum_{h=1}^{L_g} \Big(\frac{A_{gh}^2}{A_g^2} - \frac{2\,A_{gh}\,\sigma_{gh}^2}{A_g\,\sigma_g^2}\Big) + \frac{1}{L_g - 1}\Big]\,\sigma_g^2$$

$$+ \sum_{g=1}^{G} \frac{L_g}{L_g - 1} \sum_{h=1}^{L_g} \Big(Y_{gh} - \frac{A_{gh}}{A_g}\,Y_g\Big)^2, \qquad (2.7)$$

where $\sigma_{gh}^2 = \text{Var}(\hat{Y}_{gh})$, $\sigma_g^2 = \sum_{h=1}^{L_g} \sigma_{gh}^2$, $Y_{gh} = E(\hat{Y}_{gh})$ and $Y_g = E(\hat{Y}_g)$.

Two commonly used values for A_{gh} for the nonself-representing strata for surveys such as CPS are $A_{gh} = 1$ and $A_{gh} = p_{gh}$, where p_{gh} is the population of the gh-th stratum from the most recent census. $A_{gh} = 1$ is the natural choice if, ignoring the original stratification, the L_g PSUs in the g-th group are treated as independent selections from a single stratum. In this case only the second term in (2.7) is present; that is, the bias would consist only of a between strata component. If Y_{gh} is well correlated with p_{gh}, then the second term in (2.7) can generally be reduced by the use of $A_{gh} = p_{gh}$ and would disappear if Y_{gh} is proportional to p_{gh}. The first term, however, would no longer be zero. Furthermore with $A_{gh} = 1$, the bias must always be upward, while with $A_{gh} = p_{gh}$ it is possible for the bias to be downward since the first term can be negative. For a nonlinear estimator computed using (2.2) and (2.3), no such blanket statements can be made about the direction of the bias.

For the self-representing strata, $A_{gh} = 1$ is always used, since the two panels correponding to each such stratum have the same expected size.

In this paper variance estimates are computed using both $A_{gh} = 1$ and $A_{gh} = p_{gh}$ for nonself-representing strata, and compared.

2.4 Values for d

The standard form of the replicate variance estimator, as presented in Wolter (1985), only considers expressions like (2.3) for $d = 1/(L_g - 1)$. The more general form was introduced in Dippo, Fay and Morganstein (1984), with the following motivation. In (2.2) the factor multiplying the estimated $\hat{Y}_{gt}$ if the gh-th stratum is selected, $h \neq t$, is

$$1 - \frac{L_g}{(L_g-1)^{1/2} d^{1/2}} \frac{A_{gh}}{A_g} . \qquad (2.8)$$

For $d = 1$, this factor is 0 with $L_g = 2$ and $A_{gh} = 1$, and can, be negative for other combinations of L_g and A_{gh}. A negative value for (2.8) can result in negative values for replicate estimates computed using the Reweighting method even when the full sample estimate cannot be negative, an undesirable situation. Furthermore, as noted in Dippo, Fay and Morganstein, (2.8) must be strictly positive to ensure that complex functions built from ratios would be defined for each replicate whenever the function could be computed for the whole sample. To avoid these difficulties, Fay suggests $d = 4$ as an alternative. For $d = 4$, $L_g = 2$, (2.8) is positive for any set of positive A_{gh}. For $d = 4$, $L_g = 3$, (2.8) is positive for $A_{gh} = 1$, and also for $A_{gh} = p_{gh}$ as long as $A_{gh} < 2^{3/2} A_g/3$ for all g and h, as it is in this study.

Variance estimates obtained from (2.3) are clearly the same for all d for linear estimators. Furthermore, even for nonlinear estimators, under appropriate conditions, the variance estimators, treated as a function of d, asymptotically converage to the same estimators for all d.

In this paper the effects of different d on the variance estimates for the characteristics of interest are studied for the Reweighting method only, since variance estimates obtained using the Base Weights and Parent Sample Weights methods are identical for all d. Variance estimates were computed for d=1, 4, 100 and 10,000. d=100 and d=10,000 are included to provide some insight on the effects of large values of d.

2.5 Random Replication Versus Partially Balanced Replication

As explained in Wolter (1985), for a linear estimator the replicate variance estimator is identical to the standard variance estimator when a fully balanced set of replicates is used. However, the number of replicates k in a fully balanced set must always be at least G, and the cost of processing may be too high for this many replicates. If a smaller number of replicates is required, the selected strata in each replicate may be chosen at random, or, alternatively, the set of replicates may be constructed to yield a partially balanced set, as described in Wolter. For linear estimators, both approaches result in unbiased variance estimators. However, the variances of the variance estimators with either method are in general higher than for a fully balanced set because of the presence of cross-product terms. Furthermore, as explained in Wolter, many of the cross-products terms are removed by partial balancing, and as a result, for linear estimators at least, replicate variance estimators using this method generally have higher precision then with random replication.

In this paper these two methods of obtaining a set of replicates are compared to each other and also evaluated as a function of the number of replicates. Each combination of the three weighting methods and two sets of A_{gh} is used in this comparison. For each combination, variance estimates were computed several times in order to obtain estimates of the standard errors of the variance estimates over all possible random replications, and all possible groupings of the collapsed strata for partial balancing. For random replication, different random replicates were generated for each repetition, while for partial balancing, the arrangement of the collapsed strata into groups of collapsed strata was randomized.

For the Reweighting method, computing variance estimates more than once in this way also serves another purpose. For the numerical comparisons to be presented in Section 3 for the topics discussed in Sections 2.2 - 2.4, the variance estimates were averaged over the trials, thereby reducing the variability of the variance estimates arising from the cross-product terms, for both partial balancing and random replication. For the other two weighting methods, instead of averaging the variance estimates, this source of variability in the variance estimates was completely removed by computing the expected values of the variance estimates directly from (2.1). Consequently, it is the expected values of the variance estimates that are used for the Parent Sample Weights and Base Weights methods for the numerical comparisons of the topics described in Sections 2.2 - 2.4. The computation of variance estimates by partial balancing and random replication was done for the sole purpose of estimating the variances of the variance estimates arising from the use of a set of replicates that are not fully balanced.

3. EMPIRICAL RESULTS

We first describe the variance estimates that were computed. As detailed in the previous section, the following were varied.

1. Weighting methods: Reweighting (with 1, 2, 3, and 6 raking cycles), Parent Sample Weights, Base Weights.

2. A_{gh}: 1, p_{gh}

3. d: 1, 4, 100, 10,000

4. Set of replicates methods: Partial balancing, random replication.

5. k: 12, 24, 48

For the Parent Sample Weights and Base Weights methods, variance estimates were computed for each combination of the other aspects listed, with the exception that only one value of d was used, since variance estimates for these weighting methods are independent of d. For each combination, 50 estimates were obtained, with different groupings of the strata for the partially balanced method, and different random replications. In addition, for these two weighting methods, the variance estimates corresponding to a fully balanced set of replicates were computed directly from (2.1) for both sets of A_{gh}.

For the Reweighting method, the combinations for which variance estimates were computed are presented in Table 1. For each of the indicated combinations, 10 estimates were obtained. The principal reason that all combinations were not considered for the Reweighting method and that more estimates were not computed for each combination is simply that it is much more expensive to compute variance estimates for this method. Also, combinations for which $A_{gh} = p_{gh}$ and d=1 were omitted because of the potential problems discussed in Section 2.4.

The estimates for which variance estimates were computed are all estimates of population totals. The specific characteristics estimated are the same for all aspects of the study, and are listed in Tables 2-7.

The first comparisons are for the three weighting methods, with the computations summarized in Table 2 for each weighting methods and A_{gh} combination. For the Parent Sample Weights and Base Weights methods, the variance estimates listed are those computed directly from (2.1), so that the variability in the replicate variance estimates that would otherwise arise from the cross-product terms has been eliminated. For the Reweighting method, the variance estimates listed for $A_{gh} = p_{gh}$ are the simple average of the 20 repetitions for which k=24, d=4 and either partial balancing or random replication was used. For $A_{gh} = 1$, the estimates are averaged over the 10 repetitions for which d=4. (Refer to Table 1.) Variance estimates from other possible combinations were not used in computing the average, because they were not independent of the repetitions that were used. The standard errors of the variance estimates arising from the choice of the set of replicates for the Reweighting method for each set of A_{gh} is also presented in Table 2. For $A_{gh}=1$ the estimates of the standard errors of the variance estimates were computed by considering the 10 repetitions to be independent, equal probability selections, while for $A_{gh} = p_{gh}$, the sets of partially balanced and randomly selected replications were considered separate strata in this computation.

Note that the estimates of the standard errors of the variance estimates reflect the variability in the variance estimates for the Reweighting method arising from the variability in the chosen set of replicates, but does not reflect any of the other possible sources of error in the computation of the variance estimates. For example, the bias in the collapsed stratum variance estimator, and the variability in the variance estimates that would result from a different CPS sample, are not measured. Furthermore, these sources of error in the variance estimates affect all three weighting methods. Consequently, the results in the tables must be interpreted with caution.

The following are key observations from Table 2 concerning the weighting methods. For those characteristics possessed either by a large proportion of the total population, or a large proportion of a demographic subgroup which is controlled to in the second-stage adjustment, the variance estimates appear to be lower for the Reweighting method than the Parent Sample Weights method. This includes total, black and teenage employed, and in labor force. This is in accord with the results in Section 2.2. For other characteristics, such as the unemployment characteristics, for which the proportion of the total population or the indicated demographic subgroup possessing the characteristic is small, differences between the variance estimates computed with the two weighting methods are generally not as dramatic.

The Parent Sample Weights and Base Weights methods were also compared. For each A_{gh} and characteristic combination, the entry in Table 2 for the Base Weights method is lower than for the Parent Sample Weights method. If this is indicative of significant differences between these two methods, it may be do to the following. As noted in Section 2.2, the gains in actual variances arising from the second-stage adjustment, may not be reflected in the variance estimates when the Parent Sample Weights methods is used. In fact, variance estimates for this method are computed in the same manner as the Base Weights method, but the weights used with the Parent Sample Weights method are more variable due to the second-stage adjustment, and generally larger due to the undercoverage that the second-stage adjustment seeks to correct. More variable and larger weights tend to increase variance estimates, although in the case of larger weights, not necessary relative variances. Thus, ironically, by performing the second-stage adjustment, which has increased precision of the estimates as one of its goals, and then using the Parent Sample Weights methods to compute variance estimates, larger variance estimates may result than if the second-stage adjustment had not been done at all.

The results when using the Reweighting method with fewer than six cycles of raking are presented in Table 3. The variance estimates for two cycles and even possibly for one cycle appear to be close enough to the variance estimates for six cycles to be viable approximations.

We next consider the effect of the choice of A_{gh} on the variance estimates. Examining Table 2 again, we note that most of the entries for $A_{gh} = p_{gh}$ are lower than the corresponding entries for $A_{gh} = 1$. For the Reweighting method, however, the differences would generally not be significant, even if the standard errors of the variance estimates given in Table 2 are assumed to be the only source of error. We suspect that this is at least partly due to the small number of repetitions done for the Reweighting method.

For the Base Weights method, an estimator of total is a linear estimator, and consequently (2.7) is an exact expression for the bias of the variance estimator. If the variance estimates are actually smaller for $A_{gh} = p_{gh}$ and (2.7) is positive, then $A_{gh} = p_{gh}$ does result in lower biases than $A_{gh} = 1$. Furthermore, for estimates for which it is additionally true that the second-stage adjustment does lower the variances, but for which this is not reflected in the variance estimates computed with the Parent Sample Weights method, $A_{gh} = p_{gh}$ results in smaller biases for this weighting method also.

There is a further complication in comparing the two sets of A_{gh}. Different sets of collapsed strata were used for the two sets of A_{gh} for the variance estimates summarized in Table 2. This arose because collapsing was done in an attempt to minimize an average over several key characteristics of the bias expression (2.7). This is described fully in Ernst, Huggins and Grill (1986). Since (2.7) involves A_{gh}, different A_{gh} lead to different optimal collapsings. Consequently, Table 2 reflects not only the effect of the different A_{gh} but also the different sets of collapsed strata.

In an attempt to learn something about this matter, variance estimates were also computed with the A_{gh} and the sets of collapsed strata reversed, with the results presented in Table 4. That is, variance estimates were obtained with $A_{gh} = p_{gh}$ for the collapsed strata optimal for $A_{gh} = 1$ and vice versa. Comparing Tables 2 and 4 for the Base Weights and Parent Sample Weights methods, we note that for the same A_{gh} entries in Table 2 are generally lower than the corresponding entries in Table 4; that is the entries are lower for the set of collapsed strata that was chosen to be optimal with the particular A_{gh}, as one might expect. The most striking observation, however, is that for these two weighting methods, for characteristics possessed by a large proportion of the total population, that is total employed, and total in labor force, the entries in Table 4 for $A_{gh} = 1$ are much larger than the corresponding entries in Table 2 for $A_{gh} = p_{gh}$. That is, for these characteristics at least, the substitution of $A_{gh} = 1$ for $A_{gh} = p_{gh}$ with the set of collapsed strata optimal for $A_{gh} = p_{gh}$ may produce variance estimates that are severely biased upward. An explanation for this is that the optimal collapsing for $A_{gh} = p_{gh}$ tends to group strata together with total populations that vary more than the optimal collapsing for $A_{gh} = 1$, since the use of $A_{gh} = p_{gh}$ in the variance estimates can compensate for the biases that otherwise would result from the grouping of strata with different population totals. That is, for fixed g, the variability of Y_{gh}/A_{gh} with h, which is reflected in the second term of (2.7), will only arise for $A_{gh} = p_{gh}$ from differences in the proportions of the population possessing the characteristic among the strata collapsed together, not any differences in total population (assuming the strata populations remain in the same proportion from the point in time that p_{gh} was computed). However, when $A_{gh} = 1$ is used instead, the possibly large variability in the population of the strata collapsed together can increase the variability of the Y_{gh}/A_{gh}, and hence increase (2.7), particularly for characteristics possessed by a large proportion of the total population.

For the same two weighting methods, the effect of the opposite substitution, that is the use of $A_{gh} = p_{gh}$ instead of $A_{gh} = 1$ with the collapsing optimal for $A_{gh} = 1$, is not at all apparent. In fact, for many characteristics the substitution of $A_{gh} = p_{gh}$ for $A_{gh} = 1$ results in lower values for the entries in Table 4 with $A_{gh} = p_{gh}$ than for the corresponding entries in Table 2 with $A_{gh} = 1$.

Thus, it appears that for these weighting methods, Table 4 provides some evidence that it is the $A_{gh} = p_{gh}$ rather than the particular set of collapsed strata that lowers the variance estimates.

For the Reweighting method, the large variances of the estimated variances again severely limits what can be inferred from Table 4. There is, however, no evidence of any large increase in the variance estimates with this method when $A_{gh} = 1$ is substituted for $A_{gh} = p_{gh}$, as there is with the other weighting methods. A possible explanation is that the increase in the variability of the estimate of the total population that occurs for the other two weighting methods as a result of this substitution is completely removed by the reweighting.

We next consider the effects of different values of the d parameter on the variance estimates for the Reweighting method, with the results summarized in Table 5. Each entry in this table is obtained by taking the simple average of the 10 repetitions for which k=12 and partial balancing was used. For $A_{gh} = 1$ the table entries are all lower for d=4 than d=1. For $A_{gh} = p_{gh}$ the entries are lower for d=100 than d=4, while the entries for d=10,00 are close to d=100. Although these differences are generally not significant, it appears that the variance estimates are generally decreasing functions of d which converge to positive limits as d approaches ∞ . This is consistent with the findings in Judkins (1987) who provides an explanation for this relationship.

The results for the effects of partial balancing versus random replication and the number of replicates, k, on the population variances of the variance estimates is given in Table 6. The estimate of the variance of the variance estimate in a cell in this table is obtained by treating each of the 10 variance estimates for the Reweighting method and the 50 variance estimates for the other weighting methods computed for each cell as if they were obtained independently, with equal probability, from the set of all possible variance estimates. For the Parent Sample Weights and Base Weights methods,

$$\frac{1}{50} \sum_{i=1}^{50} (v_i - v_E)^2 \qquad (3.1)$$

is then the estimator used to estimate the variances of the variance estimates, where v_i, i=1,...,50 are the 50 variance estimates, and v_E is the expected value of the v_i, computed directly from (2.1). An estimator of the standard error of (3.1), which is used to produce Table 7, is easily obtained, since

$$\frac{1}{245}\sum_{j=1}^{50}[(v_j - v_E)^2 - \frac{1}{50}\sum_{i=1}^{50}(v_i - v_E)^2]^2$$

estimates the variance of (3.1).

For the Reweighting method an exact value for v_E cannot be obtained and, therefore,

$$\frac{1}{9}\sum_{i=1}^{10}(v_i - \frac{1}{10}\sum_{j=1}^{10} v_j)^2 \qquad (3.2)$$

was used to estimate the variances of the variance estimates. We were unable to compute a standard error for (3.2).

For the Parent Sample Weighting and Base Weights methods there appears, as expected, to be a general downward trend in the variances of the variance estimates with increasing k, although they remain relatively high even for 48 replicates.

Somewhat surprisingly, the data in Table 6 does not appear to support the generally held belief that the variances of the variance estimates are higher for random replication than partial balancing. We have no complete explanation for this. A possible partial explanation is that although many of the cross-product terms drop out when partial balancing is used, those that remain appear with the same sign in each of the k replicates. The large variability in these estimates arising from the small number of repetitions, as reflected in the large standard errors listed in Table 7, is a second possible partial explanation.

For the Reweighting method, the small number of repetitions of the variance estimates again make it difficult to draw any conclusions about the effects on the variances of the variance estimates of the number of repetitions or the method of obtaining the set of replicates.

REFERENCES

Bean, Judy A. (1975), "Distribution and Properties of Variance Estimators for Complex Multistage Probability Sample," Vital and Health Statistics, Series 2, No. 65, National Center for Health Statistcs, Public Health Service, Washington, D.C.

Borack, Jules I. (1971), "A General Theory of Balanced 1/N Sampling," unpublished Ph.D. dissertation, Cornell University, Ithaca, N.Y.

Dippo, C.S., Fay, R.E., and Morganstein, D.H. (1984), "Computing Variances from Complex Samples with Replicate Weights," Proceedings of the Section on Survey Research Methods, American Statistical Association, 489-494.

Ernst, L.R., Huggins, V.J., and Grill, D.E., (1986), "Two New Variance Estimation Techniques," Proceedings of the Section on Survey Research Methods, American Statistical Association, 400-405.

Frankel, Martin R. (1971), Inference From Survey Samples, Ann Arbor: Institute of Social Research, University of Michigan.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953), Sample Survey Methods and Theory, 2 Volumes, New York: John Wiley and Sons.

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

Jones, Charles D. (1984), "1980 CPS Redesign: Specifications for the Second-Stage Ratio Adjustment," memorandum to Thomas C. Walsh, Bureau of the Census.

Judkins, David T. (1987), "Modified Balanced Repeated Replications," Proceedings of the Section on Survey Research Methods, American Statistical Association, to appear.

Krewski, D. and Rao, J.N.K. (1981), "Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods," Annals of Statistics, 9, 1010-1019.

Lemeshow, Stanley (1979), "The Use of Unique Statistical Weights for Estimating Variances With the Balanced Half-Sample Technique," Journal of Planning and Inference, 3, 315-323.

Raj, Des (1968), Sampling Theory, New York: McGraw-Hill.

Rust, Keith F. (1984), "Techniques for Estimating Variances for Sample Surveys," unpublished Ph.D. dissertation, University of Michigan, Ann Arbor, Michigan.

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

Table 1. Combinations for Which Variance Estimates Computed for Reweighting Method (Indicated by "X")

	$A_{gh} = 1$	$A_{gh} = p_{gh}$			
	k = 12	k = 12		k = 24	
d	Partial Balancing	Partial Balancing	Random Replication	Partial Balancing	Random Replication
1	X				
4	X	X	X	X	X
100		X			
10,000		X			

Table 2. Variance Estimates ($x10^9$) for Each A_{gh} and Weighting Method Combination

Characteristic	$A_{gh} = 1$				$A_{gh} = P_{gh}$			
	Base Weights	Parent Sample Weights	Reweighting	Standard Error of Variance Estimates for Reweighting	Base Weights	Parent Sample Weights	Reweighting	Standard Error of Variance Estimates for Reweighting
Labor Force, Total	218.318	248.650	59.908	10.448	146.025	164.369	55.529	2.964
Black	19.581	24.269	6.736	0.843	21.351	24.375	6.625	0.452
Teenager (16-19)	10.463	12.351	4.724	0.676	8.436	10.103	4.793	0.321
Employed, Total	199.537	228.686	68.084	9.748	131.916	151.066	59.678	3.990
Black	14.873	20.002	9.016	1.082	16.089	19.734	7.722	0.521
Teenager (16-19)	8.517	10.131	5.295	0.633	7.173	8.609	5.013	0.374
Agriculture	7.653	9.273	7.912	0.821	4.882	6.106	6.519	0.453
Manufacturing wage & salary	51.363	61.946	45.912	7.309	42.127	50.418	40.405	2.604
Unemployed, Total	13.483	16.955	13.550	1.531	12.522	16.018	13.587	0.761
Black	3.234	4.200	3.406	0.503	3.134	4.220	2.967	0.154
Teenager (16-19)	2.283	2.864	2.355	0.291	2.015	2.592	2.133	0.170
15 weeks or more	3.749	5.068	3.701	0.275	3.019	4.261	3.548	0.236

Table 3. Variance Estimates ($x10^9$) with Reweighting for 1,2,3 and 6 Raking Cycles

Characteristic	A_{gh} = 1				A_{gh} = P_{gh}			
	Number of Cycles				Number of Cycles			
	1	2	3	6	1	2	3	6
Labor Force, Total	59.772	59.849	59.884	59.908	55.767	55.560	55.560	55.529
Black	6.691	6.736	6.737	6.736	6.640	6.629	6.626	6.625
Teenager (16-19)	4.763	4.738	4.731	4.724	4.812	4.789	4.789	4.792
Employed, Total	67.698	68.017	68.066	68.084	60.161	59.754	59.721	59.678
Black	8.970	9.015	9.017	9.016	7.776	7.729	7.723	7.722
Teenager (16-19)	5.325	5.284	5.284	5.295	5.014	5.001	5.004	5.013
Agriculture	7.805	7.904	7.911	7.912	6.401	6.509	6.517	6.519
Manufacturing wage & salary	45.441	45.833	45.895	45.912	39.890	40.322	40.394	40.405
Unemployed, Total	13.444	13.543	13.551	13.550	13.499	13.576	13.585	13.587
Black	3.426	3.409	3.406	3.406	2.985	2.969	2.967	2.967
Teenager (16-19)	2.340	2.358	2.359	2.355	2.119	2.132	2.133	2.133
15 weeks or more	3.668	3.700	3.701	3.701	3.524	3.546	3.548	3.548

Table 4. Variance Estimates ($x10^9$) for Each A_{gh} and Weighting Combination with the Sets of Collapsed Strata Reversed

Characteristic	$A_{gh} = 1$				$A_{gh} = P_{gh}$			
	Base Weights	Parent Sample Weights	Reweighting	Standard Error of Variance Estimates for Reweighting	Base Weights	Parent Sample Weights	Reweighting	Standard Error of Variance Estimates for Reweighting
Labor Force, Total	709.884	812.021	58.619	8.287	181.785	207.402	54.968	10.026
Black	26.982	31.866	7.573	1.306	20.656	25.311	6.010	0.863
Teenager (16-19)	12.729	14.894	5.879	0.849	10.211	12.110	4.616	0.570
Employed, Total	601.316	689.753	72.457	8.465	166.386	191.169	60.984	9.716
Black	19.460	24.549	7.755	1.114	15.497	20.511	7.983	0.838
Teenager (16-19)	9.957	11.787	5.741	0.818	8.329	9.949	5.213	0.579
Agriculture	11.130	13.350	13.979	1.650	7.794	9.473	5.226	0.917
Manufacturing wage & salary	74.106	88.725	52.243	5.652	48.328	58.801	42.227	7.857
Unemployed, Total	18.391	22.690	16.837	2.367	13.526	17.026	12.750	1.668
Black	3.469	4.502	3.124	0.343	3.336	4.342	3.337	0.495
Teenager (16-19)	2.447	3.048	2.295	0.344	2.267	2.847	2.309	0.225
15 weeks or more	3.944	5.282	4.518	0.543	3.814	5.148	3.432	0.363

Table 5. Variance Estimates ($\times 10^9$) for Each Reweighting Method Using A_{gh}, d Combinations

Characteristic	$A_{gh} = 1$		$A_{gh} = p_{gh}$		
	d=1	d=4	d=4	d=100	d=10,000
Labor Force, Total	64.462	59.908	49.080	46.994	46.698
Black	7.297	6.736	6.355	6.198	6.198
Teenager (16-19)	4.942	4.724	4.970	4.723	4.685
Employed, Total	74.665	68.804	58.636	55.795	55.381
Black	10.131	9.016	6.983	6.768	6.758
Teenager (16-19)	5.659	5.295	4.833	4.621	4.590
Agriculture	8.953	7.912	6.521	6.136	6.080
Manufacturing wage & salary	49.945	45.912	40.834	40.699	40.793
Unemployed, Total	14.623	13.550	13.908	13.556	13.531
Black	3.718	3.406	3.054	2.993	2.994
Teenager (16-19)	2.447	2.355	2.176	2.189	2.196
15 weeks or more	3.921	3.701	4.022	3.979	3.980

Table 6a. Estimates of the Variance of the Variance Estimates ($x10^{18}$) for A_{gh} = 1 Using Base Weights

Characteristic	Partial Balancing			Random Replication		
	k=12	k=24	k=48	k=12	k=24	k=48
Labor Force, Total	7471.025	3627.144	1856.658	5555.133	2710.554	1819.147
Black	53.506	27.297	12.325	66.455	26.009	18.471
Teenager (16-19)	14.453	12.477	5.993	17.802	10.869	7.493
Employed, Total	5003.361	2812.225	1334.405	4520.554	2068.048	1548.066
Black	42.422	17.486	6.875	34.344	13.824	9.749
Teenager (16-19)	13.150	8.252	3.500	11.627	6.696	3.601
Agriculture	8.888	4.299	2.284	7.336	5.222	2.598
Manufacturing wage & salary	295.233	268.440	109.023	372.482	192.229	120.917
Unemployed, Total	56.892	12.751	6.008	22.394	17.582	7.962
Black	1.161	1.036	0.510	1.728	0.767	0.446
Teenager (16-19)	0.902	0.496	0.157	0.717	0.374	0.201
15 weeks or more	3.142	1.077	0.842	2.733	1.228	0.625

Table 6b. Estimates of the Variance of the Variance Estimates ($x10^{18}$) for A_{gh} = 1 Using Parent Sample Weights

Characteristic	Partial Balancing			Random Replication		
	k=12	k=24	k=48	k=12	k=24	k=48
Labor Force, Total	10672.751	5494.403	2583.725	7179.971	3996.465	2477.443
Black	100.942	40.344	20.012	78.701	38.589	27.407
Teenager (16-19)	25.369	16.881	7.858	24.109	14.301	10.166
Employed, Total	7659.932	4169.440	1740.086	6002.506	3131.297	2039.908
Black	75.100	35.449	12.697	47.792	22.800	15.348
Teenager (16-19)	18.931	11.850	5.093	17.957	10.389	5.128
Agriculture	10.341	5.516	3.300	10.574	7.685	3.752
Manufacturing wage & salary	636.386	398.568	152.995	590.619	289.720	174.500
Unemployed, Total	96.779	21.630	11.283	33.946	28.495	13.084
Black	2.770	1.638	0.829	3.058	1.504	0.753
Teenager (16-19)	1.558	0.754	0.261	1.049	0.626	0.330
15 weeks or more	6.289	2.169	1.592	5.432	2.223	1.135

Table 6c. Estimates of the Variance of the Variance Estimates ($\times 10^{18}$) for $A_{gh} = p_{gh}$ Using Base Weights

Characteristic	Partial Balancing			Random Replication		
	k=12	k=24	k=48	k=12	k=24	k=48
Labor Force, Total	4269.654	1778.145	923.486	4151.389	1674.688	971.470
Black	57.129	30.739	16.842	59.666	36.827	25.753
Teenager (16-19)	13.687	4.972	3.949	12.305	5.275	3.358
Employed, Total	3590.862	1484.842	676.747	3302.816	1147.245	685.054
Black	36.810	21.090	9.926	38.362	23.895	14.308
Teenager (16-19)	7.981	2.489	2.508	10.855	4.119	2.278
Agriculture	3.484	1.774	1.284	4.722	2.172	0.935
Manufacturing wage & salary	290.585	118.152	85.585	226.603	134.294	75.205
Unemployed, Total	23.637	13.964	5.780	27.028	15.240	6.999
Black	1.525	0.749	0.406	1.309	1.422	0.497
Teenager (16-19)	0.424	0.255	0.102	0.592	0.411	0.161
15 weeks or more	1.615	0.823	0.414	1.578	0.684	0.396

Table 6d. Estimates of the Variance of the Variance Estimates ($\times 10^{18}$) for $A_{gh} = p_{gh}$ Using Parent Sample Weights

Characteristic	Partial Balancing			Random Replication		
	k=12	k=24	k=48	k=12	k=24	k=48
Labor Force, Total	5161.178	2093.203	1130.569	5615.510	1890.675	1182.437
Black	68.522	33.064	22.282	108.463	53.050	39.162
Teenager (16-19)	21.922	7.614	6.024	17.024	7.153	4.657
Employed, Total	4457.248	1811.693	876.709	4419.209	1372.310	817.886
Black	49.927	25.589	15.491	75.681	36.957	22.854
Teenager (16-19)	12.972	4.327	4.124	15.114	5.919	3.355
Agriculture	6.243	3.123	2.196	7.682	3.400	1.532
Manufacturing wage & salary	395.928	182.533	123.821	328.039	212.700	110.095
Unemployed, Total	35.898	21.495	9.270	46.939	22.037	12.475
Black	2.587	1.295	0.725	2.972	2.723	1.037
Teenager (16-19)	0.695	0.434	0.199	0.836	0.695	0.277
15 weeks or more	3.209	1.683	0.848	2.916	1.356	0.879

Table 6e. Estimates of the Variance of the Variance Estimates ($x10^{18}$) for $A_{gh} = p_{gh}$ with Reweighting

Characteristic	Partial Balancing		Random Replication	
	k=12	k=24	k=12	k=24
Labor Force, Total	285.854	81.636	237.451	269.894
Black	12.667	4.620	5.025	3.543
Teenager (16-19)	6.080	2.307	2.273	1.824
Employed, Total	358.666	191.104	341.353	445.681
Black	8.563	4.656	11.052	6.200
Teenager (16-19)	4.254	2.088	4.364	3.494
Agriculture	7.873	1.867	2.254	6.328
Manufacturing wage & salary	109.193	110.561	180.089	160.743
Unemployed, Total	26.627	13.395	45.285	9.768
Black	0.778	0.279	1.657	0.672
Teenager (16-19)	1.141	0.577	0.613	0.583
15 weeks or more	2.210	1.569	0.522	0.659

Table 7a. Estimated Standard Errors ($x10^{18}$) for Estimates in Table 6a

Characteristic	Partial Balancing			Random Replication		
	k=12	k=24	k=48	k=12	k=24	k=48
Labor Force, Total	2228.126	958.031	248.086	1514.590	518.107	326.923
Black	11.909	4.784	3.487	19.973	4.729	2.921
Teenager (16-19)	2.472	2.542	1.043	3.760	2.466	2.062
Employed, Total	1108.955	521.847	175.911	1047.272	358.620	285.750
Black	7.778	3.025	1.341	11.605	2.476	1.819
Teenager (16-19)	2.050	1.754	0.689	2.146	1.533	0.679
Agriculture	1.500	0.758	0.498	1.961	0.976	0.520
Manufacturing wage & salary	41.411	73.941	30.511	64.263	31.823	23.207
Unemployed, Total	14.350	2.656	0.911	4.395	3.641	1.328
Black	0.206	0.182	0.096	0.340	0.141	0.089
Teenager (16-19)	0.180	0.102	0.028	0.127	0.064	0.037
15 weeks or more	0.731	0.177	0.206	0.600	0.284	0.158

Table 7b. Estimated Standard Errors ($x10^{18}$) for Estimates in Table 6b.

Characteristic	Partial Balancing			Random Replication		
	k=12	k=24	k=48	k=12	k=24	k=48
Labor Force, Total	3565.401	1557.826	371.902	1528.628	885.593	461.606
Black	17.818	6.395	5.532	25.267	6.518	4.459
Teenager (16-19)	4.636	3.773	1.350	5.239	3.249	3.020
Employed, Total	1923.707	880.727	243.736	1124.663	594.024	351.710
Black	11.601	7.542	2.579	18.427	4.442	2.822
Teenager (16-19)	3.796	2.597	1.028	3.985	2.480	1.000
Agriculture	1.951	0.922	0.737	2.805	1.613	0.865
Manufacturing wage & salary	214.960	108.267	39.830	122.191	46.648	32.702
Unemployed, Total	21.273	4.693	1.686	5.973	6.069	2.201
Black	0.665	0.272	0.144	0.585	0.250	0.141
Teenager (16-19)	0.371	0.165	0.056	0.185	0.102	0.059
15 weeks or more	1.909	0.399	0.347	1.350	0.528	0.260

Table 7c. Estimated Standard Errors ($x10^{18}$) for Estimates in Table 6c.

Characteristic	Partial Balancing			Random Replication		
	k=12	k=24	k=48	k=12	k=24	k=48
Labor Force, Total	844.524	493.522	230.856	640.481	298.012	193.670
Black	11.088	7.066	2.908	11.584	8.446	5.509
Teenager (16-19)	3.043	0.993	0.819	3.195	1.472	1.149
Employed, Total	660.004	346.681	140.232	771.021	315.065	130.485
Black	10.115	5.049	1.753	8.658	5.866	2.574
Teenager (16-19)	1.631	0.373	0.525	2.690	0.895	0.597
Agriculture	1.018	0.412	0.273	0.779	0.349	0.188
Manufacturing wage & salary	87.962	21.283	13.869	48.815	19.938	11.775
Unemployed, Total	4.754	3.400	1.014	5.094	2.718	1.157
Black	0.226	0.137	0.062	0.318	0.502	0.115
Teenager (16-19)	0.067	0.048	0.020	0.123	0.081	0.027
15 weeks or more	0.353	0.234	0.095	0.329	0.119	0.071

Table 7d. Estimated Standard Errors ($x10^{18}$) for Estimates in Table 6d.

Characteristic	Partial Balancing			Random Replication		
	k=12	k=24	k=48	k=12	k=24	k=48
Labor Force, Total	1138.143	623.606	274.634	787.062	285.299	257.453
Black	11.650	6.109	3.838	20.456	10.815	8.388
Teenager (16-19)	5.358	1.549	1.267	4.510	2.060	1.558
Employed, Total	872.559	413.730	183.047	919.058	334.060	160.058
Black	12.171	5.827	2.504	22.306	7.759	3.838
Teenager (16-19)	2.677	0.752	0.890	3.522	1.303	0.872
Agriculture	1.934	0.690	0.497	1.280	0.551	0.310
Manufacturing wage & salary	118.518	31.296	20.266	78.841	34.172	16.434
Unemployed, Total	7.785	5.882	2.020	8.731	3.460	1.997
Black	0.370	0.206	0.125	0.863	0.995	0.261
Teenager (16-19)	0.113	0.084	0.039	0.185	0.139	0.043
15 weeks or more	0.657	0.466	0.202	0.561	0.291	0.146