BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-88/13

SOME COMMENTS ON SCHIRM AND PRESTON

by

*Michael Lee Cohen
University of Maryland
and Bureau of the Census
ASA/Census Program
Washington, D.C.

and

**Xiao Di Zhang
Bureau of the Census
ASA/Census Program
Washington, D.C. 20233

*ASA/Census Fellow
**ASA/Census Associate

Recommended by: Kirk M. Wolter

Report completed: March, 1988

Report issued: March 8, 1988

Some Comments on Schirm and Preston

by

Michael Lee Cohen
University of Maryland and Bureau of the Census

and

Xiao Di Zhang
Bureau of the Census

Abstract: Schirm and Preston (1987) have shown in a Monte Carlo study that the decision on whether to adjust the census for undercoverage can depend strongly on the loss function used to make the decision. The purpose of this comment is to explore the dependence of their findings on various features of their simulation study. Specifically, when one uses three demographic groups rather than two, or one uses different loss functions, the dependence of the decision to adjust on the choice of the loss function is considerably reduced. Furthermore, the degree of agreement of various loss functions on whether to adjust is empirically shown to be a smooth function of the correlation between the adjustment factor used and the true adjustment factor.

1. Introduction

The census is used for a variety of applications. Each application has an implicit loss function measuring the deleterious effects of differences between the true counts and the counts used. If it were the case that the relative superiority or inferiority of adjusted counts compared to census counts depended on the application for which the counts were intended, then the benefit of adjustment would be more difficult to argue. On the other hand, if superiority of adjusted counts for a given loss function generally implied superiority for other loss functions (at least for important applications) then adjustment would be more supportable. The dependence of adjustment on the choice of loss functions has been investigated by the authors for two particular loss functions in Cohen and Zhang (1988).

Schirm and Preston (1987), hereafter referred to as S-P, examined the dependence of the superiority of adjusted counts on the choice of loss function (as well as other adjustment issues). To accomplish this they designed a Monte Carlo study using the 1980 U.S. Census counts for the 50 states and the District of Columbia (hereafter all referred to as "states"), along with information from demographic analysis of the 1980 census (U.S. Bureau of the Census, 1982) and some other selected studies by the Bureau of the Census (see U.S. Bureau of the Census, 1977). They made use of the following notation:

N_{ji}^{C} = observed population count for demographic group j in state i

N_{ji}^{T} = true population count for demographic group j in state i

u_{ji} = undercount rate for demographic group j in state i

$$= \exp\{\mu_j + \sigma_j V_{ji}\}$$

for μ_j and σ_j discussed below.

In addition $N^T_{ji} = N^C_{ji}\, u_{ji}$. The V_{ji} are independent standard normal random variables. Finally, the index j, when set equal to one in S-P, represented the black population, and when set equal to two represented the white and other population.

In their paper S-P examined four Scenarios. Here we examine only Scenario I where the adjustment factors are assumed equal to $N^T_{j\cdot}/N^C_{j\cdot}$ and the adjusted counts are equal to:

$$N^A_{ji} = (N^T_{j\cdot}/N^C_{j\cdot})\, N^C_{ji} \quad .$$

Scenarios II, III, and IV investigated the addition of systematic bias and random variation to the adjustment factors.

The parameters μ_j and σ_j of the undercount rates were chosen so that the expectation of the u_{ji} matched the values obtained from demographic analysis (using the assumption that six million undocumented aliens were counted in the 1980 census), and also so that either 50, 75, or 95% of the distribution of the u_{ji} was represented by specified intervals. Since we can have 50, 75 or 95% coverage for u_{1i} and u_{2i}, nine situations result. The values of μ_j and σ_j for these nine situations are presented in Table 1.

Table 1. Parameters Defining Cases of Scenario I

	μ_1	σ_1^2	μ_2	σ_2^2
Case 11	.0440	.0058	-.0114	.0007
Case 12	.0440	.0058	-.0112	.0003
Case 13	.0440	.0058	-.0111	.0001
Case 21	.0460	.0018	-.0114	.0007
Case 22	.0460	.0018	-.0112	.0003
Case 23	.0460	.0018	-.0111	.0001
Case 31	.0466	.0006	-.0114	.0007
Case 32	.0466	.0006	-.0112	.0003
Case 33	.0466	.0006	-.0111	.0001

In the following, we will not present the results for each of the nine cases, instead we present the average over the nine cases. Generally, the pattern for the individual cases will mimic that for the means.

S-P defined P_i^K as $N_{\cdot i}^K / N_{\cdot\cdot}^K$. Replacing K by T, C, or A results in P_i^K representing the share of the total population in state i using true counts, census counts, or adjusted counts, respectively. For each repetition of their Monte Carlo study the following statistics were calculated:

$$\text{RPSAE} = \sum | P_i^C - P_i^T | / \sum | P_i^A - P_i^T |$$

$$\text{RPSSE} = \sum (P_i^C - P_i^T)^2 / \sum (P_i^A - P_i^T)^2$$

Thus, if RPSAE is less than one the census counts are preferred to the adjusted counts for absolute error, and if RPSSE is less than one the census counts are preferred to the adjusted counts for square error. S-P computed the following four aggregate statistics: 1) PADJSAE - the proportion of repetitions for which RPSAE is greater than one, 2) PADJSSE - the proportion of repetitions for which RPSSE is greater than one, 3) PADJBOTH - the proportion of repetitions for which both RPSAE and RPSSE are greater than one (where adjustment is indicated for both loss functions), and 4) PAGREE - the proportion of repetitions for which RPSAE and RPSSE are simultaneously less than or greater than one.

S-P calculated 500 repetitions for each of the nine situations of Scenario I. So that we could achieve complete comparability with later computations we repeated the calculations using 1000 repetitions for each of the nine cases. (In no case did any of the four statistics listed above for any of the nine cases differ from the values given by S-P by more than .06). The means over the nine situations for these four statistics are:

PADJSAE	PADJSSE	PAGREE	PADJBOTH
.79	.67	.77	.62

2. Change of Loss Functions

In Scenario I the adjustment factors used have no bias or random variation, and so are in some sense as good as one could expect within the context of the simple estimator investigated. Also, the simulation assumes undercounts about the size that are believed to have obtained in 1980. Finally, the two underlying loss functions, LABS = $\sum | P_i^K - P_i^T |$ and LSQUARE = $\sum (P_i^K - P_i^T)^2$ are fairly closely related (this will be discussed more fully below). Thus the main results of S-P that only 62% of the repetitions found both criteria recommending adjustment and only 77% of the repetitions found both criteria even agreeing were surprisingly pessimistic towards the prospect of adjustment. It is therefore of interest to determine how sensitive the results presented above are to the circumstances of the simulation study. We focus on two factors: (1) the loss functions used, and (2) the use of three rather than two demographic strata. We make use of the following definitions.

Let

$$\text{LOSS1} = \sum (N^K_{\cdot i} - N^T_{\cdot i})^2/N^T_{\cdot i} \quad \text{and}$$

let

$$\text{LOSS2} = \sum (N^K_{\cdot i}/N^T_{\cdot i} - N^K_{\cdot\cdot}/N^T_{\cdot\cdot})^2 \; N^T_{\cdot i}$$

be two new loss functions. LOSS1 was put forward as a reasonable loss function in the adjustment context by Fellegi (1980) and Tukey (1983). LOSS2 was put forward as a reasonable loss function in the adjustment context by Tukey (1983) among others. Analogous to the definitions of RPSAE and RPSSE, we have:

$$\text{RLOSS1} = \sum (N^C_{\cdot i} - N^T_{\cdot i})^2/N^T_{\cdot i} \;/\; \sum (N^A_{\cdot i} - N^T_{\cdot i})^2/N^T_{\cdot i}$$

$$\text{RLOSS2} = \sum (N^C_{\cdot i}/N^T_{\cdot i} - N^C_{\cdot\cdot}/N^T_{\cdot\cdot})^2 \; N^T_{\cdot i} \;/\; \sum (N^A_{\cdot i}/N^T_{\cdot i} - N^A_{\cdot\cdot}/N^T_{\cdot\cdot})^2 \; N^T_{\cdot i}$$

As in S-P, let PADJLOSS1 be the percentage of repetitions where RLOSS1 is greater than one, and let PADJLOSS2 be the percentage of repetitions where RLOSS2 is greater than one. Finally, let PAGREE(1) be the percentage of repetitions where RLOSS1 and RLOSS2 are either both greater than or less than one, and let PADJBOTH(1) be the percentage of repetitions where RLOSS1 and RLOSS2 are both greater than one. One thousand repetitions were computed for each of the nine cases of Scenario I. The means of the results over all nine cases are given below.

PADJLOSS1	PADJLOSS2	PAGREE(1)	PADJBOTH(1)
.95	.87	.92	.87

Since LOSS1 is a criterion closely allied with the simple adjustment methodology examined by S-P it is not surprising that the mean for PADJLOSS1 is high (95%). We also find that the mean for PAGREE(1) is 92% and the mean for PADJBOTH(1) is 87%. These values are certainly much more encouraging for adjustment than the corresponding values of 77% and 62% found above.

Therefore, it seems that the results of S-P are strongly affected by the choice of loss function. We will argue below that one of the loss functions used by S-P, namely LSQUARE, weights small areas too highly, a property which is not shared by LABS, LOSS1, or LOSS2. As a result, there is a justification for preferring the results found here.

3. Increasing the Number of Demographic Strata

Next we examined the dependence of the results of S-P on the choice of using only two demographic strata in the simulation. To do this we needed to expand the parameter set of Table 1 to include an undercount factor for the third demographic stratum, Hispanics. In Table 2 we have presented the parameters we used.

Table 2. Parameters Used in Three Demographic Strata Simulation

	Black		Hispanic		White & Other	
	μ_1	σ_1^2	μ_2	σ_2^2	μ_3	σ_3^2
Case 11	.0440	.0058	.0330	.0086	-.0114	.0007
Case 12	.0440	.0058	.0330	.0086	-.0112	.0003
Case 13	.0440	.0058	.0330	.0086	-.0111	.0001
Case 21	.0460	.0018	.0358	.0030	-.0114	.0007
Case 22	.0460	.0018	.0358	.0030	-.0112	.0003
Case 23	.0460	.0018	.0358	.0030	-.0111	.0001
Case 31	.0466	.0006	.0368	.0010	-.0114	.0007
Case 32	.0466	.0006	.0368	.0010	-.0012	.0003
Case 33	.0466	.0006	.0368	.0010	-.0111	.0001

Note that the situations of Table 2 are not completely crossed, which would have resulted in 27 cases. Instead, the parameters for the Hispanic undercount parallel the parameters for the black undercount. Although demographic analysis in 1980 yielded only very imprecise information about the Hispanic undercount we have made use of the available research (see Cowan and Bettin, 1982) and correspondingly assumed that the Hispanic undercount was roughly similar to that for blacks, i.e., that the expected undercounts were similar. In addition we have assumed that the Hispanic undercount had a larger variability than the black undercount. Table 3 gives the 50, 75, and 95% coverage intervals implied by our choices of μ_j and σ_j^2 .

Table 3. Intervals of Specified Probability Implied by Parameters of Table 2

μ_1	σ_1^2	Undercount Range	Coverage %	$Exp(u_{ji})$	$Var(u_{ji})$
.0440	.0058	(-.74, 10.01)	50	1.048	.0064
.0460	.0018	(-.28, 9.94)	75	1.048	.0020
.0466	.0006	(-.14, 9.92)	95	1.048	.0007
μ_2	σ_2^2	Undercount Range	Coverage %	$Exp(u_{ji})$	$Var(u_{ji})$
.0330	.0086	(-2.92, 10.03)	50	1.038	.0093
.0358	.0030	(-2.68, 10.38)	75	1.038	.0032
.0368	.0010	(-2.49, 10.38)	95	1.038	.0011
μ_3	σ_3^2	Undercount Range	Coverage %	$Exp(u_{ji})$	$Var(u_{ji})$
-.0114	.0007	(-2.88, 0.65)	50	0.989	.0007
-.0112	.0003	(-3.06, 0.88)	75	0.989	.0003
-.0111	.0001	(-3.02, 0.85)	95	0.989	.0001

Using three rather than two demographic strata, we otherwise repeated the analysis of Scenario I of S-P. Let us define PADJSAE and PADJSSE as in S-P except they now refer to a situation where there are three demographic strata, and to distinguish we designate PAGREE(2) and PADJBOTH(2) as the analogous quantities to PAGREE and PADJBOTH for the original S-P simulation. The results averaged across the nine cases of Scenario I are:

PADJSAE	PADJSSE	PAGREE(2)	PADJBOTH(2)
.81	.75	.84	.70

The results of 84% for PAGREE(2) and 70% for PADJBOTH(2) are both approximately 7% higher than the corresponding quantities for S-P. This difference has a p-value of less than 1%. Since the proposed adjustment routine used by the Bureau of the Census is likely to stratify the population demographically into more than three subpopulations, it would seem that these higher values are more appropriate than those of S-P. It is reasonable to hypothesize that this upward trend will continue as more important stratification is done, an issue that is further addressed in Section 5.

4. The Effect of Both Changes Simultaneously

A natural question to ask is what effect both changes, the change of loss functions and the addition of another demographic stratum, would have jointly on the statistics PAGREE and PADJBOTH. Using the statistics PADJLOSS1 and PADJLOSS2 and the parameters of Table 2 we computed 1,000 repetitions for each of the nine situations of Scenario I. We designate PAGREE(3) and PADJBOTH(3) as the analogous quantities to PAGREE and PADJBOTH for the original S-P simulation. The averages across the nine cases of Scenario I are:

PADJLOSS1	PADJLOSS2	PAGREE(3)	PADJBOTH(3)
.93	.87	.94	.87

The results of 94% for PAGREE(3) and 87% for PADJBOTH(3) are not significantly different from the values obtained from the change of loss functions alone. Thus, there appears to be little joint impact of the two changes.

5. Dependence of Agreement on Quality of Adjustment

The dependence of Schirm and Preston's results on the number of demographic groups raises the possibility that the agreement of decisions is strongly dependent on the quality of the adjustment. This is clearly true in the limit, as the adjusted counts closely approach the true counts, since then all reasonable loss functions should have lower loss for the adjusted counts than for the census counts. To examine this dependence of the agreement of the decision to adjust on the quality of the adjustment we performed the following simulation.

Let $\overline{P}$ be the average of $P_i = 1 - N^C_{\cdot i}/N^T_{\cdot i}$ over 10,000 repetitions over the 51 states (51 x 10,000 values). Let the standard deviation of these 510,000 values be denoted SD_P. Then let:

$$Q_i = SD_P\ (\sqrt{12})\ t_i + \overline{P}$$

where t_i is uniformly distributed on [-1/2, 1/2]. Finally let:

$$r_i^\alpha = \alpha P_i + (1 - \alpha)\ Q_i$$

Then, as $\alpha \to 0$, r_i^α becomes purely random with the same first and second moments as the rates of undercoverage, P_i, over all states. As $\alpha \to 1$, r_i^α becomes the true undercoverage rate for the i-th state. Given r_i^α we define:

$$N^A_{\cdot i}(\alpha) = N^C_{\cdot i}\ (1/(1 - r_i^\alpha))$$

When α is equal to one, $N^{A}_{\cdot i}(\alpha) = N^{T}_{\cdot i}$. It is easy to show that the correlation between P_i and $r^{\alpha}_{i} = \alpha / \sqrt{2\alpha^2 - 2\alpha + 1}$.

In a new simulation experiment, $N^{C}_{\cdot i}$ and $N^{T}_{\cdot i}$ were calculated precisely as in S-P and $N^{A}_{\cdot i}(\alpha)$ was calculated as given above. Using these, RPSSE and RPSAE were computed. This was done 1,000 times for each choice of α, allowing us to compute versions of PADJSAE, PADJSSE, and PADJBOTH(4). The results for PADJBOTH(4) are presented in Table 4. In addition, using RLOSS1 and RLOSS2, we computed versions of PADJLOSS1 and PADJLOSS2, as well as PADJBOTH(5). The results for PADJBOTH(5) are also presented in Table 4.

Table 4. Dependence of Agreement of Adjustment Decisions on Quality of Adjustment

α	Corr(P_i, r^{α}_{i})	PADJBOTH(4)	PADJBOTH(5)
.00	.00	.01	.00
.09	.10	.03	.02
.17	.20	.09	.05
.24	.30	.17	.17
.30	.40	.32	.38
.37	.50	.52	.66
.40	.55	.61	
.43	.60	.71	.89
.46	.65	.81	
.49	.70	.90	.98
.53	.75	.95	
.57	.80	.99	1.00
.62	.85	.99	
.67	.90	1.00	1.00
.75	.95	1.00	
.88	.99	1.00	1.00

Examining Table 4 we see that for the loss functions of S-P, correlations between the adjustment factors used and the true adjustment factors of .70 and higher yield values of PADJBOTH(4) of 90% and higher. Similarly, using the two loss functions RLOSS1 and RLOSS2, correlations of .60 and higher result in values for PADJBOTH(5) of 89% and higher. Therefore, if the adjustment is effective, but not necessarily exceptionally effective, loss functions with fairly distinct properties can all be expected to recommend adjustment with high probability. We point out that increased use of important stratification would cause the correlation of the adjustment factors to the true factors to increase. Thus the results of section 3 and the hypothesis following the results are consistent with this finding.

6. Weighting of Loss Functions

Many (but certainly not all) of the loss functions in the area of adjustment can be categorized by three characteristics. These are: (1) the key element, (2) the positive, symmetric function used, and (3) the weighting used in the aggregation.

The key element is what is measured for each of the states. This is commonly the estimated population of a state, here $N^K_{.i}$ for K = A, T, or C, or the population share of a state, p^K_i . The positive, symmetric function, typically $f(x) = |x|$ or $f(x) = x^2$, is used to measure the distance between the key element and the true key element resulting when K is replaced by T, which is called the loss. Finally, when we aggregate these losses over states, a weight can be used. Weights are often the true population of the state or the reciprocal of the true population of the state.

It is not immediately obvious what criterion should be used to decide if a non-constant weight should be included in a loss function. One rather compelling criterion used by Tukey (1983) is the invariance of the loss function to equal disaggregation. More precisely, we believe that any reasonable loss function should have the following property: When each state is disaggregated into m identical substates with estimated population $N^K_{.i}/m$ and with true population $N^T_{.i}/m$, the overall loss should not change. It is easy to show that LABS, LOSS1, and LOSS2 have this property, but LSQUARE does not. We demonstrate this for the loss functions LABS and LSQUARE.

$$\text{LABS (with disaggregation)} = \sum_i m|(N^K_{\cdot i}/m)/N^K_{\cdot\cdot} - (N^T_{\cdot i}/m)/N^T_{\cdot\cdot}|$$

$$= \sum_i |N^K_{\cdot i}/N^K_{\cdot\cdot} - N^T_{\cdot i}/N^T_{\cdot\cdot}|$$

$$= \text{LABS (without disaggregation)}$$

$$\text{LSQUARE (with disaggregation)} = \sum_i m[(N^K_{\cdot i}/m)/N^K_{\cdot\cdot} - (N^T_{\cdot i}/m)/N^T_{\cdot\cdot}]^2$$

$$= \sum_i (1/m)\ [(N^K_{\cdot i}/N^K_{\cdot\cdot}) - (N^T_{\cdot i}/N^T_{\cdot\cdot})]^2$$

$$= (1/m)\ \text{LSQUARE (without disaggregation)}$$

This indicates that LSQUARE should be weighted by a factor inversely proportional to either the true or estimated population of the state (or some related quantity), e.g., $1/N^T_{\cdot i}$. The use of weights $1/N^T_{\cdot i}$ produces a loss function not importantly different from LOSS2. Clearly, then LSQUARE gives too much weight to small states. We know that the smaller states are generally those with proportionately smaller minority populations. Therefore these states, with excess weight, are states with generally small undercounts. This may be the explanation of the small value (67%) of PADJSSE in S-P's simulation. We note that PADJLOSS2 is equal to 87%.

7. Conclusions

There are two primary findings in this investigation of the results of Schirm and Preston. First, the choice of loss function can make a significant and substantial impact on whether adjusted counts are preferred to census counts. When two properly weighted loss functions are used the probability that adjusted counts are preferred simultaneously for the two loss functions was 87%, substantially higher than S-P's 62%.

Furthermore, the probability of simultaneous preference of adjusted counts was empirically shown to be a smooth function of the quality of the adjustment. Even for fairly moderate levels of correlation between the adjustment factor and the true factor, the probability of simultaneous preference for adjusted counts was found to be 90% or higher.

References

Cohen, Michael L., and Zhang, Xiao Di (1988), "A Comparison of Absolute and Proportional Loss Functions in Adjustment Using Artificial Populations, manuscript.

Cowan, Charles D. and Bettin, Paul J. (1982), "Estimates and Missing Data Problems in the Post Enumeration Program. Unpublished paper, Statistical Methods Division, Bureau of the Census, Washington DC.

Fellegi, Ivan (1980), "Should the census count be adjusted for allocation purposes--equity considerations." Pp. 193-203 in Proceedings of the 1980 Conference on Census Undercount, Bureau of the Census, Washington, DC: U.S. Department of Commerce.

Schirm, Allen L., and Preston, Samuel H. (1987), "Census Undercount Adjustment and the Quality of Geographic Population Distributions," Journal of the American Statistical Association, 82, 965-977.

Tukey, J. W. (1983). Affidavit Presented to District Court, Southern District of New York, Mario Cuomo, et al., Plaintiff(s), Malcolm Baldrige, et al., Defendants, 80 Civ. 4550 (JES).

U.S. Bureau of the Census (1977), Developmental Estimates of the Coverage of the Population of States in the 1970 Census: Demographic Analysis, Current Population Reports, Special Studies, Ser. P-23, No. 65, Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census (1982), Coverage of the National Population in the 1980 Census by Age, Sex, and Race: Preliminary Estimates by Demographic Analysis, Current Population Reports, Special Studies, Ser. P-23, No. 115, Washington, DC: U.S. Government Printing Office.