REPRESENTING LOCAL AREA ADJUSTMENT BY REWEIGHTING OF HOUSEHOLDS

Alan M. Zaslavsky, Massachusetts Institute of Technology
M.I.T. Statistics Center E40-111, 77 Massachusetts Ave, Cambridge, MA 02139

Abstract

Suppose that undercount rates in a census have been estimated and that block-level estimates of the undercount have been computed. It may then be desirable to create a new roster of households incorporating the estimated omissions. It is proposed here that such a roster be created by weighting the enumerated households. The household weights are constrained by linear equations representing the desired total counts of persons in each estimation class and the desired total count of households. Weights are then calculated that satisfy the constraints while making the fitted table as close as possible to the raw data. The procedure may be regarded as an extension of the standard "raking" methodology to situations where the constraints do not refer to the margins of a contingency table. Continuous as well as discrete covariates may be used in the adjustment, and it is possible to check directly whether the constraints can be satisfied.

1. Household-level adjustment by weighting.

A major research effort has been devoted to methods for estimation of the undercount in the 1990 Census in the United States. (National Academy of Sciences 1985.) In one of the primary methodologies that has been proposed, a Post Enumeration Survey (PES) would be conducted shortly after the Census in a sample of blocks. The fraction of persons in the PES who were omitted from the Census enumeration yields an estimate of Census undercoverage. Estimates of the undercount would be carried down to some geographical level (possibly the smallest geographical unit used by the Census, the block). These estimates would apply to classes formed on the basis of characteristics of persons, as well as possibly some household or block-level characteristics. (The term "class" will be used henceforth to refer to estimation or adjustment classes or cells; the term "block" will refer to the smallest geographical unit for which undercount estimates are calculated. The 1980 Census found approximately one hundred million households in two to four million blocks, depending on the definitions used.)

For each block, the outcome of the processes described above would be a vector of estimated undercounts, with S components corresponding to the adjustment, or estimated number of persons omitted from the census in that block, from each of S adjustment classes. The methods by which these estimates are arrived upon are beyond the scope of this paper. However, in our examples we shall assume that for each class within each block there is an undercount rate, expressing estimated omissions as a fraction of enumerated persons in that class and block (In this paper, the term "adjustment" refers to any process which incorporates the estimated undercount into the enumeration. The adjustment classes might be, but would not necessarily be, the same as the post-strata formed in analysis of a Post-Enumeration Program.) For forming simple marginal tabulations of persons by characteristics, this information might well be adequate. In particular, small-area counts used for various official and commercial purposes could be calculated from block totals.

However, for some purposes it would be desirable to place the added persons in households. We assume for these purposes that there is also an estimate of the number of omissions of whole households on each block. There might also be information distinguishing omissions of persons within enumerated households from those in omitted households.

If the resulting adjusted records are to be meaningful, the composition of the added households and the relationships of its individual members must be logically consistent and typical of the types of households found in that area. (The term "composition" will be used to refer to the number of household members from each adjustment class.) Thus, for example, a household consisting of a 20-year old white female head of household, a 75-year-old Chinese male, and a 10-year-old black daughter would not be a very plausible household, even if all of its members were from classes that are well represented in the block. Yet abstractly to describe these patterns and create new households that fit them is a daunting task.

The essence of the method proposed in this paper is to assign weights to the households enumerated in the census lists for the block, so that the weighted totals of persons in each adjustment class and the weighted total number of households are precisely equal to the corresponding adjusted totals. Thus, although the weighting changes the proportionate composition of the block, all of the households are real and possess characteristics and relationships that are logically consistent and reasonable for that block. (This weighting methodology is similar to the standard raking adjustment, in which the weight applied to counts in a cell of a contingency table is the adjusted count divided by the original count.) The household weights are calculated *after* the block totals have been adjusted and will be consistent with those totals. For most Census purposes, the weighted records would be an adequate basis for forming published tables and sampled lists.

This proposal might be contrasted with imputation methods, in which undercounted units are represented by whole units added to the roster. The imputed units may be either persons or households. Although individual persons may be imputed into the block, the problem of fitting these persons into plausible households remains unsolved. Placing them in fictitious "group quarters," as was done in some tests of adjustment procedures, sidesteps this problem at the cost of creating a skewed picture of relationships in the block.

Another approach to imputation starts with probability models for omissions of households and of persons within households, and draws imputed households from the posterior distribution of the omissions given the enumerated households. This methodology is suited to the multiple imputation approach (Rubin 1987), in which the entire imputation process is repeated several times to represent the variability introduced by the underenumeration. However, in each block roster that is created, totals based on enumerated and imputed households would not necessarily be precisely equal to the desired adjusted totals. In this paper, our concern is with methods that give an exact fit to population estimates derived at a preceding stage.

2. Objectives and mathematical formulation of a weighting plan.

It is an essential goal of the proposed plan that the population of the block be assigned to valid household units, so that statistics for which the unit is the household are unambiguously defined. Thus, weights are assigned to households; the same weights apply to every person within the household.

In order that the counts in the weighted roster be those which are given by the predetermined adjustment, the following constraints must be satisfied:

(A1) Within each block, the sum of household weights equals the adjusted number of households.

(A2) Within each adjustment class and each block, the sum of weights for persons equals the adjusted number of persons.

In order that the weighted block roster be as similar as possible to the original block roster, we further require that:

(B) The weights should be, in some sense, as close to each other as possible.

With unit (or equal) weights, the composition of the block remains unchanged. If the weights are not very unequal, the census composition of the block is nearly preserved by the weighting scheme. To the extent that information about the undercount does not require a drastic revision of our view of the make-up of the block such a drastic revision should be avoided, consistently with good survey practise regarding weights.

We now turn to the mathematical formulation of these criteria. Suppose that in the block under consideration, there are S adjustment classes and I enumerated households, and household i contains C_{is} members from class s. Suppose that H is the desired total number of households in the adjusted roster for the block and D_s is the desired total number of persons in class s. Let W_i, $i=1,2,...I$, be the weights corresponding to the households. (A1) requires that

$$\sum_{i=1}^{I} W_i = H$$

and (A2) requires that

$$\sum_{i=1}^{I} W_i C_{is} = D_s, \quad s=1,2,...S.$$

These constraints can be represented by a matrix equation of the form $AW=B$, where

$$A=\begin{bmatrix} 1 \\ C' \end{bmatrix}, \; B=\begin{bmatrix} H \\ D \end{bmatrix}, W'=[W_1 \; W_2 \ldots W_I]$$

$$\text{and } D'=[D_1 \; D_2 \ldots D_S]$$

and 1 is a row of 1's.

Objective (B) is represented by selecting some objective function that represents the distance between the weights W and uniform weighting, and minimizing it. We will use the objective function $T=\Sigma W_i \log(W_i)$. This measure is proportional to the discriminant information (Kullback-Liebler information) of the discrete probability distribution (over households) with relative weights W_i with respect to the probability distribution with equal weights, and is the same objective function that underlies the traditional "raking" (iterative proportional fitting) procedure for adjusting contingency tables (Deming and Stephan 1940; Ireland and Kullback 1968; Oh and Scheuren 1978 have a larger bibliography). Thus, our procedure may be regarded as an extension of raking.

In the context of raking, initial counts X are given for cells in a contingency table, and new cell counts Y are calculated to minimize the objective function $\Sigma Y_i \log(Y_i/X_i)$. Then the weights of the original observations are the ratios $W_i=Y_i/X_i$. In our context, if X_i households happened to have exactly the same composition we could regard them, in the same way, as forming a single entry in the roster with initial count X_i and fit an adjusted count Y_i. However, with a large number of adjustment classes, it would be unusual for several households in the same block to have exactly the same composition. Thus we will not attempt to group households; rather, it is notationally and computationally simpler to list the households separately so that for each enumerated household composition the initial count $X_l=1$ and $Y_l=W_l$. Aside from this notational difference, the mathematical formulation here differs from that of a raking adjustment only in that the linear constraints do not have the special structure of margins in a contingency table.

Our procedure differs from raking in that the linear constraints do not necessarily refer to margins in a contingency table. Our methodology includes raking as a special case, as well as the raking generalization of Oh and Scheuren (1978) in which different tables are used to fit each margin. In fact, constraints may be imposed on continuous as well as discrete covariates. Furthermore, the algorithms that are set forth allow direct determination of whether there are in fact any weights that are consistent with all of the given constraints. It is possible then to select constraints that must be relaxed in order to fit weights. These features give these methods potential applicability extending beyond the area of representing undercount.

3. Fitting the weights.

The problem now is to determine weights satisfying the constraints $AW=B$, $W \geq 0$, minimizing the objective function $T=\Sigma W_i \log(W_i)$. (To make T a continuous function of W, we adopt the usual convention $0 \log 0 = 0$.)

We will call any weight vector that satisfies the linear constraints (the equations and the inequalities) a feasible solution. As long as there is a constraint on the total weight of the households, the set of feasible solutions is bounded and therefore T assumes a minimum value on it; furthermore, since T is strictly convex, the solution is unique.

The problem of calculating weights then naturally is divided into three tasks: (1) determining whether the linear constraints $AW=B$ are consistent; (2) determining whether there are *any* feasible solutions; and (3) finding the feasible solution minimizing T. We will suppose that there are I households and p constraints, so A is a $p \times I$ matrix.

3.1. Consistency of linear constraints. As long as the rows of A are independent, the constraints $AW=B$ will be consistent. If any row is dependent on the others, the corresponding constraint is either inconsistent or redundant, depending on the values in B. Dependent rows can be identified by applying the Q-R decomposition to A'. If the corresponding constraints are redundant, they may be deleted without any loss of information; if they are inconsistent, the constraints must be reformulated in some way.

3.2. Existence of feasible solutions. Determining the existence of feasible solutions is equivalent to determining an initial feasible solution in a linear programming problem, and the standard algorithms can be used. Suppose our problem is to find a positive solution W to $AW=B$, where $B \geqq 0$. (If the latter condition does not hold it can be made true by reversing the sign of negative elements of B and the corresponding rows in A.) Create an augmented problem $[A \mid I]\,[W' \mid Z']' = B,\ W, Z \geqq 0,$ where I is a $p \times p$ identity matrix and Z is a p-element vector variable. This problem automatically has an initial solution $W=0$, $Z=B$. Then apply the simplex method to minimize ΣZ_i. If that sum can be reduced to 0, the corresponding W values are a solution to the original problem, while if it cannot, the original problem has no solution.

3.3. Optimizing the objective function. By the method of Lagrange multipliers, the minimizing solution must satisfy the equations $\partial T/\partial W_i = \log W_i + 1 = a_i'\lambda$, where a_i is the i-th column of A and $\lambda' = (\lambda_1, \lambda_2, \ldots \lambda_p)$. Then $W_i = exp(a_i'\lambda\text{-}1)$; thus the model for the weights is log-linear in form, like that for a conventional raking adjustment. λ_s represents the additional log-weight increment associated with a unit increment in the corresponding constraint coefficient a_{is}, e.g. adding an additional household member from adjustment class s to the household.

We can solve for λ by Newton's method to satisfy $AW=B$. The iterative scheme we use is

$$\lambda^{(t+1)} = \lambda^{(t)} - (AW^*A')^{-1}(AW-B),$$

where W^* is the matrix with the elements of $W = W(\lambda^{(t)})$ on the diagonal. A good starting value for λ is $\lambda^{(0)} = (AA')^{-1}B$, which can be derived from a linear approximation around equal starting weights. (There is also a cyclic descent procedure for solving these equations, which is a generalization of iterative proportional fitting.)

3.4. An example. The following is a completely worked example of the fitting procedure. We will assume that there are three adjustment cells (men, women, children).

The census roster for nine households (Table 1) is represented by a table showing household compositions (Table 2).

Suppose that we are to add households and persons to the block as shown in Table 3.

Then we must find weights satisfying $AW=B$, that is:

$$\begin{bmatrix} 1&1&1&1&1&1&1&1&1 \\ 1&1&1&0&1&0&1&1&0 \\ 1&2&1&1&0&1&1&2&1 \\ 0&1&2&1&0&1&1&0&2 \end{bmatrix} \mathbf{W} = \begin{bmatrix} 10 \\ 7 \\ 10.5 \\ 9.5 \end{bmatrix}$$

Then the steps of the procedure are as follows:

Step 1: checking consistency. There is no problem, since the rows of A are independent.

Step 2: checking feasibility. There is a feasible (but not optimal) solution using only four households.

Step 3: find the optimal solution. We can solve by Newton's method. The fitted weights are shown under (*W*) in the table of households and the census roster above. The weighted totals of households and of persons by class equal the specified adjusted totals.

Table 1: Raw census data (enumeration)

House-hold #	Name	Sex	Age	(*W*)
1	Martel, John	M	45	*1.042*
1	Martel, Karen	F	43	*1.042*
2	Chen, Shiao-chi	M	27	*1.001*
2	Chen, Betty	F	26	*1.001*
2	Chen, George	M	4	*1.001*
2	Chen, Yu-ling	F	55	*1.001*
3	Chavez, Rosa	M	33	*1.548*
3	Chavez, Miguel	F	34	*1.548*
3	Chavez, Anton	M	3	*1.548*
3	Chavez, Noemi	F	7	*1.548*
..	... ...	...		
..	... ...	...		

Table 2: Household compositions

Household #	Persons by class			(*W*)
	Men	Wo-men	Child-ren	
1	1	1	0	*1.042*
2	1	2	1	*1.001*
3	1	1	2	*1.548*
4	0	1	1	*0.932*
5	1	0	0	*1.321*
6	0	1	1	*0.932*
7	1	1	1	*1.270*
8	1	2	0	*0.821*
9	0	1	2	*1.136*
Totals	6	10	8	*10.000*

Table 3: Original and adjusted totals

	House-holds	Persons by class		
		Men	Wo-men	Child-ren
Original totals	9	6	10	8
Added for adjustment	1	1	0.5	1.5
New totals	10	7	10.5	9.5
Adjustment rate	11.1%	16.7%	5%	1.875 %

4. Whole- and within-household adjustments.

We now consider the distinction between within-household adjustments (that is, adjustments for omissions of persons within enumerated households) and whole-household adjustments (that is, adjustments for omissions of whole households). This distinction has previously been made for purposes of analysing the causes of undercount (Fay 1986). Our concern here is to use it to more accurately represent the undercount in an adjustment.

Within-household adjustments do not involve adding any households to the roster, but only shifting weight between households to increase the weighted totals of persons in the various classes. (That is, households with few or no persons in a particular class are downweighted and those with many are up-weighted, so that the total household weight remains constant.) Thus, in this portion of the adjustment, some households will inevitably have their weights reduced. Whole-household adjustments, on the other hand, correspond to households that were omitted entirely from the census. These adjustments do not reflect on the accuracy of the enumerated households; thus they should be represented by adding households to the roster without taking weight away from the households that were enumerated.

We propose to separate these two portions of the adjustment by writing two separate sets of constraints. After fitting the two corresponding sets of weights, the two weights for each household are added to obtain weights that incorporate both parts of the adjustment. The distinction between whole- and within-household adjustments contains information which generally leads to a different set of adjusted weights than would be calculated if the adjustments were combined. However, if this distinction is not made in the estimation of the undercount, an adjustment can still be calculated in a single step.

5. Feasibility of constraints.

In the preceding sections we have assumed that feasible solutions exist to the constrained optimization problem. Here we will consider situations in which the solutions will not exist or will be unsatisfactory, and some alternative methods to deal with these situations.

5.1. When will constraints be non-feasible? There are three ways in which the constraints may fail to allow of satisfactory solutions: (1) when the constraints are actually inconsistent, (2) when the constraints are consistent but there are no positive weights that satisfy them, and (3) when there is a feasible solution but it involves an extreme adjustment to some household weights. The issues associated with these three failure modes are fairly similar.

One could write down constraints that are intrinsically inconsistent, for example that all classes of men are adjusted upward by 2% while men in total are adjusted upward by 4%. In our procedure each constraint applies to the number of persons in a distinct adjustment class and so there are no inconsistencies of this sort. However, a contingent inconsistency is still possible, that is to say one that depends on the particular collection of household compositions that appears in a block. The following are examples of contingent inconsistency or infeasibility:

(1) Proposed undercount estimation methods envision defining over 100 adjustment classes. In a small but diverse block the number of classes represented might be larger than the number of households; hence the number of constraints would be larger than the number of weights to be fitted. An inconsistency is then almost inevitable.

(2) The adjustment of the number of households may be too large or small to accomodate the adjustment of persons in some class. (This may represent a failure of the model for adjustment of the number of households.) For example, suppose that the number of men to be added by the whole-household adjustment is greater than the number of households to be added, but no household in the block has more than one man. The constraints then might be consistent but infeasible, since they could be satisfied only by assigning negative weight to some households without men.

(3) The block may have had omission rates atypical of blocks in the PES on which omission rates were estimated. For example, suppose that in most blocks (including most of the PES sample blocks), adult males with certain characteristics tend to be heavily undercounted, but the block being adjusted is atypical in having adult males of this class present in most households and well counted. The class undercount estimate might lead to an extreme upward adjustment that could not be accomodated within the existing households.

Problems of infeasibility may also arise where the difficulty cannot be so easily traced to a particular inconsistency in the adjustment.

5.2. Making the constraints feasible. Regardless of the stage of the fitting procedure at which the infeasibility is discovered, several methods are available to relax the constraints and make them feasible. In this section, we survey several such methods, drawing out both the intuitive logic of each choice and the computational methods required.

5.2.1. Methods based on dropping rows (constraints) of A. When checking for consistency of constraints, some rows may be found to be linearly dependent on the previous rows and hence either redundant or inconsistent. If these rows are simply dropped from the A matrix, a consistent set of constraints is obtained; thus, no further computational effort is required.

If the constraints are arranged in sequence from the most important to the least important, than the less important constraints will be dropped when they are inconsistent with the more important ones. This ordering makes the most sense if the original constraints on distinct adjustment classes (defined by a multi-way classification of the population) are reframed in an ANOVA-like manner as constraints on total population ("grand mean"), classes defined by one classification variable ("main effects"), and classes defined by interactions. For example, if there are ten adjustment classes defined by two sexes and five age ranges, the reframed constraints in order of importance might be: total population (1 constraint), population by sex (1 more constraint), population by age (4 more constraints), age-sex interactions (the remaining 4 constraints). The 4 age constraints could be further broken down as old-vs.-young (1 constraint) and 3 further constraints within those larger groups.

A similar procedure can be applied at the stage of checking feasibility of the constraints. If it is not possible to make all of the $Z_i = 0$, the objective function in the linear programming problem can be modified to be $\Sigma c_i Z_i$, with the coefficients $c_i > 0$ corresponding to the most important constraints made larger. Then a maximal set of feasible constraints can be identified, and the remaining constraints dropped.

The outcome of this procedure would be weights that give the correct block totals on the coarser classifications of persons, while failing to be correct on all cross-tabulations.

5.2.2. Methods based on adding columns (households) to A. When constraints are only contingently infeasible (in the previous sense that infeasibility depends on the particular set of household compositions in the block), they become feasible when households are added that have the required composition. The simplest application of this principle is to work at a higher level of geographical aggregation than a block. A few adjacent blocks may be combined when problems arise in fitting, or the entire roster may be grouped at, for example, the enumeration district level before weighting. The larger the unit, the broader the range of household compositions that will be represented and the less likely that problems of infeasibility will arise.

A more sophisticated procedure would use a hot-deck of households from adjacent "donor" blocks to enrich the pool of households to which weight can be assigned. Computational simplicity is important here since it may be necessary to scan through a long list of households to find the one or ones which will make the constraints feasible. In the consistency-checking stage, if row j of A is dependent on the previous rows, then if the column for the added household is independent of the columns of A (with regard only to the first j rows), row j of the augmented A will be independent. In the stage of checking for feasibility, if the algorithm halts because no reduction can be made in the objective function ΣZ_i, the search for basic columns can be extended to columns corresponding to households in the hot deck. Finally, if some household's fitted weight is extremely high, the hot deck can be scanned for other households that would also receive high weights with the current values of λ (that is, columns a such that $a'\lambda$ is large). If these are added to the block they will draw off some of the weight from the overweighted households when the weights are refitted, since they are likely to also have members in the same adjustment classes.

The intuition behind this method is that the household compositions that are enumerated in a block are only a sample of those which actually could have appeared there had the enumeration been complete. The observed distribution of household compositions is smoothed by mixing it with the distribution for adjacent blocks, which contain households that are also typical for that area. Thus, conceptually this method is related to Bayesian smoothing methods that improve estimation of some quantity for one unit by borrowing strength from its distribution in similar units.

The donor blocks could be chosen by a sequential hot deck procedure; then, the donor blocks would tend to be geographically close to the adjustment block and no particular set of blocks would have undue influence on the entire census. By detailed stratification of blocks, the donor blocks could be selected to be similar to the block being adjusted on characteristics such as mean income, types of housing units, and racial balance.

5.2.3. Combined methods. The two types of methods outlined above can be combined by an appropriate reframing of constraints. The principle here is to satisfy <u>all</u> constraints in the larger geographical units, while satisfying only the more important constraints in the smaller units. This type of compromise may make it possible to get a fairly good fit to the desired distribution without having to add additional records to the roster.

Suppose that the A matrices for several blocks have been reframed similarly as sequences of rows representing main and interaction constraints. Then a single large A matrix representing all of the constraints can be formed. The rows for the more important constraints can be kept separate, while rows for subsidiary constraints can be combined across blocks. For example, suppose there are ten adjustment classes, defined by sex (2 levels)and age (5 levels), and two blocks. Altogether there are eleven constraints (one for number of households and one for each adjustment class) in each block. If these are combined into a single matrix, keeping main effects and two-way interactions, the constraints are: block household counts (2 constraints), block populations (2 constraints), sex (1 constraint), age (4 constraints), block $\times$ sex interaction (1 constraint), block $\times$ age interaction (4 constraints), and sex $\times$ age interaction (4 constraints) in the combined blocks. Here 4 constraints have been eliminated (block $\times$ sex $\times$ age interaction); in a more realistic problem with more blocks, classification variables, and levels, the reduction would be much greater.

6. Simulation results.

Simulations were performed to answer two classes of questions: For these simulations, real households (from a Public Use Microdata Sample) and undercount estimates (from the Census Bureau's Test of Adjustment Related Operations) were used.

6.1. Feasibility simulations. The first set of questions is concerned with evaluation of the success of the algorithm in terms of its own constraints and objectives. Does the reweighting algorithm give an answer? In real problems, is there a solution to the weighting constraints?

To answer these questions, "feasibility simulations" were performed in which the weighting algorithm was applied to simulated blocks made up of real households, using real adjustment rates. This procedure thus closely parallels the practical application of the algorithm.

The algorithm was almost always able to yield a set of weights satisfying all constraints when there were 50 households per block for each racial group. (Each race was subdivided into 20 adjustment classes.) However, because one racial group (Asians) was only lightly represented in the area used for the simulations, even a large block would not usually have enough Asian families to make the constraints consistent; thus it would be necessary to use some of the methods described in Section 5 for this population.

6.2. Inference simulations. The second set of questions is concerned with evaluation of the success of the algorithm in improving the quality of inferences based on a micro-data set: does the weighted micro-data set more accurately describe the real world than the raw, unweighted data?

To answer these questions, simulated blocks made up of real households were drawn, representing the true (but unobserved) compositions of households in blocks. For each "true" block, omissions were imposed using real estimated undercount rates and a plausible model for the distribution of undercount among households. The weighting algorithm was applied to the "enumerated" blocks generated in this way. Summary statistics describing household composition were calculated for the simulated "true" blocks and for the simulated observed blocks with undercount, both unweighted and weighted for undercount adjustment. The goal of these "inference simulations" was to determine whether the reweighting brought the statistics closer to their values in the "true" blocks; in other words, did reweighting correct the biases caused by the undercount?

Several sets of statistics were used in evaluation of the reweighting procedure. These were all chosen because they summarized household characteristics that are *not* functions of the populations by adjustment class. The first set was the distribution of sizes (number of members) of households. Note that the mean number of persons per household, like any function of the class totals and household count, will automatically be adjusted to the correct (pre-undercount) values; the distribution of sizes, however, is not controlled by the adjustment procedure.

The second set of statistics was the distribution of number of *adult* (over 14 years old) members in households with one or more *children* (up to 14 years old). The last two sets of statistics were the distribution of the age group (coded from 1 to 5 as in the formation of the adjustment classes) of the *oldest male* in the household (coded 0 if no male is present), and the same distribution for households with one or more children. In these cases, the mean is not automatically adjusted to the correct value, since it depends on the joint distribution of counts from different classes within households as well as on marginal totals.

The results of these simulations are summarized in Table 4. The lines of each table are labelled "true" (for the original pseudo-blocks), "enum" (for the simulated enumerated blocks, i.e. after omissions due to undercount), and "adjust" (enumerated blocks after adjustment for undercount).

Household size distribution was biased downwards in the enumerated blocks. As well as correcting the mean, adjustment brought the estimated percentage for every size substantially closer to the true percentage.

The distribution of number of adults in households with children was also biased downwards. The majority of these households had contained two adults, so this size category was most understated by the enumerated statistics. Due to the log-linear structure of the adjustment, however, the most extreme adjustments were made to the largest and smallest households. Thus, the highest size categories were slightly overadjusted and intermediate categories were underadjusted; the "size 2" category was adjusted a small amount in the wrong direction. Nonetheless, the mean of the adjusted distribution was much closer to the "true" value than the unadjusted mean was.

The story is similar for the distributions of age of oldest male. Although these statistics are only indirectly related to the counts by class, in almost every case the adjusted distributions and means are closer to the "truth" than are the unadjusted distributions and means.

In summary, these simulations suggest that these weighting adjustments can improve estimates of measures of household structure as well as the aggregate counts for which they were intended. However, reweighting does not provide accurate adjustments with certain configurations of the data; to deal with these situations may require a model-based imputation method such as that outlined by Zaslavsky (1989).

6. Adjustment of covariate information

Covariates that are classification variables for formation of adjustment classes are automatically adjusted by the procedure. For example, if "sex" is used in forming adjustment classes, the sex ratio is automatically adjusted.

More complicated structural measures are usually (but not necessarily) improved by reweighting. Examples are given in the preceding section.

Other covariates may not be properly adjusted. For example, "income" may require further adjustment, since reweighting will not necessarily upweight families with the right income levels.

This is a social science question as well as a statistical question: how do households that are (partially or completely) omitted from census compare with enumerated households? Is omission "ignorable?"

If omitted households have the same income as enumerated households with same composition, then we may reweight with no further adjustment. (This implies an underenumeration of income.) On the other hand, if omitted households systematically lag behind enumerated households, then there should be a regression adjustment for the income differential, or we should constrain the weights so that reweighted mean incomes equal an estimated adjusted mean.

7. 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.

FAY, R.E. (1986), "Implications of the 1980 PEP for Future Census Coverage Evaluation," U.S. Bureau of the Census, unpublished.

IRELAND, C.T. and KULLBACK, S. (1968), "Contingency Tables with Given Marginals," *Biometrika* 55:179-188.

NATIONAL ACADEMY OF SCIENCES (1985), *The Bicentennial Census: New Directions for Methodology in 1990*, Washington, National Academy Press.

RUBIN, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York, Wiley.

OH, H.L. and SCHEUREN, F.J. (1978), "Multivariate Ratio Raking Estimation in the 1973 Exact Match Study," *Proc. Section on Survey Research Methods, ASA*, pp. 716-722.

ZASLAVSKY, A.M. (1988), "Representing Local Area Adjustments by Reweighting of Households," forthcoming in *Survey Methodology*.

ZASLAVSKY, A.M. (1989), "Representing Census Undercount at the Household Level," Ph. D. thesis, Department of Mathematics, M.I.T.

Table 4. Inference simulation results.
(All rows except means are percentages.)

Size distribution

		true	enum	adjust
All Households	size 1	7.240	10.349	7.372
	size 2	16.200	19.631	16.421
	size 3	20.240	21.772	20.690
	size 4	22.600	20.690	21.392
	size 5+	33.720	27.558	34.219
	mean	3.971	3.632	3.871
Adults, Households with children	size 0	0.000	1.736	0.924
	size 1	6.925	18.309	13.277
	size 2	58.404	49.874	48.557
	size 3	17.214	15.965	18.223
	size 4	9.125	7.677	9.810
	size 5+	8.322	6.439	9.209
	mean	2.585	2.323	2.562

Age of oldest male (5 age groups)

		true	enum	adjust
All Households	none	7.080	9.981	7.853
	age 1	4.000	7.388	5.989
	age 2	28.680	26.296	26.307
	age 3	33.800	30.972	33.439
	age 4	21.960	21.160	21.931
	age 5	4.480	4.203	4.480
	mean	2.730	2.585	2.690
Households with children	none	3.602	5.809	4.272
	age 1	6.214	11.723	9.069
	age 2	30.744	27.321	27.242
	age 3	42.649	39.096	42.038
	age 4	15.843	15.158	16.418
	age 5	0.949	0.894	0.962
	mean	2.638	2.488	2.601