Using Equalization Constraints to Find Optimal Calibration Weights

Reid A. Rottach and David W. Hall, U.S. Bureau of the Census

Abstract[*]

Weights that are used for direct estimation using survey data are often constructed using calibration methods, such as benchmarking to auxiliary totals and weight equalization. Benchmarking (forcing certain estimates to match known totals) has been shown to reduce variances for statistics correlated with the auxiliary characteristics, and weight equalization (forcing the weights within higher-level units to be equal) has been shown to further reduce variances for statistics measured on the higher-level units. We will examine the effect of adding a family equalization constraint to person level calibration weights using data from the 2001 Panel of the Survey of Income and Program Participation (SIPP). We also consider weights that are an average of those with and without equalization constraints, such that the combined weights minimize variance of certain statistics. Note that the combined weights will maintain the benchmarking constraints.

Key words: weight equalization, integrated weighting, raking, variance estimation, SIPP

1. Introduction

The SIPP is a national household survey, providing detailed information on the economic situation of persons, households, and families. SIPP data is often used to study wealth, well-being, and participation in national assistance programs. Many of the statistics that are of particular interest are correlated with living in a low-income household.

At the design stage, we use stratified sampling and weighting adjustments in an effort to improve efficiency of such statistics. Currently, the weighting adjustments at the benchmarking stage depend on demographic characteristics of the individual, without consideration of the makeup of the individual's family. The purpose of this research is to consider whether we can further improve the precision of key SIPP statistics by incorporating information about sampled families, such that an individual's final estimation weight depends in part on the characteristics of other family members. Primarily, we consider family-level weight equalization, which was studied by Lemaitre and Dufour [1987], who described their approach as an integrated method of constructing weights for both persons and families, and is now often referred to as "integrated weighting." These authors found that the precision of estimates on numbers of persons were not substantially affected by the use of equalized weights, but estimates of numbers of families generally had improved precision when compared to "principal person" methods. The principal person method of estimation uses the weight of someone selected in the family (or household) to assign to the family, without first requiring that all persons in the family have the same weight.

In our application to SIPP data, the results are similar: the variance of an estimated number of families is lowered, but there is little effect on any other statistic we considered. Furthermore, there was little or no improvement in efficiency resulting from averaging the two calibration weights.

In this paper, we will describe the construction of weights, provide numerical results in our application to SIPP, and discuss practical concerns that could affect the feasibility of equalizing weights within families.

2. Constructing Weights for Estimation

2.1 Design Weights

The adjustments we describe are used to fine-tune design weights, which are valid estimation weights that reflect the sampling design of the survey. In particular, the design weight of a sampled element is the inverse of its selection probability. The use of design weights in practice may be just hypothetical if we distinguish between the initial probability of selection and the probability of selecting a sample respondent. For this paper, we assume the design weight is the inverse probability of selecting a sample respondent, ignoring the fact that this parameter is not known, but estimated. It is an important distinction in discussing weight equalization, which depends on the level at which the design weights are equal.

The ultimate sampling unit clusters (USUs) in a complex sample design consist of elements that are selected together, and therefore have the same

[*] **Disclaimer:** This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical or methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau.

probability of selection. In our case, we will use this term to refer to the clusters that are not only sampled together, but undergo the same adjustment for non-response.

2.2 Calibration

Calibration weighting offers a way to incorporate auxiliary information into survey estimates so that, in general, characteristics that are correlated with the auxiliary variables are estimated with greater precision. The information required for calibration is a set of population control totals $\Sigma_U \mathbf{x}_k$, where U is the finite population universe and the $\mathbf{x}_k$ are vectors of auxiliary information that are known individually only for elements in the respondent sample. Calibration uses this information by constructing weights such that $\Sigma_s w_k \mathbf{x}_k = \Sigma_U \mathbf{x}_k$, where s represents the respondent sample and w_k is the calibrated weight for element k. Typically, there are many possible choices of weights that satisfy this benchmarking constraint. Calibration, by its classical definition, produces the one that is closest to the design weights, with closeness determined by a suitable distance function (See Deville, Särndal, and Sautory [1993] for details). So selecting a calibration estimator reduces to the selection of a distance function.

There are two well-established estimators that can be expressed as a weighted sum with the weights defined by calibration: the generalized regression estimator (GREG) and the raking ratio estimator. The value of calibration is not that it offers an alternative method of producing these estimators, but that its flexibility may offer improvements to them. While the weighted-sum form of the GREG may have negative weights, for example, calibration allows for range restrictions, which can force weights to be positive. And the procedure for constructing raking ratio weights, as described by Deming and Stephan [1940], is used in cases where the auxiliary vectors are indicators of membership in control groups. Calibration techniques allow the use of continuous auxiliary vectors, which we will show is the primary reason weight equalization can be accomplished with calibration and not the raking ratio algorithm as currently implemented in SIPP.

2.3 Weight Equalization

When two sample elements have the same design weight and the same values of their corresponding auxiliary vectors, their calibrated weights will be the same. This property leads to the method of equalizing weights that we use.

The weights are constructed so they satisfy:

(a1) $\Sigma_s w_k \mathbf{x}_k = \Sigma_U \mathbf{x}_k$

as well as many added constraints of the form:

(a2) $w_k = w_l$

The trick is not to calibrate using the x-vectors, but using alternative auxiliary vectors that lead to conditions (a1) and (a2). The new vectors are represented by $\mathbf{z}_k$, for k in U. To calibrate using these, $\Sigma_U \mathbf{z}_k$ must be known, and $\mathbf{z}_k$ must be known for all k in s. And to satisfy (a1) and (a2), they are constructed so that:

(b1) $\Sigma_U \mathbf{z}_k = \Sigma_U \mathbf{x}_k$

(b2) $\Sigma_s w_k \mathbf{z}_k = \Sigma_s w_k \mathbf{x}_k$

(b3) $\mathbf{z}_k = \mathbf{z}_l$ wherever we want $w_k = w_l$, only allowing the constraint if elements k and l are in the same USU

The construction is simple - if we would like all elements in group p to have the same weight after calibration, define $\mathbf{z}_k$, for k in p, to be the average of the x-vectors in p. That is, $\mathbf{z}_k = \Sigma_p \mathbf{x}_k / n_p$, where n_p is the number of elements in p. It is straightforward to show (b1), (b2), and (b3) are satisfied in this case, and therefore, so are (a1) and (a2).

There are two main reasons to require that elements be in the same USU if we constrain their weights to be equal. The first reason is that their design weights will be equal, which was one of our assumptions; this is a necessary condition for the approach to lead to equal calibration weights. Secondly, the z-vectors must be known for all elements in sample. So, if $\mathbf{z}_k$ is a function of $\mathbf{x}_k$ and the x-vectors of other elements, those other elements should be guaranteed to be in sample whenever k is in sample. Constraining these other elements to be in the same USU as element k would satisfy this requirement.

In our application, we look at equalizing weights within each family. This will be referred to as family-level weighting, as compared with weighting without the constraint, which will be referred to as person-level weighting.

2.4 Range Restrictions

One of the benefits of calibration is that it offers control over the distribution of weights, specifically the g-weights. A g-weight is the ratio of the calibration and design weights.

Each range-restricted estimator described in Deville, Sarndal, and Sautory [1993], and similarly Singh and Mohl [1996], forms a family of distance functions defined by the upper and lower bounds imposed on the g-weights. Their distance functions generalize those of other calibration estimators. The linear truncated estimator, for example, gives the GREG if the range is wide enough to be effectively unrestricted.

The choice of upper and lower bounds requires a balancing of the desire to have all g-weights close to one, while ensuring a solution to the calibration equation exists and can be found in a reasonable number of iterations.

2.5 Optimal-Average Weights

We consider a case where the weights are not necessarily equal within each family, but will have less variation within each family than the person-level weights. This is accomplished by averaging the person- and family-level weights. In particular, we look at an optimal-average weight (equation 1, below). In this expression, $w_{p,k}$ is the person-level weight and $w_{f,k}$ is the family-level weight. The optimal choice of α is the one that minimizes the variance of an estimated total; this will have the form given in (equation 2), where $\hat{t}_p = \Sigma_s w_{p,k} y_k$, and $\hat{t}_f = \Sigma_s w_{f,k} y_k$. Its variance is given in (equation 3), where $\hat{\sigma}_p^2$ is the estimated variance of $\hat{t}_p$, $\hat{\sigma}_f^2$ is the estimated variance of $\hat{t}_f$, and $\hat{\sigma}_{p\,f}$ is the estimated covariance of the two statistics. The choice of α that minimizes this variance is given in (equation 4).

$$w_{opt,k} = \alpha\, w_{p,k} + (1 - \alpha)\, w_{f,k} \qquad (1)$$

$$\hat{t}_{opt} = \alpha\, \hat{t}_p + (1 - \alpha)\, \hat{t}_f \qquad (2)$$

$$\hat{\sigma}^2 = \alpha^2\ \hat{\sigma}_p^2 + (1 - \alpha)^2\, \hat{\sigma}_f^2 + 2\,\alpha\,(1 - \alpha)\ \hat{\sigma}_{p\,f} \qquad (3)$$

$$\alpha = \frac{\hat{\sigma}_f^2 - \hat{\sigma}_{p\,f}}{\hat{\sigma}_p^2 + \hat{\sigma}_f^2 - 2\,\hat{\sigma}_{p\,f}} \qquad (4)$$

2.6 Linearizing the Weight Adjustments

All calibration estimators are asymptotically equivalent, and in practice, evidence has shown only minor differences among the estimators even for modest sample sizes, with the possible exception of instances where tight range restrictions are imposed [Singh and Mohl, 1996]. So it is often assumed that the variance of any calibration estimator is reasonably well approximated by treating it as though it were the GREG. Suitable variance estimators for the GREG are usually easy to compute, compared with other calibration estimators, due to the fact that the GREG does not require iterations to calculate. We use a residual technique of variance estimation, which was developed for the GREG.

In this case, the variance of a calibration estimator, such as $\hat{t}_p$ or $\hat{t}_f$, is approximated by the variance of a weighted sum of residuals, $\Sigma_s w_k e_k$, which is treated as a linear statistic. This may also be expressed as $\Sigma_s d_k g_k e_k$, where d_k and g_k are the design and g-weights, respectively. Since a linear statistic is defined as one that is linear in the design weights, this form more clearly identifies $g_k e_k$ as the linearized *variable*. See Binder [1996] for a derivation of this result.

3. Application

3.1 SIPP Sampling and Estimation

The 2001 panel of SIPP has a two-stage sampling design, in which primary sampling units (PSUs) are selected in the first stage from regionally defined strata, and consist of groups of counties. Some strata are comprised of a single PSU, which is selected with certainty. Since the sample from these PSUs is not weighted to represent any other geographic area, the PSU is called self-representing (SR). The remaining strata consist of multiple PSUs, from which two are selected with probability proportional to size, following Durbin's plan of sampling without replacement [Durbin, 1967]. These PSUs represent a larger stratum of PSUs and are therefore non-self-representing (NSR). In the second stage of sampling, clusters of housing units are selected systematically from a list sorted on demographic variables derived from the decennial census.

The variance estimator we use treats the first stage selection of NSR PSUs as though it were with replacement – that there was a nonzero probability of choosing the same PSU twice. Variance estimation that is unbiased for "with replacement" selection generally incurs a small positive bias when sampling occurs without replacement. We choose this biased estimator due to its computational simplicity over the alternative unbiased estimator. In this case, the two PSUs form replicate clusters (half-samples) selected from the strata. In SR PSUs, variance estimation for the systematic sample is based on the paired selections model of Kish [1965]. This results in stratum and half-sample assignments, similar to the NSR PSUs.

With two half-samples per stratum as we described, a suitable variance estimator of a linear statistic, in particular the weighted sum of residuals, is:

$$\hat{\sigma}_i^2 = \Sigma_h (E_{i\,h\,1} - E_{i\,h\,2})^2 \qquad (5)$$

where i indexes the two calibration estimators (p and f); h represents stratum; and $E_{i\,h\,j}$ represents the $w_{i,k}$ - weighted sum of residuals in stratum h, half-sample j.

Further details on SIPP weighting and applying the residual technique of variance estimation to SIPP estimates can be found in Rottach and Hall [2003].

3.2 Meeting All Constraints

The three sets of constraints imposed on the estimation weights – benchmarking, equalization, and range restrictions – are substantial, and it is possible not all can be met. It would be desirable to impose constraints that can be met with a high likelihood, with some mechanism for loosening the constraints if a solution cannot be found. This is especially important if a replicate weighting approach to variance estimation is used, where the procedure should be repeatable without replicate-by-replicate modifications.

One option is to form the final control groups (cells) such that they meet certain requirements based on the collected sample. When a given cell does not meet the requirements, it is collapsed with another to form a new cell. The SIPP employs this option, requiring that each cell have some minimum sample size, and that each population control divided by the estimate of that control using design weights be between 0.67 and 2. This should reduce the number of occurrences of especially large or small final weights relative to design weights, but does not strictly impose upper and lower bounds on the g-weights as would a range restriction.

Another possible approach would be that of ridge regression in calibration, as described in Rao and Singh [1997]. This procedure is only approximate calibration, in that it allows for some discrepancies between the population controls and their estimates. We will not apply this method in our examples.

Range restrictions have a substantial affect on the rate of convergence. There does not appear to be much advantage to imposing very tight bounds on these. We chose a range of 0.2 to 5 since it disallowed extremely small or large weights, but was loose enough to allow convergence in a small number of iterations for our application.

4. Numerical Results

4.1 Programming

The methods we implemented were programmed in SAS/IML using the algorithms described in Singh and Mohl [1996]. The results presented here use linear truncated calibration.

4.2 Differences in the Estimates

One of the assumptions we make for variance estimation is that our calibration estimator is a close approximation to the GREG. For validation, we compared the estimates of poverty, health insurance coverage, social security, and foodstamp participation. In no case did the estimates using our calibration estimator and the GREG differ by more than 0.1%. This was true for calibration with and without family-level equalization.

The relative precision (RP) of these estimates is calculated as a ratio of standard errors, with the numerator being that of person-level weighting. The relative difference (RD) of the estimates is measured relative to those with person-level weights.

Table 1
Relative Differences and Relative Precision[a]

Statistic	RD (%)	RP
Persons in Poverty		
FW	-1.93	.99
OW (α=.60)	-.76	1.01
Families in Poverty		
FW	-3.34	1.02
OW (α=.28)	-2.28	1.03
Number of Families		
FW	.13	2.15
OW (α=0)	.13	2.15
Number Uninsured		
FW	.11	.99
OW (α=.62)	.04	1.00
Social Security Recipiency		
FW	-.47	.97
OW (α=.97)	-.01	1.00
AFDC/TANF		
FW	-2.63	1.01
OW (α=.38)	-1.63	1.02
Food Stamps		
FW	-1.93	1.04
OW (α=.11)	-1.27	1.04

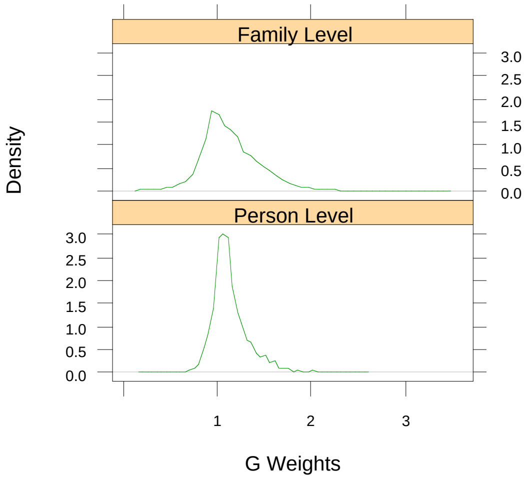
[a] FW=family-level weighting; OW=optimal weighting

4.3 Distribution of Weights

The person-level calibration weights were in some way the ones closest to the design weights that satisfied the benchmarking constraint, so it follows that the family-level weights will be further from the design weights in this respect. The greater dispersion of g-weights resulting from equalization constraints is clearly seen in Figure 1. In general, when estimation weights have added variability, so do the statistics computed using the weights. So, the reduction in variance we might expect by incorporating information about families could be counteracted by the effect of a wider distribution of g-weights.

The wider dispersion we saw in our application was different for different populations, such as the age groups shown in Figure 2 (last page).

Figure 1
Distribution of g-weights



5. Conclusion

In our application, we found that weight equalization improved the precision of estimates of numbers of families, reducing the standard errors by half. All other statistics were only minimally affected, including the estimated number of families in poverty. We do not consider the statistic "number of families" itself of particular importance in our study, but a reduction in its variance could indicate a reduction in variance of (possibly correlated) statistics that are fundamental to the survey.

So the advantage of using this approach for improved efficiency of key SIPP statistics remains a conjecture, and one that would need to be weighed against practical concerns, such as those related to convergence.

6. References

Binder, D. (1996), "Linearization Methods for Single Phase and Two-Phase Samples: A Cookbook Approach," Survey Methodology, 22, 17-22

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, pp. 427-444

Deville, J., Särndal, C.E., and Sautory, O. (1993), “Generalized Raking Procedures in Survey Sampling Journal of the American Statistical Association, 88, 1013-1020

Durbin, J. (1967), "Design of multi-stage surveys for the estimation of sampling errors," Applied Statistics, 16, 152-164

Jayasuriya, Bodhini R., and Valliant, Richard (1996), "An Application of Restricted Regression Estimation in a Household Survey," Survey Methodology, 22, 127-137

Kish, L. (1965), *Survey Sampling*, New York: John Wiley & Sons, Inc.

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

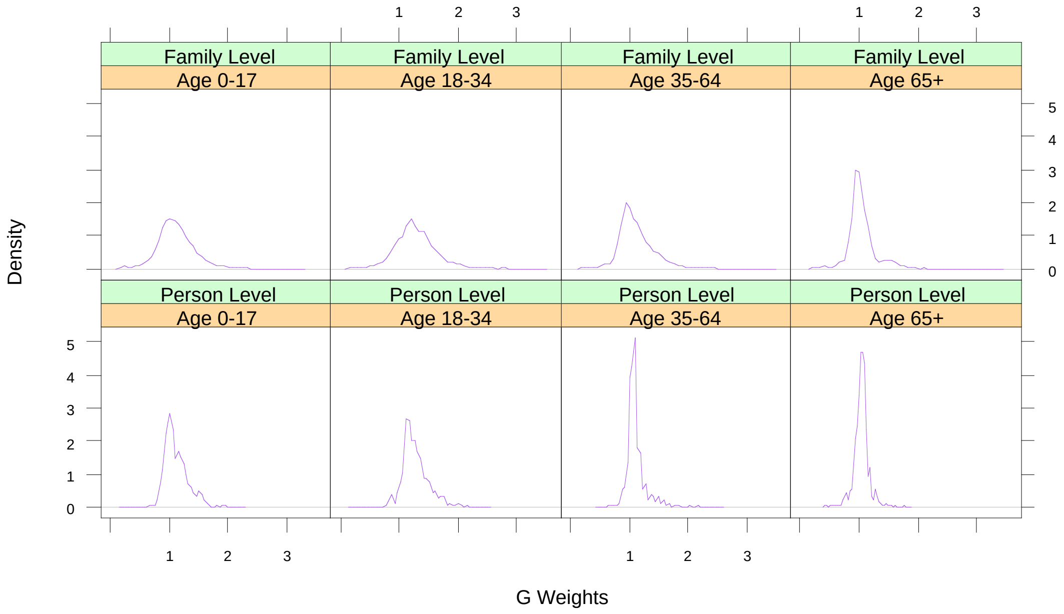
Luery, D.M., (1986), “Weighting Sample Survey Data Under Linear Constraints on the Weights,” Proceedings of the Section on Social Statistics, American Statistical Association, Chicago, IL.

Rao, J.N.K. and Singh, A.C., (1997), "A Ridge-Shrinkage Method for Range-Restricted Weight Calibration in Survey Sampling," Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 57-65

Rottach, R., and Hall, D., (2003), “Methods of Statistical Inference for the Survey of Income and Program Participation that are Suitable for an Online Data Analysis Tool,” Proceedings of the Section on Survey Research Methods, American Statistical Association, San Francisco, CA.

Singh, A.C. and Mohl, C.A., (1996), "Understanding Calibration Estimators in Survey Sampling," Survey Methodology, 22, 107-115

Figure 2
Distribution of g-weights by Age Group


Family Level
Age 0-17
Family Level
Age 18-34
Family Level
Age 35-64
Family Level
Age 65+
Person Level
Age 0-17
Person Level
Age 18-34
Person Level
Age 35-64
Person Level
Age 65+
Density
G Weights