WHERE WILL IT ALL END?
SOME ALTERNATIVE SASS ESTIMATION RESEARCH OPPORTUNITIES

Steven Kaufman and Fritz Scheuren
Fritz Scheuren, The George Washington University, Washington, D.C. 20052-0001

KEY WORDS: GENERALIZED REWEIGHTING, MASS IMPUTATION

1. INTRODUCTION

For 1993-94, both NCES's Schools and Staffing Survey (SASS) and its Private School Survey (PSS) were conducted. SASS has a large private school component as part of its overall sample. PSS is essentially a census of private schools -- but with considerably less item content than SASS. Over the last two years at these Joint Statistical Meetings, a modified Generalized Least Squares (GLS) estimator has been explored in SASS to see if we could achieve simultaneous intersurvey consistency in comparable totals for schools, teachers, and students between SASS and PSS.

We were successful (Scheuren and LI 1996); but, in the end, the results were ultimately disappointing in that so little use had been made of the exceptionally rich PSS data. If our goal had been more general -- say, improving SASS estimates -- not just trying to achieve a limited consistency with PSS, then alternatives to GLS become attractive. In this year's paper we look generally at what can be done when conducting a survey in a data rich setting. These general observations are then applied to private school surveys. Among the methods advocated in this setting, "mass imputation" is given special attention, since it has not been used as often as its considerable strengths might warrant.

Organizationally, the present paper is divided into four parts: This introduction begins the discussion (Section 1). A summary of our results with Generalized Least Squares (GLS) estimators makes up Section 2. Section 3 presents an alternative to reweighting SASS which has been called "mass imputation." This technique, now roughly 20 years old (Colledge et al. 1978), imputes records from a survey back to its sampling frame; and, in a sense, operates in making estimates as if there had been a census. The final section discusses some "What Nexts."

2.GLS ESTIMATION IN SASS

As already noted, a Generalized Least Squares (GLS) technique was used to achieve simultaneous consistency or near consistency in totals for schools, teachers, and students between the private school component of the 1993-4 SASS and the 1993-4 PSS (Scheuren and Li, 1996). This technique is described briefly below (subsections 2.1 and 2.2), then our 1993-4 results summarized (subsection 2.3). A discussion of our (possibly misplaced) expectations concludes the section (in subsection 2.4).

2.1 Generalized Least Squares. -- Advocated by Deville and Särndal (1992), GLS can be used (as in Imbens and Hellerstein 1993) to achieve consistency between SASS and PSS. To see how GLS works in this setting it is necessary to define some notation; in particular --

$\mathbf{w}_i$	is the original SASS Private School base weight for the ith SASS observation, i=1,...,n.
$\mathbf{t}_i$	is the SASS total of teachers for ith SASS observation, i=1,...,n.
$\mathbf{s}_i$	is the SASS total of the students for the ith SASS observation, i=1,...,n.
N	is the total estimated number of schools, as given by PSS.
T	is the total estimated number of teachers, as given by PSS.
S	is the estimated total number of schools, as given by PSS.

In reweighting SASS three constraints are imposed on the new weights $\mathbf{u}_i$,

$$\sum \mathbf{u}_i = \mathbf{N}$$

$$\sum \mathbf{u}_i \mathbf{t}_i = \mathbf{T}$$

$$\sum \mathbf{u}_i \mathbf{s}_i = \mathbf{S}$$

For our application the new weights $\mathbf{u}_i$, subject to these constraints, are to be chosen, as in Burton (1989), to minimize a loss function which can be written as the sum of squares, $\sum (\mathbf{u}_i - \mathbf{w}_i)^2$. Motivating this loss function here is outside our present scope, except to say that the sensitivity of the final results to the loss function chosen (e.g.,Deville and Särndal 1992; Deville et al. 1993) seems not to be too great. Now the usual Lagrange multiplier formulation of this problem yields after some algebra that the new weights are of the form $\mathbf{u}_i = \mathbf{w}_i + \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2\mathbf{t}_i + \boldsymbol{\lambda}_3\mathbf{s}_i$, where the $\boldsymbol{\lambda}$'s are obtained from the matrix expression $\underline{\mathbf{d}} = \mathbf{M}\underline{\boldsymbol{\lambda}}$ with the vector $\underline{\mathbf{d}}$ consisting of three elements, each a difference between the corresponding PSS and SASS totals for schools (first component), teachers (second component), and students (third component); in particular

$$\mathbf{N} - \sum \mathbf{w}_i$$

$$\mathbf{T} - \sum \mathbf{w}_i\mathbf{t}_i$$

$$\mathbf{S} - \sum \mathbf{w}_i\mathbf{s}_i$$

where the summations are over the SASS sample observations and the quantities: **N**, **T**, and **S** are known PSS totals for schools (**N**), teachers (**T**), and students (**S**) respectively.

The matrix **M** is given by:

$$\begin{matrix} \mathbf{n} & \sum \mathbf{t}_i & \sum \mathbf{s}_i \\ \sum \mathbf{t}_i & \sum \mathbf{t}^2_i & \sum \mathbf{t}_i\mathbf{s}_i \\ \sum \mathbf{s}_i & \sum \mathbf{t}_i\mathbf{s}_i & \sum \mathbf{s}^2_i \end{matrix}$$

and $\underline{\boldsymbol{\lambda}}$ is the vector of unknown GLS adjustment factors obtained from $\underline{\lambda} = \mathbf{M}^{-1}\underline{\mathbf{d}}$. Notice that the **M** matrix is based solely on the unweighted sample relationships among schools, teachers and students. This is not an essential feature of our approach; and, indeed, had we used another loss function, a weighted version of the **M** matrix could have been employed.

2.2 Olkin Modified GLS. -- Based on concerns about negative weights raised in our pilot application of GLS (Scheuren and Li 1995), it seemed worthwhile to see if a ratio adjustment step could be introduced before the GLS algorithm was employed. An old idea of Olkin (1958) formed our starting point. Assume we have a total $\boldsymbol{\tau}$, say, of student enrollment in the current application. Suppose further, as is the case here, that this is to be estimated from a sample. Following Olkin we tested a multivariate ratio estimator for $\boldsymbol{\tau}$ of the form $\mathbf{Y} = \mathbf{a}_1 \mathbf{R}_1 \mathbf{w}_. + \mathbf{a}_2 \mathbf{R}_2 \mathbf{t}_. + \mathbf{a}_3 \mathbf{R}_3 \mathbf{s}_.$ where the $\mathbf{a}_i$ are positive and add to 1, where the **w.**, **t.**, and **s.** are sample totals, as before; and the $\mathbf{R}_i$ are the ratios $\mathbf{R}_1 = \mathbf{S/N}$, $\mathbf{R}_2 = \mathbf{S/T}$, and $\mathbf{R}_3 = \mathbf{S/S}$. In our application, the $\mathbf{a}_i$ were simply chosen to be equal to one-third; however, a more natural approach would be to select them so as to minimize the variance of **Y**. Given the complex sample design of SASS, though, this has been left for the future.

In principle, an Olkin adjustment to the original weights could be produced within whatever domain is desired; then in order to determine the "new" weight for that domain, all the cases would be adjusted such that they would have new weights $\mathbf{u}_i = \mathbf{r}\mathbf{w}_i$ where the overall ratio **r** is obtained by taking **Y** and dividing it by the corresponding estimate obtained from the original sample. The intuition is that if the Olkin estimation is first carried out for small (appropriate) subdomains, then there would be a direct benefit from this step in those subdomains; and, among other things, the number of negative weights reduced..

2.3 Results of 1993-94 GLS Reweighting. -- For the nine typologies that make up the private school population separate GLS and Olkin GLS reweighting attempts were made. Sometimes this was straightforward; sometimes extremely difficult. In each case the complex nature of the PSS and SASS sample designs was considered, operational problems were documented, and independent comparisons were made to PSS school size and community type information. Measures of benefit and harm could be developed because of the comparisons possible. Extensive tabular, graphical, and analytic material have been looked at in making the assessments required (Scheuren and Li 1996). Table A at the end of this paper summarizes the results.

Our operational assessment of the Olkin GLS adjustment to SASS was judged to be good to excellent. In only one case, that for the Other Unaffiliated typology was the evidence unclear. We consider this typology "unclear" because the Olkin GLS did not work without a considerable amount of *ad hoc* tinkering.

Based on the independent assessment by community type and school size, the Olkin GLS seemed to do no apparent harm and may have even been of benefit -- beyond the basic consistency achieved with PSS. However, in the independent assessment we never judged the Olkin GLS as "excellent" because, especially by community type, the Olkin GLS was never best overall. Regularly, it did a "good" job, usually by school size, but even here the performance was less than hoped. In three cases, the Olkin GLS was judged only "fair." These were instances where very mixed results were achieved: Some estimates much improved, others quite adversely impacted.

2.4 Discussion. -- While a detailed comparison was also conducted for the Basic GLS, we have omitted any detailed comment on it here, because generally the Basic GLS was inferior to the Olkin GLS.

Frankly, as already noted, our expectations were really not satisfied, even in the Olkin GLS case. Upon reflection, we feel that part of our problem is that the expectations were misplaced. After all, why should introducing just three totals from PSS make a big improvement in SASS. Conversely, why should such a seemingly small change sometimes be so hard?

3 MASS IMPUTATION ALTERNATIVE

Because the positive benefits of the Olkin GLS were often disappointing, we began to question the entire reweighting approach. In order to avoid negative and small weights we had already set aside a few of the largest schools (about a half dozen to a dozen per typology) to be imputed rather than reweighted. We asked ourselves why not do more than just a few? In fact, why not impute the entire SASS file to the PSS in order to take full advantage of the opportunity that having PSS and SASS fielded for the same year offered? In other words, why not do "mass imputation"?

In particular, let us suppose that mass imputation were to be conducted as part of an overall change in SASS estimation. How would it be done? Suppose, for the sake of discussion, that we had PSS and SASS done in the same year. What would the steps be?

Take a specific typology, "Other Religious Unaffiliated" Schools. For the 1993-94 round of SASS, there were 329 schools in the SASS sample with this designation. In the corresponding PSS for the same period, there were 3,193 such schools. The original SASS estimate of students in other religious unaffiliated schools was 462,934. From PSS, the estimate was 37,578 smaller-- at 425,356 students.

An Olkin GLS reweighting approach was taken to this problem to "solve it." However, as noted earlier, we were not satisfied that we had done enough to use the PSS data to improve SASS. If number of students was the major predictive variable, a sensible mass imputation method that could be applied would be to simply impute the SASS records to nearby PSS cases where nearness is defined simply by student enrollment. For parts of the distribution where the SASS sample is sparse, the SASS observation could be used over and over as a donor perhaps up to, say, 1.5 times its original SASS weight.[1] Conversely, in parts of the distribution where there were lots of SASS cases relative to those in the PSS, the SASS cases would be used as donors less often than their original SASS weights would suggest. The SASS observation would always be used at least once, of course, to represent itself.

It may be useful to think of choosing a mass imputation approach after successively imputing SASS to each of the PSS variables separately and looking at how often each SASS observation was used as a donor. If this range of donor use is not too large, then a single, perhaps nearest neighbor, imputation model could work well. Widely discrepant values in terms of donor use would suggest that the imputation is sensitive to one's beliefs as to the predictive power of the variables being used in the imputation. In such settings a case can be made for doing several different imputations that might be made available to the final users for possibly different purposes. It may even make sense simply to use some convex combination of the separate imputations

In Kovar and Whitridge (1995), there is an excellent discussion of mass imputation. Among other things, they comment on the parallels that can exist between weighting and imputation. They call attention to the work of Folsom (1981) in this connection. Evidence that imputation model sensitivity can be a serious problem exists, as they point out -- citing Cox and Cohen (1985), among others.

Difficulties exist in calculating variances and covariances when using mass imputation. A multiple imputation approach to their estimation has been advocated (Rubin 1996) and could be workable since by design the missing data is missing at random.

In another application, by Wong and Ho (1991), bootstrapping was employed successfully. We think a form of bootstrapping might be the best approach for SASS. The presentation by Kaufman (1996), also given at these meetings, presents related work.

4. POSSIBLE NEXT STEPS

In this paper we have brainstormed at length about possible improvements in SASS that could be undertaken -- only one small aspect of which has been summarized in these Proceedings (See Scheuren 1996 for further details). We are not of one mind as to next steps. For one of us, some effort to try out mass imputation seems warranted. Maybe a single typology should be taken and alternative approaches tried out. We both agree that mass imputation could be a lot of fun.

On the other hand, efforts at reweighting SASS have not all been explored and an additional look at some form of poststratification, using GLS or some other calibration estimator makes a lot of sense too, especially since it is unlikely that PSS and SASS will ever be fielded again in the same year.

[1] This factor, about 1.5, is clearly arbitrary and depends on how much of a potential variance price one is willing to pay to get the "nearness" desired. In many weighting settings (e.g., Oh and Scheuren 1987), truncating factors under the square root of 2 work well.

REFERENCES

Burton, R. (1989). Unpublished Memorandum, National Center for Education Statistics.

Colledge, M., Johnson, J., Pare, R. and Sande, I. (1978). "Large Scale Imputation of Survey Data," *Proceedings of the Survey Research Methods Section, American Statistical Association.*

Cox, B. and Cohen, S. (1985). *Methodological Issues for Health Care Surveys*, Marcel Decker: New York.

Deville, J.C., and Särndal, C.E. (1992). "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association,* 87, 376-382.

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

Folsom, R. (1981). "The Equivalence of Generalized Double Sampling Regression Estimators, Weight Adjustments and Randomized Hot Deck Imputations, " *Proceedings of the Survey Research Methods Section, American Statistical Association.*

Imbens, G.W. and Hellerstein, J.K. (1993). "Raking and Regression," *Discussion Paper Number 1658*, Cambridge, MA, Harvard Institute of Economic Research, Harvard University.

Kaufman, S. (1996). *Properties of the Schools and Staffing Survey's Bootstrap Estimator in Nearest Neighbor Matching.* Paper given at the Chicago meetings of the American Statistical Association.

Kovar, J. and Whitridge, P. (1995). "Imputation of Business Survey Data," in Cox, Binder, Chinnappa, Christianson, Colledge, and Kott, eds. *Business Survey Methods*, Wiley: New York.

Oh, H.L. and Scheuren, F. (1987). "Modified Raking Ratio Estimation in the Corporate Statistics of Income," *Survey Methodology*.

Olkin,I. (1958). "Multivariate Ratio Estimation for Finite Populations," *Biometrika*, 45,154-165.

Scheuren, F., and Li, B. (1995). *Intersurvey Consistency in NCES Private School Surveys.* Prepared for U.S. Department of Education, National Center for Education Statistics.

Scheuren, F. and Li, B. (1996). *Intersurvey Consistency in NCES Private School Surveys* for 1993-94. Prepared for U.S. Department of Education, National Center for Education Statistics.

Whitridge, P. Bureau, M. and Kovar, J. (1990). *Mass Imputation at Statistics Canada.* U.S. Bureau of the Census, Sixth Annual Research Conference.

Wong, W. and Ho, C. 1991. "Bootstrapping Post-Stratification and Regression Estimates from a Highly Skewed Distribution," *1991 Proceedings oj the American Statistical Association,* Section on Survey Research Methods.

Table A.-- Olkin GLS Comparisons to Original Weighted SASS Data, By Typology

SASS Typology	Operational Assessment	Independent Assessment
Catholic Parochial	excellent	good
Catholic Diocesan	excellent	fair
Catholic Private	excellent	good
Conservative Christian	good	fair
Other Affiliated	excellent	good
Other Unaffiliated	unclear	good
Non-sectarian Regular	good	good
Non-sectarian Special Emphasis	good	good
Non-sectarian Special Education	good	fair

Notes: The admittedly subjective conventions employed in table A were devised to separate typologies by level of perceived difficulty or benefit. This was done as follows:

(1) Operationally typologies where a simple visual inspection was all that was needed to remove outliers are labelled "excellent" in the operational assessment column.

(2) Typologies labelled "good" operationally were ones where an analytic (potentially iterative) process was required to identify SASS cases that might best be treated by imputation to similar PSS cases rather than being reweighted.

(3) Operationally, only in one case, that of the "Other Unaffiliated" typology was the label "unclear" used. This was done because constructing the Olkin GLS weights proved enormously difficult and required great patience and persistence. (Parenthetically, this typology may, also, have been the most instructive in terms of learning more about how to employ the GLS.)

(4) Based on the independent assessment by community type and school size, the Olkin GLS seemed to do no apparent harm and may have even been of benefit -- beyond the basic consistency achieved with PSS. The comparisons made are to the original SASS weighted data.

(5) The independent assessment column was never coded "excellent" because, especially by community type, the Olkin GLS was never best overall. Regularly, it did a "good" job, usually by school size, but even here the performance was less than hoped. In three cases, the Olkin GLS was judged only "fair." These were instances where very mixed results were achieved: some estimates much improved, others quite negatively impacted.