Journal of Official Statistics, Vol. 19, No. 2, 2003, pp. 81–97

Weighting Methods

Graham Kalton and Ismael Flores-Cervantes[1]

Weighting adjustments are commonly applied in surveys to compensate for nonresponse and noncoverage, and to make weighted sample estimates conform to external values. Recent years have seen theoretical developments and increased use of methods that take account of substantial amounts of auxiliary information in making these adjustments. The article uses a simple example to describe such methods as cell weighting, raking, generalised regression estimation, logistic regression weighting, mixtures of methods, and methods for restricting the range of the resultant adjustments. It also discusses how auxiliary variables may be chosen for use in the adjustments and describes some applications.

Key words: Calibration; generalised regression estimation; poststratification; raking; trimming weights.

1. Introduction

Weights are commonly assigned to respondent records in a survey data file in order to make the weighted records represent the population of inference as closely as possible. The weights are usually developed in a series of stages to compensate for unequal selection probabilities, nonresponse, noncoverage, and sampling fluctuations from known population values (Brick and Kalton 1996).

The first stage of weighting for unequal selection probabilities is generally straightforward. Each sampled element (whether respondent or nonrespondent) is assigned a base weight that is either the inverse of the element's selection probability or proportional to that inverse. With probability sampling, the selection probabilities are known, and the base weights are generally readily determined. A difficulty that sometimes occurs with the base weights arises from sampling frame problems that result in some sampled elements being selected with markedly different probabilities than desired. For example, an element may be classified into the wrong stratum, and hence have a much lower selection probability than intended. If sampled, the element will be assigned a large base weight, which will result in a loss of precision for the survey estimates. To reduce this loss of precision, the base weights for such elements may be trimmed in the same manner as discussed in Section 3 for trimming nonresponse adjusted weights. Base weights will not be addressed further in this article.

[1] Westat, 1650 Research Blvd., Rockville, MD, 20850-3195, U.S.A. Graham Kalton, e-mail: grahamkalton@westat.com
Acknowledgments: We thank the Associate Editor for helpful comments on the earlier version of the article. The article was presented at the Association of Statistical Computing's conference New Methods for Survey Research 1998. See the Association's website www.asc.org.uk for details.

The second stage of weight development is usually to attempt to compensate for unit, or total, nonresponse. The base weights of responding elements are adjusted to compensate for the nonresponding elements. The general strategy is to identify respondents who are similar to the nonrespondents in terms of auxiliary information that is available for both respondents and nonrespondents, and then to increase the base weights of respondents so that they represent similar nonrespondents. In many cases little is known about the nonrespondents (often only their stratum and cluster), in which case a simple cell weighting adjustment may be used. Respondents and nonrespondents are sorted into weighting cells formed from the auxiliary information available, and the weights of the respondents in each cell are increased by a multiplying factor so that the respondents represent the nonrespondents in that cell (see Section 2). This method works well when there is limited auxiliary information available for the nonrespondents. However, when a sizeable amount of auxiliary information is available, and the researcher wants to incorporate much of it in the nonresponse weighting adjustments, then one of the alternative methods described in Section 2 may be needed.

For most surveys with only one round of data collection, nonresponse adjustments are made in a single operation. However, in panel and multiphase surveys, where nonresponse may occur at each successive round of data collection, a sequence of nonresponse adjustments may be employed. In this case, a good deal of auxiliary information is available about later round nonrespondents from their responses at earlier rounds. Complex forms of weighting adjustments may then be needed to take account of a sizeable amount of this information in carrying out the later adjustments. See Section 5 for an example relating to a panel survey, the U.S. Survey of Income and Program Participation (SIPP).

The third stage of weight development involves a further adjustment to the weights to make the resultant weighted estimates from the sample conform to known population values for some key variables. A common form of this type of adjustment forces the sample joint distribution of certain variables (e.g., the sex by age-group distribution) to match the known population joint distribution. This type of adjustment is often termed poststratification. This stage of adjustment serves two purposes: to compensate for noncoverage and to improve the precision of the survey estimates. It can also be used to compensate for nonresponse. It should be noted that the theory for poststratification presented in survey sampling texts assumes full response and perfect coverage. In this situation, the adjustments are generally relatively small provided that the sample sizes in the poststrata are reasonably large, and on average poststratification can be expected to lead to gains in precision for the survey estimates (e.g., Holt and Smith 1979; Kish 1965). However, when there is sizeable noncoverage and/or nonresponse involved, the adjustments can be substantial; in this case the adjustments are used to reduce the bias of the survey estimates, but standard errors for estimates unrelated to the adjustment variables may be increased.

Nonresponse cell weighting adjustments and poststratification adjustments have been widely used for many years. A more recent development has been the increasing use of more complex adjustment methods that can incorporate more auxiliary information than is possible with adjustment cell methods. These newer methods may be used both to compensate for nonresponse, particularly in later rounds of surveys that involve several rounds of data collection, and to force sample distributions to conform to known population distributions, as with poststratification. The methods include raking, generalised regression estimation (GREG), logistic regression modelling, and combinations of weighting cell methods with these methods.

The main purpose of weighting adjustments is to reduce the bias in the survey estimates that nonresponse and noncoverage can cause. A concern with making the adjustments, however, is that they result in increased variability in the weights and thereby lower the precision of the survey estimates. In the case of nonresponse adjustments, a useful measure of this loss of precision is $F = 1 + CV(w_i)^2$, where $CV(w_i)$ is the coefficient of variation of the weights w_i (see, for example, Kish 1992). The measure F represents the multiplying factor that is applied to the variance of a survey estimate due to the variability in the weights in the situation where equal weights are optimal. In this article F relates to the weighting adjustments after the base weights have been applied. Henceforth it will be termed the variance inflation factor. Since the magnitude of this factor is particularly affected by large weighting adjustments, methods have been developed to constrain the adjustments within specified bounds or to trim extreme weights, as discussed later. Such methods serve to improve the precision of the estimates, while accepting some possible loss in bias reduction. It should be noted that the variance inflation factor does not apply for weighting adjustments that benchmark survey estimates to population values; as noted above, absent nonresponse and noncoverage, these adjustments can be expected to increase the precision of survey estimates involving variables related to the auxiliary variables used in the adjustments. However, in this case also, large variability in weighting adjustments due to variable noncoverage can lower the precision of survey estimates unrelated to the auxiliary variables.

This article reviews a number of weighting methods that are currently being applied. Section 2 describes the methods, illustrating their application with a simple artificial example. Section 3 describes methods for constraining the variability of the weights. Section 4 discusses how auxiliary variables may be chosen for use in weighting adjustments when many such variables are available. Section 5 describes some applications in actual surveys. The article ends with some concluding remarks.

2. A Review of Weighting Methods

This section outlines a selection of methods that are used for making weighting adjustments for nonresponse and for making certain weighted sample distributions or estimates conform to distributions or estimates obtained from other sources. The same methods can be applied for each of these two purposes. Nonresponse weighting adjustments are employed to adjust the base weights so that the weighted respondent distributions for certain variables conform to the total sample distributions for these variables. Subsequent adjustments are employed to make the resultant weighted respondent sample distributions or estimates conform to distributions or estimates from an external source, such as population distributions or totals obtained from the sampling frame, from another source (such as population estimates by age and sex), or from a large high-quality survey (such as a labour force survey). Kalton and Kasprzyk (1986) term the first type of adjustments ‘‘sample weighting adjustments’’ and the second type ‘‘population weighting adjustments.’’ In both cases, the adjustments are used to bring the weighted respondent sample data in line with other data. For simplicity, the presentation here will be made in terms of making weighted sample data conform to population data, but it should be borne in mind that this can be readily converted to nonresponse adjustments by interpreting ''population data'' as ''data for the total sample.''

The approach adopted in this section is to briefly describe each adjustment method and the assumptions of its underlying statistical model and to illustrate its application with a simple example. The same example is used throughout so that results can be compared across methods. For ease of presentation, the example employs two auxiliary variables, A and B, in making the weighting adjustments. Since complex weighting adjustments are applicable mostly in situations where several auxiliary variables are to be used in making the adjustments, the example is artificial. Nevertheless, it serves as a useful illustration. For most methods A and B will be treated as categorical variables, with A having four categories (A1, A2, A3, and A4) and B having three categories (B1, B2, and B3). The initially weighted sample joint distribution and the population joint distribution of A and B are given in Table 1 below.

It may be seen from Table 1 that the overall sample and population totals are 1,000 and 1,500. The sample total is constructed to be smaller than the population total because of nonresponse (in the case of nonresponse weighting adjustments) or noncoverage (in the case of adjustments to external controls). A comparison of the numbers in corresponding cells of the sample and population tables shows some substantial differences in the relative magnitudes (for example the ratio of the population total to the sample total in Cell A1/B1 is 4 whereas the ratio in Cell A1/B2 is 1). These differences may be accounted for by differential nonresponse or noncoverage rates.

Six methods will now be described. The weighting adjustments obtained by applying each of the six methods to the data in Table 1 are presented in Table 2, Parts A to F. They are discussed below. Parts G and H of Table 2 give the weighting adjustments for two methods that constrain the range of the adjustments. These and other methods for constraining the variability of the weighting adjustments are discussed in Section 3.

2.1. Cell weighting

The standard cell weighting procedure is to adjust the sample weights so that the sample totals conform to the population totals on a cell-by-cell basis. Thus the weighting adjustment for Cell A1/B1 is $80/20 = 4.00$ and that for Cell A1/B2 is $40/40 = 1.00$. The cell weighting adjustments for sample elements in each of the cells are given in Part A of Table 2.

Table 1. The initially weighted sample and population joint distributions for the auxiliary variables A and B

Sample

	B1	B2	B3	Total
A1	20	40	40	100
A2	50	140	310	500
A3	100	50	50	200
A4	30	100	70	200
Total	200	330	470	1,000

Population

	B1	B2	B3	Total
A1	80	40	55	175
A2	60	150	340	550
A3	170	60	200	430
A4	55	165	125	345
Total	365	415	720	1,500

Table 2. Weighting adjustments obtained from various weighting methods for the data in Table 1

A: Cell weighting

	B1	B2	B3
A1	4.00	1.00	1.38
A2	1.20	1.07	1.10
A3	1.70	1.20	4.00
A4	1.83	1.65	1.79
$F = 1.24$			

B: Raking

	B1	B2	B3
A1	1.81	1.45	2.02
A2	1.08	0.87	1.21
A3	2.20	1.76	2.45
A4	1.83	1.47	2.04
$F = 1.10$			

C: Linear weighting

	B1	B2	B3
A1	1.82	1.50	1.97
A2	1.09	0.78	1.24
A3	2.19	1.88	2.34
A4	1.83	1.52	1.98
$F = 1.10$			

D: GREG weighting

	B1	B2	B3
A1	1.21	1.17	1.14
A2	1.43	1.40	1.36
A3	1.66	1.62	1.59
A4	1.88	1.85	1.81
$F = 1.02$			

E: Logistic regression weighting

	B1	B2	B3
A1	1.87	1.35	1.99
A2	1.11	1.04	1.13
A3	2.23	1.49	2.40
A4	2.05	1.42	2.20
$F = 1.10$			

F: Mixture

	B1	B2	B3
A1	1.71	1.33	2.19
A2	1.32	1.03	1.10
A3	2.12	1.65	2.71
A4	1.75	1.36	2.24
$F = 1.12$			

G: Logit weighting

	B1	B2	B3
A1	1.87	1.46	1.98
A2	1.08	0.74	1.27
A3	2.17	2.09	2.18
A4	1.89	1.49	1.99
$F = 1.10$			

H: Truncated linear weighting

	B1	B2	B3
A1	1.81	1.48	1.99
A2	1.08	0.76	1.26
A3	2.20	2.00	2.20
A4	1.83	1.50	2.00
$F = 1.10$			

The assumption underlying cell weighting adjustments for nonresponse is that the respondents within a cell represent the nonrespondents within that cell. Using a response probability model, this assumption is satisfied if every population element in a cell has the same probability of responding if sampled (Kalton and Maligalig 1991). This corresponds to the missing at random (MAR) assumption of Little and Rubin (2002). A similar model applies if the adjustment is being made for noncoverage. Unlike other methods, cell weighting makes no assumptions about the structure of the response (or coverage) probabilities across cells. Cell weighting has been used for many years to compensate for nonresponse and there is a substantial literature on the method (see, for example, Oh and Scheuren 1983; Kalton 1983; Little 1986; Elliot 1991; Brick and Kalton 1996). Applications in some U.S. surveys are described by Chapman, Bailey, and Kasprzyk (1986).

A potential disadvantage of cell weighting is that it can lead to a large variability in the distribution of the weighting adjustments, thereby inflating the variances of the survey estimates. The cell weights in Table 2 have a four-fold variation from 1.00 in Cell A1/B2 to 4.00 in Cells A1/B1 and A3/B3, leading to a variance inflation factor of $F = 1.24$, i.e., to a reduction in the effective sample size of almost 20 percent. Variance inflation is a particular concern when sample sizes in a number of the adjustment cells are small (as can occur when there are many cells), since small sample sizes give rise to an instability in the adjustments. It is this feature that leads to the use either of weight trimming methods or of alternative adjustment methods.

2.2. *Raking*

Whereas cell weighting forces the sample joint distribution of the auxiliary variables to conform to the population joint distribution, raking operates only on the marginal distributions of the auxiliary variables. Raking is an iterative proportional fitting procedure: first, the sample row totals are forced to conform to the population row totals; then the sample adjusted column totals are forced to conform to population column totals; then the row totals are readjusted to conform; and so on until convergence is reached. Reasonable convergence is usually reached fairly rapidly. However, in some cases convergence can be slow and it is not guaranteed (Ireland and Kullback 1968).

The first round of the raking iteration for the data in Table 1 is given in Table 3. Step 1 makes the sample row totals conform to the population row totals in Table 1. Thus, all the cell entries in Row A1 are multiplied by 175/100, those in Row A2 by 550/500, etc. Step 2 now adjusts the cell entries from Step 1 to make the column totals conform to the population column totals in Table 1. The cell entries in Column B1 are multiplied by 365/356.75, those in Column B2 by 415/504, and those in Column B3 by 720/639.25. As a result of Step 2, the row totals no longer conform to the population row totals. The iteration process is therefore repeated until both row and column totals conform to the population totals within an acceptable margin of error. The final weights obtained from this raking procedure are displayed in Part B of Table 2.

Table 3. The first raking iteration for the data in Table 1

Step 1

	B1	B2	B3	Total
A1	35.00	70.00	70.00	175.00
A2	55.00	154.00	341.00	550.00
A3	215.00	107.50	107.50	430.00
A4	51.75	172.50	120.75	345.00
Total	356.75	504.00	639.25	1,500.00

Step 2

	B1	B2	B3	Total
A1	35.81	57.64	78.84	172.29
A2	56.27	126.81	384.08	567.16
A3	219.97	88.52	121.08	429.57
A4	52.95	142.04	136.00	330.99
Total	365.00	415.00	720.00	1,500.00

When raking is used for nonresponse adjustments, the response probability model being assumed is that the response probabilities are equal for all population elements within cells, as with cell weighting, and also that the response probability for cell (h, k) is of the form $\phi_{hk} = \alpha_h \beta_k$, where α_h and β_k denote row and column effects (Kalton and Maligalig 1991). This additional assumption in the raking model leads to less variability in the weights and hence in the example F is reduced from 1.24 with cell weighting to 1.10 with raking. Note, however, that the additional assumption leads to substantial differences in the weights for many cells from the cell weighting method (see, for example, Cells A1/B1 and A1/B3). The appropriateness of the raking model needs to be carefully considered in such a case.

Raking has been widely used for many years for benchmarking sample distributions to external distributions (see, for example, Deming and Stephan 1940). When benchmarking to population distributions from external sources, sometimes only the marginal distributions of the auxiliary variables are available, in which case raking is possible but cell weighting is not. When the joint distribution is available, the choice between raking and cell weighting involves balancing the smaller loss in precision with raking from reduced variability of the weights against the bias resulting from a failure of the assumed structure of the response probabilities across cells. When the number of cells is relatively small and the cell sample sizes are reasonably large, cell weighting may generally be preferable. However, when there are many cells (as arises from the use of a sizeable number of auxiliary variables) raking may be the better choice.

The raking method described above is also known as the raking ratio method. Sharot (1986) calls it rim weighting. For further discussion of raking, see Oh and Scheuren (1983) and Brackstone and Rao (1979) and, for an application, see Berry, Flatt, and Pierce (1996). It is one form of the generalised raking methods of Deville, Särndal, and Sautory (1993) (see also Deville and Särndal 1992). This class of methods is designed to minimise the weighting adjustments needed to make sample marginal distributions conform to population marginal distributions according to some defined distance function; some of the methods also set upper and lower limits to the adjustments.

2.3. Linear weighting

Linear weighting is another form of generalised raking that is similar to raking except that a different distance function is used (Deville, Särndal, and Sautory 1993). Like raking, it adjusts the weights to make the sample marginal distributions agree with the population marginal distributions. Linear weighting to marginal distributions is a special case of generalised regression estimation (see below). It is treated separately here to bring out its similarity to raking.

The weighting adjustments derived for the example using linear weights are given in Part C of Table 2. The variance inflation factor F is 1.10, the same as for the raking weighting adjustments. The cell weighting adjustments in Part C of Table 2 differ somewhat from those in Part B for raking, but the differences are not large and are much less than those between either of these sets of adjustments and the adjustments with cell weighting.

Linear weighting has the undesirable feature that it can produce negative weights in some situations. As discussed in the next section, the method may be modified to place constraints on the adjustments that avoid this outcome.

2.4. GREG weighting

To this point, the weighting adjustments have been developed for equating joint or marginal distributions with the population distributions. An alternative form of benchmarking is to make weighted sample estimates for quantitative variables conform to population parameters. For example, the estimate of total output from a sample of businesses may be equated to the population total output. To illustrate this approach, assign values 1, 2, 3, and 4 to the classifications A1, A2, A3, and A4, and similarly 1, 2, 3, to B1, B2, B3. The population total for A is then 3,945 and the unadjusted sample total is 2,500; for B the corresponding totals are 3,350 and 2,270. The GREG weighting adjusts the weights so that the weighted sample totals become 3,945 for A and 3,350 for B. The weights needed to achieve these specifications are given in Part D of Table 2.

Since the unadjusted sample mean for A is 2.50 whereas the population mean is 2.63, it is to be expected that the weighting adjustments will be below the average adjustment of 1.50 for low A scores and above that average for high A scores, and this turns out to be the case. The unadjusted sample mean for B is 2.27 and the population mean is 2.24, so that here one might expect the high B values to have slightly lower adjustments, and this also turns out to be the case.

The GREG weighting adjustment derives from the standard regression estimator in survey sampling (see, for example, Cochran 1977, Chapter 7). The method involves incorporating the adjustment for the auxiliary variables used in the regression estimator as a modification of the weights (Bethlehem and Keller 1987; Bethlehem 1988; Deville and Särndal 1992; Fuller, McLoughlin, and Baker 1994; Fuller 2002). As a simple illustration, for an equal probability sample without nonresponse or noncoverage, the regression estimator of the population total Y is

$$\hat{Y}_R = \hat{Y} + b(X - \hat{X})$$

where $\hat{Y} = \sum dy_i$, $\hat{X} = \sum dx_i$, X is the known population total for the auxiliary variable x, b is the sample regression coefficient for the regression of y on x, and d is the inverse of the sampling fraction (the base weight). This estimate may be alternatively expressed as $\hat{Y}_R = \sum w_i y_i$, where the adjusted weight w_i is given by

$$w_i = d + \left[(X - \hat{X})(x_i - \bar{x}) \Big/ \sum (x_i - \bar{x})^2 \right]$$

The GREG weighting approach readily extends to cover several auxiliary variables in the regression model and to incorporate unequal weights. As with standard regression modelling, transformations of variables can be used and interaction terms can be included. Note that, although the GREG weights come from a regression model with a dependent variable y, the weights do not depend on y. They are general weights that are applied for all analyses of the survey data. The use of GREG weighting in Statistics Canada's Generalized Estimation System is described by Estevao, Hidiroglou, and Särndal (1995).

The linear weighting described above is a special case of GREG weighting when the auxiliary variables are categorical variables, or treated as such. Note that by making the sample weighted marginal distributions conform to the population marginal distributions, linear weighting also makes the weighted sample totals conform to the population totals if A and B are discrete variables, as considered in Part D of Table 2. Equating weighted sample totals to population totals imposes far fewer constraints on the weighting adjustments than the linear weighting in Part C. As a result the cell-by-cell variation in the weighting adjustments in Part C is much reduced, and the variance inflation factor is only $F = 1.02$.

2.5. *Logistic regression weighting*

Logistic regression modelling has been used to develop weighting adjustments for nonresponse. A logistic regression model is constructed to predict the probability of responding if sampled based on the auxiliary information, and each respondent's weighting adjustment is then equated to the inverse of the respondent's predicted response probability (see, for example, Iannacchione, Milne, and Folsom 1991; Lepkowski, Kalton, and Kasprzyk 1989). When the auxiliary information is a set of categorical variables and no interactions are included in the logistic regression model, logistic regression weighting is similar to raking. However, unlike raking, it cannot give weighting adjustments of less than 1. Logistic regression weighting is more flexible than raking in that it can include continuous predictors without categorising them. It can also readily include interaction terms as needed to best predict the response probabilities. Interaction terms can be included with raking by, for example, using the cells of cross-classification of two variables as a single marginal control, but the process is less automatic.

Part E of Table 2 gives the weighting adjustments for the data in Table 1 obtained using logistic regression weighting with A and B treated as categorical variables and with no interaction terms in the model. Not surprisingly, these weighting adjustments are seen to be fairly similar to those obtained by raking as presented in Part B of the table. However, the simple use of logistic regression modelling as conducted here does not ensure that the weighted sample marginal distributions conform to the population marginal distributions; Iannacchione, Milne, and Folsom (1991) address that problem.

A variant of this approach is to place the full sample on a continuum in order of their predicted response probabilities from the logistic regression model and to divide the continuum up into a number of cells. The cell weighting method described earlier is then applied with these cells (see, for example, Little and Rubin 2002). In essence, this version involves constructing a single composite variable as a linear combination of the auxiliary variables in the model, and then forming cells based on this composite variable.

2.6. *Mixture of cell weighting and another method*

Cell weighting has the attraction over other methods that it avoids assumptions apart from the missing at random (MAR) assumption that is common to all the methods under review. However, it can produce unstable weighting adjustments, particularly when the sample sizes in some cells are small. A compromise approach is to use cell weighting for cells with large sample sizes and some other weighting method for other cells. Oh and Scheuren (1987), for instance, have used a combination of cell weighting and raking in this fashion, and termed the method modified raking.

To illustrate this general approach, consider using a cell weighting adjustment for the large Cell A2/B3, and raking for the other cells, after removing Cell A2/B3. The results of this procedure are displayed in Part F of Table 2. The weighting adjustment for Cell

A2/B3 is set at 1.10 as in Part A of the table for cell weighting, thus lower than the 1.21 for overall raking in Part B. Other weighting adjustments in the A2 row and B3 column are therefore higher than the raking adjustments in compensation, and this feature, in turn, leads to lower weighting adjustments in the remaining cells.

3. Methods for Constraining Weight Adjustments

The variance inflation factor F is increased greatly by extreme weighting adjustments. To avoid this increase, several alternative procedures have been used to eliminate very large weighting adjustments.

With cell weighting, common procedures are to trim the large weighting adjustments to some specified level or to collapse cells. For example, if a maximum allowable weighting adjustment is specified, cells with adjustments larger than the maximum are assigned the maximum adjustment, with the adjustments of other cells modified to ensure the weighted total is unaltered. Alternatively, these cells are collapsed with other cells such that the weighting adjustments of the collapsed cells are not larger than the specified maximum. Potter (1988, 1990, 1993) reviews procedures for trimming weights and Kalton and Maligalig (1991), Little (1993), Lazzeroni and Little (1998), and Tremblay (1986) discuss procedures for collapsing cells.

With the adjusted weights described in Section 2, the weighted sample total – obtained by summing over the cells the product of the cell weight by the sample number in the cell – equals the population total. For example, using the cell weights in Part A of Table 2 and the sample numbers in Table 1, the weighted sample total is $(4.00 \times 20) + (1.00 \times 40) + \ldots + (1.79 \times 70) = 1{,}500$. However, if the weights for Cells A1/B1 and A3/B3 are trimmed back from 4.00 to, say, 2.50, the contribution to the weighted sample total from these two cells falls from 280 to 175. As a result, the weighted sample total is reduced to 1,395. To bring the weighted population total back up to 1,500, the weighted sample total in other cells may be increased from $(1{,}500 - 280) = 1{,}220$ to $(1{,}500 - 175) = 1{,}325$. This increase may be achieved by increasing the weights in each of these other cells by $1{,}325/1{,}220 = 1.09$. Thus, the trimming procedure leads to a redistribution of the weights, with a part of the population in Cells A1/B1 and A3/B3 being represented by sample elements in other cells.

The alternative procedure of collapsing cells has a similar effect. Consider the problem Cell A1/B1. With collapsing, the cell is combined with an adjacent cell, say A1/B2. The weighting adjustment is then the ratio of the population total (120) to the sample total (60) for the collapsed cell, i.e., an adjustment of 2.00. Thus the effect of this collapsing is that half of the population in Cell A1/B1 is being represented by sample elements in Cell A1/B2.

The techniques of weight trimming and collapsing cells can be applied with any method but there are alternatives that have the advantage of being built into the weighting adjustment process, rather than being modifications made after extreme weights are discovered. Huang and Fuller (1978), Jayasuriya and Valliant (1996), Singh and Mohl (1996), Rao and Singh (1997), and Chen, Sitter, and Wu (2002) describe methods for ensuring nonnegative weights with regression weighting. Deville, Särndal, and Sautory (1993) describe a logit method that corresponds to raking and a truncated linear method that corresponds to the linear method, in both of which the user can specify lower and upper bounds for the weighting adjustments. Both of these methods can be applied using the CALMAR (*cal*ibration on *mar*gins) program, as can unconstrained raking and the linear method. That program is available, together with documentation in French, at the INSEE website http://www.insee.fr/fr/ppp/macro/macro.htm.

The results of applying the logit and truncated linear methods, with lower and upper bounds of 0.5 and 2.20 respectively, to the data in Table 1 are displayed in Parts G and H of Table 2. As can be seen from Parts B and C of the table, only Cell A3/B3 has a weighting adjustment that exceeds 2.20 with raking and linear weighting. Placing the restriction of a maximum weighting adjustment of 2.20 reduces the weighting adjustment for that cell to that amount and distributes the excess weight to other cells, particularly to cells in the A3 row and B3 column.

4. Choice of Auxiliary Variables

When a sizeable number of auxiliary variables is available for use in making weighting adjustments, it may be that only a selection of them can be employed, even when applying the methods described in the previous section. That selection may be guided by analyses of the relationships of the auxiliary variables to the response or coverage rates and to the survey variables.

Consider the bias caused by nonresponse in using the simple unadjusted (base-weighted) sample mean ($\bar{y}$) to estimate the population mean ($\bar{Y}$). That bias can be expressed approximately as

$$B(\bar{y}) \approx \frac{1}{N\bar{\phi}} \sum (Y_i - \bar{Y})(\phi_i - \bar{\phi})$$

where ϕ_i is the response probability of element i ($\phi_i > 0$), $\bar{\phi}$ is the average response probability in the population, and N is the size of the population. As can be seen from this formula, the bias in $\bar{y}$ arises from differential response probabilities that are associated with the y-values. If $\phi_i = \bar{\phi}$ for all population elements, $\bar{y}$ is unbiased for $\bar{Y}$.

To illustrate the effect of weighting adjustments, consider cell weighting. The effect of cell weighting can be seen by expressing the bias of the unadjusted mean as

$$B(\bar{y}) \approx \frac{1}{N\bar{\phi}} \sum \sum (Y_{ci} - \bar{Y}_c)(\phi_{ci} - \bar{\phi}_c) + \frac{1}{N\bar{\phi}} \sum \sum N_c(\bar{Y}_c - \bar{Y})(\bar{\phi}_c - \bar{\phi})$$

where c denotes cell c, Y_{ci} and ϕ_{ci} are the value of y and the response probability for element i in cell c, N_c is the number of elements in the population in cell c, and $\bar{Y}_c$ and $\bar{\phi}_c$ are the mean of y and the average response probability in cell c (Brick and Kalton 1996). The effect of cell weighting is to remove the second term from the bias. Both the average response probabilities and average y-values must differ across cells for this term to be nonzero. Unless these two conditions apply, cell weighting adjustments are ineffective.

The sample mean computed with cell weighting adjustments is unbiased for $\bar{Y}$ if the first term is zero, which occurs when the y-values and response probabilities are unrelated within cells. If the y-values or response probabilities were constant within cells, the adjustment mean would be unbiased. In practice, it is of course unrealistic to expect that weighting adjustments will remove the bias, but it is hoped that they will reduce it. However, if the two terms in the above equation are of opposite sign, it is in fact possible for the adjustments to increase the absolute bias (Thomsen 1973).

To be effective, adjustment cells need to differ in average response probabilities and in y-values. Since surveys generally collect data on many variables, most emphasis is usually placed on constructing cells that differ in response probabilities since it is likely that at least some of the survey variables will differ in level across such cells. Techniques for identifying auxiliary variables to be used in forming adjustment cells with differing response probabilities include logistic regression and tree algorithms (see, for example, Rizzo, Kalton, and Brick 1996).

The logistic regression approach for nonresponse adjustments involves developing a logistic regression model to predict response status (respondent (1) versus nonrespondent (0)) from the set of auxiliary variables known for both respondents and nonrespondents. The model can then be used to predict the response probability for each of the sample members. As described in Section 2, either the respondents' predicted response probabilities can then be used directly as the basis of the weighting adjustments or the predicted response probabilities for all sample members can be used for forming cells for cell weighting adjustments.

The tree algorithm approach involves the use of a branching algorithm such as CHAID, CART, or SEARCH (see, for example, Breiman, Friedman, Olshen, and Stone 1993) to form the nonresponse adjustment cells. The algorithm searches through the auxiliary variables to find the one that can be used to split the full sample into two (or more) parts (cells) in a way that best explains response status. Next, the algorithm performs the same operation separately on each of the two cells. It then continues to do the same on each of the four resulting cells, and so on. The process continues until the splits cease to have useful explanatory power for response status or until the cell sample size reaches a specified minimum level. The final cells of the resulting tree are then treated as standard weighting adjustment cells.

When a survey collects data on only a few variables, or there are only a few variables of key interest, then appropriate statistical models can be constructed for predicting the key variables from the auxiliary variables. A combination of important predictors from these models and from the models predicting response probabilities may then be used for creating the adjustment cells.

With nonresponse adjustments the main consideration in choosing auxiliary variables is generally that they predict response probabilities. With weighting adjustments to make the sample data conform to external sources, however, often the main consideration is that the auxiliary variables predict key survey variables. This is because benchmarking to external sources that are closely related to the key survey variables increases the precision of the survey estimates based on those variables.

5. Two Illustrations

This section briefly outlines the application of complex weighting adjustments in two large-scale U.S. survey programs. One program collects data on food intakes by individuals and the other is a panel survey of income and participation in welfare programs.

5.1. Surveys of food intakes

The U.S. Department of Agriculture's (USDA) 1987–1988 Nationwide Food Consumption Survey collected food use information from sampled households with interviews spread across the seasons of the year. Since only 37 percent of occupied housing units provided the requisite data, there was a serious concern about nonresponse bias. To address these concerns extensive benchmarking to external controls was carried out by regression weighting (Fuller, McLoughlin, and Baker 1994). Fifteen auxiliary variables that were believed to be related to food behaviour – including such variables as household income as a percentage of the poverty level, household size, presence of a child under 7 years of age, age of household head, and geographic region – were used. All but two of the external controls came from the March 1987 U.S. Current Population Survey (CPS). Controls for urbanisation came from another source and control for season was an even distribution over the four seasons. All the variables except household size were treated as categorical variables. In addition to the 25 indicator variables for the categorical variables, household size and the square of household size were included in the regression. The regression weights were constrained to fall within defined limits, limiting the variance inflation factor to 1.32.

The USDA's Continuing Survey of Food Intakes by Individuals (CSFII), conducted in 1994–1996, achieved a much higher response rate than the 1987–1988 survey, but it also used extensive benchmarking to external controls, primarily taken from the CPS. In the CSFII, nonresponse adjustments were made in two stages. The first stage involved adjustments for nonresponses to a screener interview and were based on characteristics of the segment (final stage cluster) in which they fell. The second stage adjustment compensated for nonresponse to the interview about the first day's food intake. Cell weighting adjustments were made, with cells identified by a CHAID analysis and with auxiliary variables selected from the segment characteristics and screener responses (e.g., age, sex, income status). The final adjustment was raking to CPS data and to equal distributions by season and day of the week. The 16 raking dimensions were similar to those used for the 1987–1988 survey, with household size being treated as a categorical variable. Raking was conducted separately for four subgroups (males, 20 years of age or older; females, 20 years of age or older; children 0–5 years of age; persons 6–19 years of age), thereby covering some interaction effects. Further interactions were covered by crossclassifying auxiliary variables (e.g., age group and home ownership for those 20 years of age or older; age group and sex for those under 20 years of age) and treating the cells of the crossclassification as a single control variable. Chu and Goldman (1997) provide further details.

5.2. Survey of Income and Program Participation

The Survey of Income and Program Participation (SIPP) is a repeated household panel survey conducted by the U.S. Census Bureau. New panels were started each year from 1984 to 1993 and a revised panel design was introduced in 1996. The current discussion relates to the 1987 SIPP panel. That panel started with a sample of about 12,300 households, and followed the members of those households for seven waves of interviews, conducted at four-month intervals. The panel weights produced for the 1987 panel relate only to respondents to all of the seven waves for which they were eligible. Nonresponse adjustments are made for the 6.7 percent of nonresponding households at the initial wave and for the 20.8 percent of panel nonrespondents, that is, respondents at the initial wave who failed to respond for one or more of the subsequent waves for which they were eligible.

The weighting adjustments for panel nonrespondents can employ the responses to the initial wave interviews as auxiliary variables. Several approaches have been examined for developing these weighting adjustments. Rizzo, Kalton, and Brick (1996) conducted logistic regression analyses to determine which of 58 initial wave items to retain, and thereby reduced the number of items to 31. They examined the use of several alternative weighting adjustments including cell weighting with cells obtained from a CHAID analysis, logistic regression weighting, raking, and a mixture of cell weighting for cells with large sample sizes and logistic regression weighting for other cells. They found a fair degree of similarity between the weights produced with these different methods. However, all these sets of weights were somewhat different from the current weights, which were based on a different set of auxiliary variables.

Folsom and Witt (1994) examined the use of a constrained logistic regression model to compensate for panel nonresponse in the 1987 SIPP panel. Their analysis suggested that the weight adjustments should be developed separately for seven subgroups of the population defined in terms of household income, race/ethnicity, marital status and census region. Sizeable numbers of predictor variables were included in most models.

An, Breidt, and Fuller (1994) employed 97 CPS and 79 initial wave indicator variables in developing regression weighting adjustments for 1987 SIPP panel respondents. The adjustments were made in a sequence of steps. Three alternative weighting schemes involving different ordering to the use of the CPS control totals were applied, and the standard errors of a selection of estimates using the weights obtained from the three schemes were compared.

6. Concluding Remarks

Recent developments for complex weighting adjustments are now being increasingly applied. By permitting more auxiliary information to be used, they have the potential to reduce biases arising from nonresponse and noncoverage. Benchmarking to external data both addresses noncoverage bias and has a poststratification effect that works towards improving the precision of some survey estimates. It needs to be recognised, however, that when noncoverage is large and variable across the population, the overall effect of benchmarking may well be to lower the precision of some survey estimates. The use of methods that constrain the variability of the weighting adjustments or trim large weights can avoid serious losses in precision.

In general these complex methods are most useful when there is a sizeable amount of auxiliary information available. With sample adjustment for nonresponse, this situation mostly occurs in later rounds of panel surveys and surveys with several phases of data collection. With population adjustments for noncoverage and improved precision, it occurs when the survey data are benchmarked to another large-scale survey, and when a great deal is known about the population (as in some establishment surveys). When a substantial amount of auxiliary information is available, a variety of alternative methods may be used. Many of the methods are in fact fairly similar and the weighting adjustments they produce are likely to be highly correlated (Deville, Särndal, and Sautory 1993; Rizzo, Kalton, and Brick 1996). Thus the choice of auxiliary variables and of the mode in which they are employed in the adjustments may be of more significance than the choice of a particular method.

In applying a complex weighting adjustment procedure, attention should be paid to the assumptions of its underlying statistical model. Most of the models control to marginal data, using some type of ''no interaction'' assumption. This assumption needs to be carefully assessed, and modifications made if it is obviously erroneous (for example, by combining auxiliary variables into a single variable with raking, performing the adjustments separately for different subsets of the population, or including interaction terms in regression models). Similarly, methods that automatically restrict the range of the adjustments are redistributing the excess adjustments that would otherwise be given to some respondents to other respondents. The appropriateness of this redistribution should be examined. Although a complex method may be readily applied using available computer software, the need remains for a careful review of the method's applicability for a given survey.

Finally, it should be noted that survey weights that include weighting adjustments for nonresponse, for noncoverage, and for poststratification-type gains in precision should not be treated as constants when estimating the variances of survey estimates. One approach that deals with this complication is to use a jackknife variance estimation procedure with the weighting adjustments recomputed for each replicate. Another is to employ a Taylor series linearization procedure that takes account of the weight variation. See, for example, Stukel, Hidiroglou, and Särndal (1996).

7. References

An, A.B., Breidt, F.J., and Fuller, W.A. (1994). Regression Weighting Methods for SIPP Data. Proceedings of the American Statistical Association, Section on Survey Research Methods, 434–439.

Berry, C.C., Flatt, S.W., and Pierce, J.P. (1996). Correcting Unit Nonresponse via Response Modeling and Raking in the California Tobacco Survey. Journal of Official Statistics, 12, 349–363.

Bethlehem, J.G. (1988). Reduction of Nonresponse Bias through Regression Estimation. Journal of Official Statistics, 4, 251–260.

Bethlehem, J.G. and Keller, W.J. (1987). Linear Weighting of Sample Survey Data. Journal of Official Statistics, 3, 141–153.

Brackstone, G.J. and Rao, J.N.K. (1979). An Investigation of Raking Ratio Estimation. Sankhyā, Series C, 41, 97–114.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1993). Classification and Regression Trees. New York: Chapman and Hall.

Brick, J.M. and Kalton, G. (1996). Handling Missing Data in Survey Research. Statistical Methods in Medical Research, 5, 215–238.

Chapman, D.W., Bailey, L., and Kasprzyk, D. (1986). Nonresponse Adjustment Procedures at the U.S. Bureau of the Census. Survey Methodology, 12, 161–180.

Chen, J., Sitter, R.R., and Wu, C. (2002). Using Empirical Likelihood Methods to Obtain

Range Restricted Weights in Regression Estimators for Surveys. Biometrika, 89, 230–237.
Chu, A. and Goldman, J. (1997). Weighting Procedures for USDAs Continuing Survey of Food Intakes by Individuals 1994–1996. Proceedings of the American Statistical Association, Section on Survey Research Methods, 802–807.
Cochran, W.G. (1977). Sampling Techniques (3rd ed.). New York: Wiley.
Deming, W.E. and Stephan, F.F. (1940). On a Least Squares Adjustment of a Sample Frequency Table When the Expected Marginal Totals Are Known. Annals of Mathematical Statistics, 11, 427–444.
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.
Elliot, D. (1991). Weighting for Non-Response. A Survey Researcher's Guide. London: Office of Population Censuses and Surveys.
Estevao, V., Hidiroglou, M.A., and Särndal, C.-E. (1995). Methodological Principles for a Generalized Estimation System at Statistics Canada. Journal of Official Statistics, 11, 181–204.
Folsom, R.E. and Witt, M.B. (1994). Testing a New Attrition Nonresponse Adjustment Method for SIPP. Proceedings of the American Statistical Association, Section on Survey Research Methods, 428–433.
Fuller, W.A. (2002). Regression Estimation for Survey Samples. Survey Methodology, 28, 5–23.
Fuller, W.A., McLoughlin, M.M., and Baker, H.D. (1994). Regression Weighting in the Presence of Nonresponse With Application to the 1987–1988 Nationwide Food Consumption Survey. Survey Methodology, 20, 75–85.
Holt, D. and Smith, T.M.F. (1979). Post Stratification. Journal of the Royal Statistical Society, Series A, 142, 33–46.
Huang, E.T. and Fuller, W.A. (1978). Nonnegative Regression Estimation for Survey Data. Proceedings of the American Statistical Association, Social Statistics Section, 300–303.
Iannacchione, V.G., Milne, J.G., and Folsom, R.E. (1991). Response Probability Weight Adjustments Using Logistic Regression. Proceedings of the American Statistical Association, Section on Survey Research Methods, 637–642.
Ireland, C.T. and Kullback, S. (1968). Contingency Tables With Given Marginals. Biometrika, 55, 179–188.
Jayasuriya, B.R. and Valliant, R. (1996). An Application of Restricted Regression Estimation in a Household Survey. Survey Methodology, 22, 127–137.
Kalton, G. (1983). Compensating for Missing Survey Data. Ann Arbor: Institute for Social Research, University of Michigan.
Kalton, G. and Kasprzyk, D. (1986). The Treatment of Missing Survey Data. Survey Methodology, 12, 1–16.
Kalton, G. and Maligalig, D.S. (1991). A Comparison of Methods of Weighting Adjustment for Nonresponse. Proceedings of the U.S. Bureau of the Census Annual Research Conference, 409–428.
Kish, L. (1965). Survey Sampling. New York: Wiley.

Kish, L. (1992). Weighting for Unequal P_i. Journal of Official Statistics, 8, 183–200.
Lazzeroni, L.C. and Little, R.J.A. (1998). Random-Effects Models for Smoothing Poststratification Weights. Journal of Official Statistics, 14, 61–78.
Lepkowski, J., Kalton, G., and Kasprzyk, D. (1989). Weighting Adjustments for Partial Nonresponse in the 1984 SIPP Panel. Proceedings of the American Statistical Association, Section on Survey Research Methods, 296–301.
Little, R.J.A. (1986). Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review, 54, 139–157.
Little, R.J.A. (1993). Post-Stratification: A Modeler's Perspective. Journal of the American Statistical Association, 88, 1001–1012.
Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis With Missing Data (2nd ed.). New York: Wiley.
Oh, H.L. and Scheuren, F. (1983). Weighting Adjustments for Unit Nonresponse. In Incomplete Data in Sample Surveys, Volume 2: Theory and Bibliographies, W.G. Madow, I. Olkin, and D. Rubin (eds). New York: Academic Press.
Oh, H.L. and Scheuren, F. (1987). Modified Raking Ratio Estimation. Survey Methodology, 13, 209–219.
Potter, F.J. (1988). Survey of Procedures to Control Extreme Sampling Weights. Proceedings of the American Statistical Association, Section on Survey Research Methods, 453–458.
Potter, F.J. (1990). A Study of Procedures to Identify and Trim Extreme Sampling Weights. Proceedings of the American Statistical Association, Section on Survey Research Methods, 225–230.
Potter, F.J. (1993). The Effect of Weight Trimming on Nonlinear Survey Estimates. Proceedings of the American Statistical Association, Section on Survey Research Methods, 758–763.
Rao, J.N.K. and Singh, A.C. (1997). A Ridge-Shrinkage Method for Range-Restricted Weight Calibration in Survey Sampling. Proceedings of the American Statistical Association, Section on Survey Research Methods, 57–65.
Rizzo, L., Kalton, G., and Brick, J.M. (1996). A Comparison of Some Weighting Adjustment Methods for Panel Nonresponse. Survey Methodology, 22, 43–53.
Sharot, T. (1986). Weighting Survey Results. Journal of the Market Research Society, 28, 269–284.
Singh, A.C. and Mohl, C.A. (1996). Understanding Calibration Estimators in Survey Sampling. Survey Methodology, 22, 107–115.
Stukel, D.M., Hidiroglou, M.A., and Särndal, C.-E. (1996). Variance Estimation for Calibration Estimators: A Comparison of Jackknifing Versus Taylor Linearization. Survey Methodology, 22, 117–125.
Thomsen, I. (1973). A Note on the Efficiency of Weighting Subclass Means to Reduce the Effects of Nonresponse When Analyzing Survey Data. Statistisk tidskrift, 4, 278–283.
Tremblay, V. (1986). Practical Criteria for Definition of Weighting Classes. Survey Methodology, 12, 85–97.

Received April 2002
Revised December 2002