SURVEY NONRESPONSE ADJUSTMENTS

Roderick J. A. Little, University of California, Los Angeles

1. INTRODUCTION

Three approaches to the analysis of incomplete data from sample surveys can be usefully distinguished. A theoretically appealing strategy is to accept the nonrectangular structure of data subject to missing values, and to estimate parameters by methods based on a model for the incomplete data. See, for example, Little (1982). However, the practical problems of building sufficiently realistic models for large data sets with complex sample designs are not negligible. Furthermore, the availability of special software to fit such models to incomplete data is not extensive, and is limited to fairly specific models. The survey processor faced with incomplete data has a strong motivation to create rectangular files for statistical analyses or for public use tapes. Thus alternative procedures that yield rectangular data sets will continue to be important in practice. Implicit or explicit models also underpin these alternative strategies (Rubin, 1978; Oh and Scheuren, 1983), so these procedures are not incompatible with a modeling philosophy.

Two strategies leading to rectangular data sets are common in survey practice, namely weighting and imputation. In the former approach missing or incomplete units in the sample are ignored, and the sampling weights for responding units are inflated by dividing them by estimates of the probability of response, typically the response rate in a subclass of the sample. In imputation approaches incomplete or missing units are included in the sample, with missing values replaced by the imputed values. For a recent review of imputation methods in surveys, see Kalton and Kasprzyk (1982).

In large government surveys weighting is often used to handle unit response, which arises when whole questionnaires are missed because of noncontact or refusal. Here the weighting nonresponse adjustment is a natural extension of the sampling weight defined for sampled units. For example, in the Current Population Survey (CPS), unit nonresponse is handled by dividing the sample weights of respondents by the response rate within homogeneous collections of primary sampling units (Hanson, 1978). Imputation is often used to handle item nonresponse, where particular items in the interview are missing. For example, missing income items in the Income Supplement of the CPS are imputed by a flexible hot deck matching scheme (Welniak and Coder, 1980; Oh and Scheuren, 1980).

A common preliminary to weighting or imputation is to classify respondents and nonrespondents into adjustment cells. Consider the artificial example in Table 1, where 200 out of 300 sampled individuals respond to a question on y = annual income. Respondents and nonrespondents are classified into three adjustment cells defined by the variable A = Region. Within each cell, the response rates are 80/100, 70/100 and 50/100 and the respondent mean incomes (in $1,000) are 9.8, 11.6 and 13.6, respectively. Suppose that all individuals in the population have an equal chance of selection, so that sample weights are not required.

A simple estimate of the mean income in the population is the respondent mean $\bar{y}_R = 11.4$. However the table suggests that nonresponse is higher in the high income region (C=3) than in the low income region (C=1), and hence $\bar{y}_R$ may be an underestimate. Weighting the income contributions in each cell by the inverse of the response rate in the cell yields an adjusted mean $\bar{y}_A = 11.7$ that plausibly reduces the bias from restriction to the respondent sample. It is well known that the same adjusted mean ($\bar{y}_A$) can be obtained by imputing the cell respondent mean for all the nonrespondents in that cell (for example, 11.6 for all 30 nonrespondents in cell 2).

These two basic adjustment methods - weighting by reciprocal cell response rates and imputing cell means - are the focus of this article, although other methods are discussed for comparative purposes. Note that mean imputation has the disadvantage of distorting the distribution of income values in each cell (Kalton and Kasprzyk, 1982). Hence hot deck versions of the method where values from individual respondents in the same cell are imputed rather than cell means, are popular in practice. These modifications increase the variance of population estimates by an amount that depends on the method used to assign respondent values to nonrespondents (Ernst, 1980; Kalton and Kish, 1981). We do not consider hot deck methods here, although results about the large sample bias of cell mean imputation also apply to hot deck methods that impute values that average the cell mean in hypothetical repetitions.

In section 2 we consider mean squared error properties of $\bar{y}_R$, $\bar{y}_A$ and a modification of $\bar{y}_A$ that assumes the population distribution over the adjustment cells is known. Thomsen (1973, 1978) and Oh and Scheuren (1983) perform similar calculations, but consider a single predetermined choice of adjustment cells. Our focus is on how to choose adjustment cells when a large amount of information is available to form them. Two key dimensions in the space of potential stratifiers are distinguished, the response propensity $\hat{p}(x)$ and the predicted mean $\hat{y}(x)$. Stratifying on the former dimension controls large sample bias, and

Table 1. Adjustment Cell Estimator for an Overall Mean: Example

STATISTIC	A = Adjustment Cell (Region)		
	C=1	C=2	C=3
RESPONSE RATE	80/100	70/100	50/100
MEAN INCOME ($1,000)	9.8	11.6	13.6
TOTAL INCOME ($1,000)	780	815	680

Respondent Mean : $\bar{y}_R = (780 + 815 + 680)/(80 + 70 + 50) = 11.4$

Adjusted Mean : $\bar{y}_A = \dfrac{780(\frac{100}{80}) + 815(\frac{100}{70}) + 680(\frac{100}{50})}{100 + 100 + 100} = 11.7$

Table 2. Adjustment Cell Estimators for a Crossclass Mean: Example

RESPONSE RATE / MEAN INCOME (\$1,000) / TOTAL INCOME (\$1,000)		A = Adjustment Cell (Region) C=1	C=2	C=3
Z = Education	High	20/30	15/30	10/30
		12.0	14.0	16.0
		240	210	160
	Low	60/70	55/70	40/70
		9.0	11.0	13.0
		540	605	520
	Total	80/100	70/100	50/100
		9.8	11.6	13.6
		780	815	680

Respondent Mean : $\bar{y}_{jR} = (240 + 210 + 160)/(20 + 15 + 10) = 13.56$

Weighted in cell c: $\bar{y}^{(2)}_{jA} = \dfrac{240(100/80) + 210(100/70) + 160(100/50)}{20(100/80) + 15(100/70) + 10(100/50)} = 13.85$

Imputed in cell c : $\bar{y}^{(3)}_{jA} = \dfrac{240 + 10(9.8)+210+15(11.6)+160+20(13.6)}{30 + 30 + 30} = 12.82$

stratification on the latter dimension controls both bias and variance, but (unlike $\hat{p}(x)$) requires separate models and nonresponse adjustments for each y variable.

Adjustment cell weighting and mean imputation yield the same estimator $\bar{y}_A$ for population means. They also yield the same estimators for means or totals in domains of the population, which we define as collections of adjustment cells. However, weighting and imputation yield different estimators of means or totals in crossclasses of the population, which we define as subclasses of the population that cut across adjustment cells (David, Little, Samuhel and Triest, 1983). Consider, for example, Table 2, where the data in Table 1 are further classified by the crossclass variable Z = Education. Suppose the objective is to estimate the mean income in the high education group. As shown in the calculations below the table, the respondent mean is $\bar{y}_{jR} = 13.56$. Weighting the respondent income amounts by the inverse of the adjustment cell response rates yields the estimator $\bar{y}^{(2)}_{jA} = 13.85$, whereas imputing adjustment cell means yields $\bar{y}^{(3)}_{jA} = 12.82$, a value even lower than the unadjusted mean $\bar{y}_{jR}$.

In section 3 formulae are presented for the bias and variance of $\bar{y}_{jR}$, $\bar{y}^{(2)}_{jA}$, $\bar{y}^{(3)}_{jA}$ and some natural alternative estimators. We show that different choices of adjustment cells are appropriate for weighting and imputation when the objective is to control the bias of crossclass mean estimates. Specifically, when weighting adjustments are employed, adjustment cells based on the estimated **response propensity** $\hat{p}(x)$ are appropriate, whereas adjustment cells based on the predicted mean $\hat{y}(x)$ may yield biased estimates. On the other hand, when imputation is employed, adjustment cells based on $\hat{y}(x)$ are appropriate, whereas adjustment cells based on $\hat{p}(x)$ may yield biased estimates. This theory establishes links between weighting and response propensity stratification and between imputation and predictive mean stratification.

Variance properties of crossclass mean estimators are also examined in section 3. Mean squared errors properties of the estimators of domain and crossclass means in sections 2 and 3 are then explored by a simulation study, described in section 4. Finally, section 5 indicates some modifications of response propensity methods to control variance whilst retaining their bias reducing properties.

2. ADJUSTMENT CELL ESTIMATOR OF DOMAIN MEANS

2.1 Moments of Adjustment Cell Estimators in Repeated Sampling

In this section we discuss adjustment cell estimators of the population mean $\bar{Y}$ of a variable y. The theory also applies to estimators of domain means, where a domain consists of a subset of the adjustment cells. Notation is defined in Table 3 A). Population and sample quantities have upper and lower case letters, respectively. The symbols N, $\bar{Y}$, B and P are used for population counts, means, response rates and cell proportions, respectively. The suffix c refers to adjustment cell c, and the suffix R denotes restriction to respondents.

Three estimators of $\bar{Y}$ are compared. The respondent mean,

$$\bar{y}_R = \sum_{c=1}^{C} p_{cR}\, \bar{y}_{cR} \quad , \tag{1}$$

is obtained when a constant weighting adjustment b_+^{-1} is applied to respondents, or when $\bar{y}_R$ is imputed for all nonrespondents. The adjusted mean

$$\bar{y}_A = \sum_{c=1}^{C} p_c\, \bar{y}_{cR} \; , \tag{2}$$

is obtained by weighting respondents in cell c by

Table 3. Notation for Population and Sample Quantities

A) For Inferences about $\bar{Y}$.

Population	Counts: Cell c	Counts: Overall	Ratio	Means: Cell c	Means: Overall	Variances: Cell c
Respondents (R)	N_{cR}	N_{+R}	P_{cR}	$\bar{Y}_{cR}$	$\bar{Y}_R$	S^2_{cR}
Respondents + Nonrespondents (R+N)	N_c	N_+	P_c	$\bar{Y}_c$	$\bar{Y}$	
Ratio (R/(R+N))	B_c	B_+				

Sample	Counts: Cell	Counts: Overall	Ratio	Means: Cell	Means: Overall	Variances: Cell c
Respondents (R)	n_{cR}	n_{+R}	p_{cR}	$\bar{y}_{cR}$	$\bar{y}_R$	s^2_{cR}
Respondents + Nonrespondents (R+N)	n_c	n_+	p_c	$\bar{y}_c$	$\bar{y}$	
Ratio (R/(R+N))	b_c	b_+				

B) For Inferences about the Crossclass Mean, $\bar{Y}_{+j}$.

Population	Counts: Cell c	Counts: Overall	Ratio	Means: Cell c	Means: Overall	Variances: Cell c
R	N_{cjR}	N_{+jR}	P_{cjR}	$\bar{Y}_{cjR}$	$\bar{Y}_{+jR}$	S^2_{cjR}
R+N	N_{cj}	N_{+j}	P_{cj}	$\bar{Y}_{cj}$	$\bar{Y}_{+j}$	
R/(R+N)	B_{cj}	B_{+j}				

Sample	Counts: Cell c	Counts: Overall	Ratio	Means: Cell c	Means: Overall	Variances: Cell c
R	n_{cjR}	n_{+jR}	p_{cjR}	$\bar{y}_{cjR}$	$\bar{y}_{+jR}$	s^2_{cjR}
R+N	n_{cj}	n_{+j}	p_{cj}	$\bar{y}_{cj}$	$\bar{y}_{+j}$	
R/(R+N)	b_{cj}	b_{+j}				

b_c^{-1}, or by imputing $\bar{y}_{cR}$ for nonrespondents in cell c. Finally, the poststratified estimator

$$\bar{y}_S = \sum_{c=1}^{C} P_c \bar{y}_{cR} , \quad (3)$$

can be calculated when the population cell proportions $\{P_c\}$ are known. For example, adjustment cells could be based on age, race and sex, and $\{P_c\}$ may be available from census tables. The bias and variance of $\bar{y}_R$, $\bar{y}_A$ and $\bar{y}_S$ are considered by Thomsen (1973, 1978) and by Oh and Scheuren (1983), but comparisons are complicated by different assumptions and choices of reference distribution. Specifically, let $\underset{\sim}{y}=(y_1,\ldots,y_N)$ denote the vector of y values in the population; $\underset{\sim}{r}=(r_1,\ldots,r_N)$ denote the vector of response indicators, such that $r_i=1$ if unit i responds if sampled, and 0 otherwise; $\underset{\sim}{s}=(s_1,\ldots,s_N)$ denote the vector of sample indicators, such that $s_i=1$ if unit i is sampled, and 0 otherwise; $\underset{\sim}{n}=(n_1,\ldots,n_C)$ denote the vector of sample sizes in the C adjustment cells; and $\underset{\sim}{n}_R=(n_{1R},\ldots,n_{CR})$ denote the vector of respondent sample sizes in the C adjustment cells. Thomsen calculates the bias and variance of $\bar{y}_R$, $\bar{y}_A$ and $\bar{y}_S$ over the distribution of $\underset{\sim}{s}$, with $\underset{\sim}{y}$ and $\underset{\sim}{r}$ held fixed. Oh and Scheuren calculate moments over the distribution of $\underset{\sim}{r}$ and $\underset{\sim}{s}$, with i) $\underset{\sim}{y}$ held fixed and ii) $\underset{\sim}{y}$, $\underset{\sim}{n}$ and $\underset{\sim}{n}_R$ held fixed. They call the former unconditional moments and the latter conditional moments. This approach requires specification of the distribution of $\underset{\sim}{r}$; Oh and Scheuren assume the distribution corresponding to Bernoulli subsampling within the adjustment cells, an assumption they describe as <u>quasi-randomization</u>. They also include finite population corrections (fpc's), assuming simple random sampling without replacement; Thomsen ignores these corrections.

We prefer the calculations of Oh and Scheuren that condition on $\underset{\sim}{n}$ and $\underset{\sim}{n}_R$ as well as $\underset{\sim}{y}$, since they provide more precise results when the respondent sample sizes $\{\underset{\sim}{n}_R\}$ are small. The situation is analogous to that of Holt and Smith (1979), who condition on $\underset{\sim}{n}$ when comparing unstratified and poststratified estimates of a mean from complete data. On the other hand, we prefer (like Thomsen) to calculate moments conditional on $\underset{\sim}{r}$, since the validity of the quasi-randomization assumption of Oh and Scheuren is specific to a particular choice of adjustment cells, and we wish to consider a variety of choices. Hence we present in Table 4 moments conditional on $\underset{\sim}{y}$, $\underset{\sim}{r}$, $\underset{\sim}{n}$ and $\underset{\sim}{n}_R$. Expressions for bias assume an equal probability sampling design. Expressions for variance apply assuming simple random sampling with replacement. For simple random sampling without replacement the variances are modified by multiplying S_{cR}^2 by the fpc $(1-n_{cR}/N_c)$, yielding similar results to the conditional calculations of Oh and Scheuren (1983).

Note that $\bar{y}_A$ is undefined if, for one of the adjustment cells, $n_c>0$ and $n_{cR}=0$. In calculations that condition on $\underset{\sim}{n}$ and $\underset{\sim}{n}_R$, we assume that this event has not occurred. In Thomsen's unconditional calculations, the assumption is made that this event has negligible probability of occurrence.

Each of the bias components in Table 4 is written as the sum of two terms, say C and LSB. As the sample size increases, the proportions p_{cR} and p_c converge to their population analogs P_{cR} and P_c respectively, so the first terms C in these expressions tend to zero (for $\bar{y}_S$, C is identically equal to zero). We call the second terms LSB in the bias expressions <u>large sample biases</u>, since they increasingly dominate the bias as the sample size increases. The mean squared error of each estimator can be decomposed as

$$\text{mse}=(C+LSB)^2+V=C^2+2(C)(LSB)+LSB^2+V ,$$

where V is the variance. In section 2.2 we discuss the formation of adjustment cells to minimize the large sample squared bias LSB^2 of $\bar{y}_A$ and $\bar{y}_S$. In section 2.3 we consider the formation of adjustment cells to limit the size of $LSB^2 + V$, which we call the conditional mean squared error increment (ΔMSE). The cross product term 2(C)(LSB) can have either sign and is generally small in magnitude.

2.2 <u>Choosing Adjustment Cells to Control Large Sample Bias</u>

The respondent mean $\bar{y}_R$ has zero large sample bias if $\bar{Y}_R = \bar{Y}$, that is if the mean of y is the same for respondents and nonrespondents. More generally, if interest concerns the entire distribution of Y rather than simply the mean, unadjusted inferences based on the respondent sample require that y and response are independent, that is

$$y \perp\!\!\!\perp r ,$$

where $\perp\!\!\!\perp$ is Dawid's (1979) notation for independence. This assumption is usually unrealistic.

The large sample bias of the adjustment cell estimators $\bar{y}_A$ and $\bar{y}_S$ is

$$LSB = \sum_{c=1}^{C} P_c (\bar{Y}_{cR} - \bar{Y}_c) ,$$

which equals zero if $\bar{Y}_{cR} = \bar{Y}_c$ for all c. More generally, we seek adjustment cells within which the distribution of y is the same for respondents and nonrespondents. Let c denote the level of the adjustment cell variable A. Then we seek A such that y is conditionally independent of the response indicator r, given A. That is

$$y \perp\!\!\!\perp r|A . \quad (4)$$

Now suppose we have a large set of potential stratifiers x, recorded for respondents and nonrespondents in the sample, and that

$$y \perp\!\!\!\perp r|x , \quad (5)$$

Table 4. Bias and Variance of Three Estimates of $\bar{Y}$ in Repeated Sampling*

Estimator	Bias	Variance
$\bar{y}_R$	$\sum_c (p_{cR} - P_{cR})\bar{Y}_{cR} + (\bar{Y}_R - \bar{Y})$	$\sum_c p_{cR} S_{cR}^2/n_{+R}$
$\bar{y}_A$	$\sum_c (p_c - P_c)\bar{Y}_{cR} + \sum_c P_c(\bar{Y}_{cR} - \bar{Y}_c)$	$\sum_c p_c^2 S_{cR}^2/n_{cR}$
$\bar{y}_S$	$0 + \sum_c P_c(\bar{Y}_{cR} - \bar{Y}_c)$	$\sum_c P_c^2 S_{cR}^2/n_{cR}$

* See text for details of reference distribution.

so that a full stratification on x removes the nonresponse bias. In practice, adjustment cells cannot be based on x, because some variables are interval scaled, or the joint distribution of x contains cells with sampled units, none of which respond. The question is how to define A based on a restricted subset of the x information, so that (4) is approximately satisfied. Two approaches to this question can be distinguished.

The first approach is to model the distribution of y given x. Let D(x) be the distribution of population y values for respondents (and by (5), for nonrespondents) with value x of the covariates. Pooling over values of x such that D(x) is constant clearly leads to subpopulations within which y and r are still independent. More specifically, suppose we specify the model that D(x) and D(x'), the distributions of y for different values x and x' of the covariates, differ only in their location parameters, viz the population means $\bar{Y}(x)$ and $\bar{Y}(x')$. Then forming adjustment cells with constant values of $\bar{Y}(x)$ yields a variable A for which (4) is satisfied. If A is formed so that $\bar{Y}(x)$ is constant within adjustment cells, then condition (4) is satisfied and the large sample bias eliminated.

In practice the means $\bar{Y}(x)$ have to be estimated from the data. Let $\hat{y}(x)$ be the predicted mean of y from the regression of y on x, fitted to the respondent sample. Form a categorical version of $\hat{y}(x)$, say $\hat{y}_G$, by grouping $\hat{y}(x)$ into intervals, and then form adjustment cells by stratifying on $\hat{y}_G$. Values of $\bar{Y}(x)$ should be approximately constant within these cells, so the large sample bias of $\bar{y}_A$ and $\bar{y}_S$ should be nearly eliminated. We refer to this method of forming adjustment cells as <u>predicted mean</u> (PM) <u>stratification</u>.

If imputation is the chosen method of adjustments, then the adjustment cell mean is assigned to nonrespondents. An alternative method is regression imputation, where predictions $\hat{y}(x_i)$ are imputed directly, without forming adjustment cells based on $\hat{y}(x)$. If the regression equation captures the systematic variation in the y values, and the adjustment cells are large, then these methods should be quite similar. If adjustment cells contain a small number of respondents, the adjustment cell method lies somewhere between regression imputation, which imputes a conditional mean, and stochastic regression imputation, where noise is added to the predicted means (Little and Samuhel, 1983). As noted in section 1, weighting by the inverse response rates in the adjustment cell yields the same estimate $\bar{y}_A$ of $\bar{Y}$ as that obtained by mean imputation. This property establishes a link between weighting and regression imputation.

Practical limitations may inhibit PM stratification for certain problems. Note that regressions need to be developed for every y-variable subject to missing values, and these regressions yield different PM stratifications, and hence different weighting adjustments if weighting is the chosen mode of adjustment. One strategy is to estimate prediction equations for a small number m of key survey variables, and then to form joint classifications of the sample by the adjustment cell variables $A_1, A_2, \ldots, A_m$. Some pooling of the cells from this joint classification can form the basis for weighting adjustments that are relatively efficient for all the variables involved.

The second strategy for forming adjustment cells has the merit that it yields a unique set of adjustment cells for any block of variables $y_1, \ldots, y_k$ with the same response pattern. Such blocks occur notably with unit nonresponse, where the entire interview is missing for nonrespondents in the sample. Furthermore, the adjustment cells can be based on the results of a single regression, rather than requiring the fitting and combination of results from regressions on each y-variable.

The approach is suggested in David, Little, Samuhel and Triest (1983), and is a straightforward extension of the propensity score theory of Rosenbaum and Rubin (1983), developed in the context of matching in observational studies. Define the <u>response propensity</u>

$$p(x) = pr(r=1|x) ,$$

and suppose p(x) > 0 for all observed values of x. Then the theory of Rosenbaum and Rubin shows that (5) implies that

$$x \perp\!\!\!\perp r|p(x) \tag{6}$$

and

$$y \perp\!\!\!\perp r|p(x) . \tag{7}$$

That is, (4) is satisfied with A = p(x). This suggests the following strategy for choosing adjustment cells to limit nonresponse bias.

A) Estimate the propensity score p(x) by $\hat{p}(x)$, from the regression of the response indicator r on x. Forms of regression suitable for binary responses, such as logistic or probit regression, are advisable if response rates are close to zero or one.

B) Form adjustment cells based on $\hat{p}_G$, a grouped version of $\hat{p}(x)$. We describe this method of forming adjustment cells as <u>response propensity</u> (RP) <u>stratification</u>.

An alternative use of the estimated response propensity $\hat{p}(x)$ is to weight respondent i directly by the inverse of its estimated response propensity $\hat{p}(x_i)$, without forming adjustment cells as in B). This procedure avoids the choice of cutpoints required to form adjustment cells. However, respondents with very low values of $\hat{p}(x)$ receive large weights that can inflate the variance of survey estimates excessively. In the stratification approach, large weights can be dampened by a suitable choice of cutpoints for the variable $\hat{p}_G$. Another argument for stratification is that it places less reliance on correct specification of the response propensity regression, since the predictions are used only to partially order the sample, rather than to supply probabilities to be used directly in the weighting. Thus a linear regression of the response indicator may be adequate to define the adjustment cells, but inadequate for defining weights directly.

2.3 <u>Choosing Adjustment Cells to Limit Conditional Mean Squared Error Increments</u>

If response and outcome variable are independent, then the large sample bias of $\bar{y}_R$, $\bar{y}_A$ and $\bar{y}_S$ is zero, and relative precisions are measured by their respective conditional mse increments $C_R^2 + V_R$, $C_A^2 + V_A$ and V_S. Comparisons of $C_R^2 + V_R$

and V_S parallel those of Holt and Smith (1979) for the unweighted and poststratified mean, given complete response. Note that the sampling variance of the weighted linear combination $\bar{y}_W = \sum_{c=1}^{C} W_c \bar{y}_{cR}$ is minimized with weights W_c proportional to p_{cR}/S^2_{cR}. If the within cell variance S^2_{cR} is constant across cells, then this weight is equal to p_{cR}, yielding the estimator $\bar{y}_R$. Hence for any choice of adjustment cells,

$$V_R \leq V_A, \; V_R \leq V_S \; ,$$

if S^2_{cR} is constant across cells. On the other hand, in general we expect that

$$(C^2_S = 0) \leq C^2_A \leq C^2_R \; , \tag{8}$$

although the second inequality does not apply in all cases. To establish (8), note that if we treat the sample counts $\{n_c\}$ as multinomial with probabilities $\{P_c\}$ and index n_+, then averaging C^2_A over this distribution yields

$$\begin{aligned} E\{C^2_A\} &= E\{\sum_{c=1}^{C} (p_c-P_c)^2 \bar{Y}^2_{cR} + \sum_{c\neq c'} (p_c-P_c)(p_{c'}-P_{c'}) \bar{Y}_{cR}\, \bar{Y}_{c'R}\} \\ &= \sum_{c=1}^{C} P_c(1-P_c)\bar{Y}^2_{cR}/n_+ - \sum_{c\neq c'} P_c P_{c'} \bar{Y}_{cR}\bar{Y}_{c'R}/n_+ \\ &= \frac{1}{n_+} \sum_{c=1}^{C} P_c(\bar{Y}_{cR} - \bar{Y}_A)^2 \; , \end{aligned}$$

where $\bar{Y}_A = \sum_{c=1}^{C} P_c \bar{Y}_{cR}$. Similarly, if $\{n_{cR}\}$ are multinomial with probabilities $\{P_{cR}\}$ and index n_{+R}, then averaging C_R over this distribution yields

$$E\{C^2_R\} = \frac{1}{n_{+R}} \sum_{c=1}^{C} P_{cR}(\bar{Y}_{cR}-\bar{Y}_R)^2 \; .$$

If y and R are independent, then $P_{cR} = P_c$, $\bar{Y}_A = \bar{Y}_R$ and $E\{C^2_A\} = (n_{+R}/n_+)\, E\{C^2_R\}$, so the adjustment cell estimator $\bar{y}_A$ reduces the expected value of the component C^2_R of the mean squared error of $\bar{y}_R$ by a factor equal to the sample response rate.

For complete data, Holt and Smith (1979) discuss factors affecting the relative size of $C^2_R + V_R$ and V_S, and conclude that poststratification is relatively useful (that is, $V_S < C^2_R + V_R$) when the sample size is large and the ratio B/W of between to within cell variances of y is large. On the other hand, if the means of y between adjustment cells are close together, and the sample size is small, then the unweighted mean is favored. To understand the influence of B/W, note that if B/W is large, then C^2_R makes a relatively large contribution to $C^2_R + V_R$, and hence poststratification, which eliminates C_R at the expense of inflating V_R, is relatively profitable. Similar considerations apply to the weighting class estimator $\bar{y}_A$. However, conditions under which $\bar{y}_A$ is superior to $\bar{y}_R$ are more restricted, since $\bar{y}_A$ only reduces C^2_R by n_{+R}/n_+ (on the average).

The above discussion implies that adjustment cells should be chosen to maximize B/W, the ratio of between to within cell variance of y. With a large set of potential stratifiers x, this objective is achieved by PM stratification method discussed in section 2.2. Thus PM stratification has the virtue of controlling both the bias and variance of $\bar{y}_A$; RP stratification controls the large sample bias, but yields estimates $\bar{y}_A$ that may have large variance. The latter is particularly true when the response propensity is largely determined by variables that are not associated with y.

3. ESTIMATES OF CROSSCLASS MEANS

3.1 Introduction

Let $\bar{Y}_j$ denote the population mean of a variable y in a crossclass defined by the value z=j of a crossclass variable Z, assumed to be observed for all units in the sample. Other notation for population and sample quantities in crossclass j is given in Table 3 B); the notation parallels that in Table 3 A) with an additional subscript j for the crossclass.

Six estimators of $\bar{Y}_j$ are shown in Table 5, with their bias and variance properties under a restricted sampling distribution with y values, response indicators, and cell respondent and nonrespondent sample sizes in the crossclass held fixed.

Three of the estimators in Table 5, the unadjusted crossclass mean $\bar{y}_{jR}$ and the adjustment cell estimates from weighting $(\bar{y}^{(2)}_{jA})$ and from imputation $(\bar{y}^{(3)}_{jA})$, have already been introduced for the example in Table 2. The poststratified crossclass mean $\bar{y}_{jS}$ is obtained when the weights $(P_c/p_c)b_c^{-1}$ that result in the poststratified estimator $\bar{y}_S$ are applied to respondents in adjustment cell c, crossclass j, The adjusted mean $\bar{y}^{(1)}_{jA}$ is obtained by weighting or mean imputation within adjustment cells formed by the joint classification of A and Z. Finally $\bar{y}^{(4)}_{jA}$ is a model-based estimator motivated by the imputations leading to $\bar{y}^{(3)}_{jA}$. These imputations $\{\bar{y}_{c+R}\}$ pool the y values across subclasses within adjustment cells, and hence effectively assume that

$$\bar{Y}_{cjR} = \bar{Y}_{ckR} \text{ for all j, k} \; . \tag{9}$$

If the assumption (9) is firmly held, then a natural alternative is to pool across crossclasses when estimating respondent as well as nonrespondent means in cell c, crossclass j. This leads to the estimate $\bar{y}_{c+R}$ for $\bar{Y}_{cj}$ in cell c, crossclass j. Weighting $\bar{y}_{c+R}$ by p_{cj} yields $\bar{y}^{(4)}_{jA}$, as given in the table. With complete response, $\bar{y}^{(4)}_{jA}$ reduces to the so-called synthetic estimator, sometimes used when c represents census classifiers such as age, race and sex, and Z represents a small area classification (e.g., Gonzalez and Hoza, 1978).

3.2 Large Sample Bias of Crossclass Mean Estimators

The following results are obtained by considering the expressions for bias in Table 5 when the sample size becomes large:

1) The LSB of $\bar{y}_{jR}$ is $\bar{Y}_{+jR} - \bar{Y}_{+j}$, which is zero

when response and y are independent within crossclasses.

2) If $y \perp\!\!\!\perp r|x$, then $\bar{y}_{jA}^{(1)}$ has zero LSB under predictive mean or response propensity stratifications for A. To see this, note that $\bar{Y}_j$ is a domain mean when adjustment cells are based on A <u>and</u> Z, so the arguments of section 2 apply here.
3) If $y \perp\!\!\!\perp r|x$, then $\bar{y}_{jA}^{(2)}$ and $\bar{y}_{jS}$ have zero LSB with RP stratification, but in general non-zero LSB with PM stratification. To see this, note that the LSB of these estimators differs from the LSB of $\bar{y}_{jA}^{(1)}$ by the quantity $Q=\sum_c(\hat{P}_{cj}-P_{cj})\bar{Y}_{cjR}$, where $\hat{P}_{cj}=P_{cj}B_{cj}B_c^{-1}/\sum_{c'}P_{c'j}B_{c'j}B_{c'}^{-1}$. If A is formed by RP stratification, then response rates are homogeneous within A (that is, expression (6) holds), so $B_{cj}=B_c$ and $\hat{P}_{cj}=P_{cj}$ for all c, and hence Q=0. In general, $Q\neq 0$ for PM stratification.
4) If $y \perp\!\!\!\perp r|x$, then $\bar{y}_{jA}^{(3)}$ and $\bar{y}_{jA}^{(4)}$ have zero LSB with PM stratification, but in general non-zero LSB with RP stratification. To see this, note that these estimators have zero LSB with $\bar{Y}_{cjR}=\bar{Y}_{ckR}$ for all k, or more generally when $y \perp\!\!\!\perp Z|A$. This condition is satisfied by PM stratification, but not in general by RP stratification.

3.3 <u>Conditional Mean Squared Error Increments for Crossclass Mean Estimators</u>

The conditional mean squared error increments (ΔMSE) for the estimators in Table 5 have quite complicated expressions. Some general statements can be made, however:

1) Comparisons of $\bar{y}_{jR}$ and $\bar{y}_{jA}^{(1)}$ parallel comparisons of $\bar{y}_R$ and $\bar{y}_A$, except that they apply to quantities calculated within the crossclass. Thus $\bar{y}_{jA}^{(1)}$ dominates $\bar{y}_{jR}$ with respect to ΔMSE except when a) variability of y across the adjustment cells is small, and b) the sample sizes n_{cjR} are small. Note that b) is more likely for crossclass means than for domain means, since subclassification by crossclass reduces the sample sizes. Indeed, joint stratification by Z and A may yield cells with $n_{cj} > 0$ and $n_{cjR} = 0$, in which case $\bar{y}_{jA}^{(1)}$ cannot be calculated.
2) RP stratification yields estimators with smaller ΔMSE than PM stratification, since the within-cell variance of y is minimized.
3) With PM stratification, the model estimator $\bar{y}_{jA}^{(4)}$ should have lower ΔMSE than $\bar{y}_{jA}^{(3)}$ or $\bar{y}_{jA}^{(1)}$, since the distribution of y within adjustment cells is homogeneous. However, as the simulations in section 4 show, the model estimator is more sensitive to departures from homogeneity than the other estimators, so its use requires careful modeling of the regression of y on x when forming the adjustment cells.
4) Unlike $\bar{y}_{jA}^{(1)}$, the weighting estimators $\bar{y}_{jA}^{(2)}$ and $\bar{y}_{jS}$ do not require respondents in all cells (c,j) where $n_{cj} > 0$. This property suggests that $\bar{y}_{jA}^{(2)}$ and $\bar{y}_{jS}$ may have lower ΔMSE than $\bar{y}_{jA}^{(1)}$ when the respondent sample sizes

Table 5. Estimators of Crosstclass Means

Estimator	Bias	Variance
$\bar{y}_{jR} = \sum_c p_{cjR}\,\bar{y}_{cjR}$	$\sum_c (p_{cjR}-P_{cjR})\,\bar{Y}_{cjR}+(\bar{Y}_{+jR}-\bar{Y}_{+j})$	$\sum_c p_{cjR}\,S_{cjR}^2/n_{+jR}$
$\bar{y}_{jA}^{(1)} = \sum_c p_{cj}\,\bar{y}_{cjR}$	$B_1 = \sum_c (p_{cj}-P_{cj})\,\bar{Y}_{cjR}+\sum_c P_{cj}(\bar{Y}_{cjR}-\bar{Y}_{cj})$	$\sum_c p_{cj}^2\,U_{cj}$
$\bar{y}_{jA}^{(2)} = \sum_c \hat{p}_{cj}\,\bar{y}_{cjR}$	$B_1 + \sum_c (\hat{p}_{cj}-p_{cj})\,\bar{Y}_{cjR}$	$\sum_c \hat{p}_{cj}^2\,U_{cj}$
$\bar{y}_{jS} = \sum_c \hat{p}_{cj}^{\star}\,\bar{y}_{cjR}$	$B_1 + \sum_c (\hat{p}_{cj}^{\star}-p_{cj})\,\bar{Y}_{cjR}$	$\sum_c \hat{p}_{cj}^{\star 2}\,U_{cj}$
$\bar{y}_{jA}^{(3)} = \sum_c p_{cj}\,\tilde{y}_{cjR}$	$B_1 + \sum_c p_{cj}(1-b_{cj})D_{cj}$	$\sum_c p_{cj}^2\{b_{cj}^2(1 + \frac{2(n_{cj}-n_{cjR})}{n_{c+R}})U_{cj}+(1-b_{cj})^2V_{cj}\}$
$\bar{y}_{jA}^{(4)} = \sum_c p_{cj}\,\bar{y}_{c+R}$	$B_1 + \sum_c p_{cj}\,D_{cj}$	$\sum p_{cj}^2\,V_{cj}$

Notes: the following quantities in the table require definition:

(1) $\hat{p}_{cj} = p_{cj}b_{cj}b_c^{-1}/\sum_{c'}p_{c'j}b_{c'j}b_{c'}^{-1}$

(2) $\hat{p}_{cj}^{\star} = p_{cj}^{\star}b_{cj}b_c^{-1}/\sum_{c'}p_{c'j}^{\star}b_{c'j}b_{c'}^{-1}$,

where $p_{cj}^{\star} = p_{cj}P_c/p_c$

(3) $\tilde{y}_{cjR} = b_{cj}\bar{y}_{cjR} + (1-b_{cj})\bar{y}_{c+R}$

(4) $D_{cj} = \sum_k n_{ckR}\,n_{c+R}^{-1}\,(\bar{Y}_{ckR} - \bar{Y}_{cjR})$

(5) $U_{cj} = S_{cjR}^2/n_{cjR}$

(6) $V_{cj} = \sum_k n_{ckR}\,S_{ckR}^2/n_{c+R}^2$

$\{n_{cjR}\}$ are small. We examine this possibility in the next section.

4. A SIMULATION STUDY

4.1 Description of the Study

A simulation study was carried out to explore the mean squared errors of the estimators in Tables 4 and 5. Six factors affecting the mean squared error of the estimators were chosen as parameters in the study:

B^A: variation of population response rates $\{B_c\}$ between adjustment cells;
B^Z: variation of population response rates $\{B_{cj}\}$ between crossclasses, within adjustment cells;
M^A: variation of population means $\{\bar{Y}_{c+}\}$ between adjustment cells;
M^Z: variation of population means $\{\bar{Y}_{cj}\}$ between crossclasses, within adjustment cells;
R : correlation between response rates $\{B_c\}$ and cell means $\{\bar{Y}_{c+}\}$;
S : sample size.

Each of these factors was assigned two levels (1=Low, 2=High) and the factors varied in a 2^6 factorial design, yielding 64 problems. For each problem, 100 independent sets of sample sizes $\{n_{cj}\}$ and respondent sample sizes $\{n_{cjR}\}$ were generated, and for each set the root mean squared error (rmse) of each estimator in Tables 4 and 5 was calculated, using the formulae in the tables. Distributions of relative rmse's over the 100 data sets were then computed and summarized to yield measures of comparative performance. The simulations are similar to those of Holt and Smith (1979) directed at the effect of poststratification on estimates from completely observed simple random samples. However, our simulation design is considerably more complex since additional factors are involved.

Populations were constructed with twelve cells, formed by the joint classification of a six category adjustment cell variable A and a two category crossclass variable Z. The percentage distribution $\{100\ p_{cj}\}$ of the population across these cells was fixed, with values shown in Table 6A). Four sets of response rates $\{B_{cj}\}$ were determined by the levels of the factors B^A and B^Z. Table 6B) shows two choices (B^A=1,2) for the marginal response rates $\{B_c\}$, averaged over crossclass. Note that the variation of the response rates is small when B^A=1 (60% to 70%) and large when B^A=2 (40% to 90%). The factor B^Z determines the response rates for each crossclass within the adjustment cells. When B^Z=1, $\{B_{cj}\}$ are calculated so that $B_{c2}=1.02\ B_{c1}$ for all c, so variation of response rates between crossclasses is small. When B^Z=2, $B_{c2}=1.2\ B_{c1}$ for all c, so the variation of response rates between crossclasses is large.

Means $\{\bar{Y}_{cj}\}$ for the outcome variable y were determined by levels of the factors M^A and M^Z. Table 6C) shows two choices (M^A=1,2) for the marginal means $\{\bar{Y}_{c+}\}$ averaged over crossclass. Note that variation of these means is smaller when M^A=1 (30 to 40) than when M^A=2 (10 to 60). These ranges can be compared with the within cell variances $\{V_{cj}\}$ of the y values, which were set to 100 for all the cells. The factor M^Z determines the means $\{\bar{Y}_{cj}\}$ for each crossclass within the adjustment cells. When M^Z=1, $\{\bar{Y}_{cj}\}$ are calculated so that $\bar{Y}_{c2}=\bar{Y}_{c1}+2$ for all c, so variation of means between crossclasses is small. When M^Z=2, $\{\bar{Y}_{cj}\}$ are calculated so that $\bar{Y}_{c2}=\bar{Y}_{c1}+10$ for all j, so variation of means between crossclasses is large. The respondent means $\{\bar{Y}_{cjR}\}$ were set equal to $\{\bar{Y}_{cj}\}$ for all c, j. However, $\bar{Y}_{c+R} \neq \bar{Y}_{c+}$, since these marginal means are composed of different linear combinations of the cell means $\bar{Y}_{c1}$ and $\bar{Y}_{c2}$.

Table 6. Parameters for Simulation Study*

A) Population Distributions, $\{100\ P_{cj}\}$, $\{100\ P_c\}$

		Cell, A 1	2	3	4	5	6	
Crossclass, Z	1	3.66	7.32	9.76	12.20	13.41	12.20	58.54
	2	8.54	9.76	7.32	6.10	6.10	3.66	41.46
	ALL	12.20	17.07	17.07	18.29	19.51	15.85	100.00

B) Population Response Rates, Averaged over Crossclass, $\{B_c\}$

FACTOR	Cell, A 1	2	3	4	5	6
B^A = 1	0.60	0.62	0.64	0.66	0.68	0.70
B^A = 2	0.40	0.50	0.60	0.70	0.80	0.90

C) Population Means, Averaged over Crossclass, $\{\bar{Y}_{c+}\}$

FACTOR	Cell, A 1	2	3	4	5	6
M^A = 1	30	32	34	36	38	40
M^A = 2	10	20	30	40	50	60

*Other parameters are described in the text.

Two other factors, R and S, complete the description of the simulation design. When R=2, the response rates and population means are arranged as in Table 6, so they have a strong positive association: both the response rates and the cell means increase across the adjustment cells. When R=1, the marginal response rates $\{B_c\}$ and the corresponding within adjustment cell rates $\{B_{cj}\}$ are permuted so that rates indexed 1, 2, 3, 4, 5 and 6 are assigned to cells 1, 3, 5, 2, 4 and 6 respectively. This change largely eliminates the association between the means and the response rates. Finally, two sample sizes are chosen, n=240 (S=1) and n=2400 (S=2). These sample sizes are reduced by about 40% by nonresponse, and of course are further reduced for crossclass mean estimates other than $\bar{y}_{jR}^{(4)}$.

The sample sizes $\{n_{cj}\}$ are selected by multinomial random number generator GGMTN in the IMSL subroutine library, (IMSL, 1980), under the assumption that they have a multinomial distribution with index n and probabilities $\{P_{cj}\}$ given in Table 6A). The respondent sample sizes $\{n_{cjR}\}$ are selected by the binomial random number generator GGBN in the IMSL library, under the assumption that they have independent binomial distributions with index n_{cj} and probabilities B_{cj} determined by values B^A and B^Z. To avoid indeterminacy in the estimators in Table 5, samples were restricted to outcomes where $n_{cjR} > 0$ for all c and j, a similar strategy to that adopted by Holt and Smith (1979).

4.2 Summary Results

For each sample, the root mean squared error for each method was calculated as the square root of the sum of the bias^2 and variance in Table 4 or 5. Relative rmse's were then calculated with $\bar{y}_A$ and $\bar{y}_{jA}^{(2)}$, the estimators obtained by weighting respondents by the inverse of the response rates in the adjustment cells, as baseline estimators.

Hence

$$\mathrm{rel}(\bar{y}_R)=100(\mathrm{rmse}(\bar{y}_R)/\mathrm{rmse}(\bar{y}_A)-1),$$
$$\mathrm{rel}(\bar{y}_S)=100(\mathrm{rmse}(\bar{y}_S)/\mathrm{rmse}(\bar{y}_A)-1)$$

for estimators in Table 4, and

$$\mathrm{rel}(\bar{y}_{jR})=100(\mathrm{rmse}(\bar{y}_{jR})/\mathrm{rmse}(\bar{y}_{jA}^{(2)})-1),\ldots,$$
$$\mathrm{rel}(\bar{y}_{jA}^{(4)})=100(\mathrm{rmse}(\bar{y}_{jA}^{(4)})/\mathrm{rmse}(\bar{y}_{jA}^{(2)})-1),$$

for estimators in Table 5.

The average relative rmse over the 100 generated samples is used to summarize relative performances of the estimators for each problem. Crude rankings of the methods are obtained by further averaging over the 64 problems, which yields the following results:

Relative rmse's(rel)

$\bar{y}_R$	$\bar{y}_S$	$\bar{y}_{jR}$	$\bar{y}_{jA}^{(1)}$	$\bar{y}_{jS}$	$\bar{y}_{jA}^{(3)}$	$\bar{y}_{jR}^{(4)}$
112	-12	68	-3	-2	66	254

Thus $\bar{y}_R$ has on average 112% higher rmse than $\bar{y}_A$, reflecting its large bias for some problems. Within crossclasses, $\bar{y}_{jR}$ has on average a 68% higher rmse than $\bar{y}_{jA}^{(2)}$, a large increase but smaller than that for $\bar{y}_R$, reflecting the fact that the benefits of adjustment increase with the sample size.

Comparisons of $\bar{y}_S$ with $\bar{y}_A$ and $\bar{y}_{jS}$ with $\bar{y}_{jA}^{(2)}$ show the effects of poststratifying on the population proportions $\{P_c\}$ when available. The rmse is reduced by an average of 12% for estimates of the overall mean, and by an average of 2% for estimates of the crossclass mean. Poststratification is useful here, as in the complete data simulations of Holt and Smith (1979). The estimator $\bar{y}_{jA}^{(1)}$, obtained by further subdivision of the adjustment cells by crossclass, performs slightly better than $\bar{y}_{jA}^{(2)}$ on average, with a 3% reduction in rmse.

Imputation within adjustment cells ($\bar{y}_{jA}^{(3)}$) is markedly worse than weighting within adjustment cells ($\bar{y}_{jA}^{(2)}$), with an average increase of 66% in rmse. The reason for the poor performance of imputation and its model-based relative $\bar{y}_{jA}^{(4)}$ is discussed below.

4.3 Detailed Analysis of Root Mean Squared Errors

Detailed performance of the estimators is summarized in Table 7. A preliminary six-way analysis of variance of the average relative rmse's allowed two of the six factors to be eliminated with minor loss of information, viz. B^Z and M^Z for analyses of rel($\bar{y}_R$) and rel($\bar{y}_{jR}$), and B^A and R for analyses of the other estimators. The first sixteen rows of Table 7 give average relative rmse's for each of the sixteen combinations of the reduced factor set, ranked from low to high on the first estimator presented. The following eight rows give marginal means for the two levels of each factor, averaged over the other factors (M denotes average in the table). These means are omitted when the differential is small. Finally, the last row gives the overall means, as presented in section 4.2. The main features of Table 7 are as follows:

1) The adjusted means $\bar{y}_A$ and $\bar{y}_{jA}^{(2)}$ dominate the unadjusted means $\bar{y}_R$ and $\bar{y}_{jR}$ except when S=1, R=1, B^A=2 and M^A=1. The benefits of adjustment increase markedly when S, R, B^A and M^A are set to high levels, reflecting conditions where nonresponse bias of the unadjusted means is large.
2) $\bar{y}_{jA}^{(1)}$ has slightly lower rmse than $\bar{y}_{jA}^{(2)}$, with greatest reductions when M^A, S or B^Z are set to high levels. In comment 4) of section

Table 7. Average Relative Root Mean Squared Errors, Expressed as Percent Deviations, Classified by Four Most Important Factors

FACTOR*				Rel($\bar{y}_R$)	Rel($\bar{y}_{jR}$)	FACTOR*				Rel($\bar{y}_S$)	Rel($\bar{y}_{jS}$)	FACTOR*				Rel($\bar{y}_{jA}^{(3)}$)	Rel($\bar{y}_{jA}^{(4)}$)	FACTOR*				Rel($\bar{y}_{jA}^{(1)}$)
S	R	B^A	M^A			S	B^Z	M^A	M^Z			S	B^Z	M^A	M^Z			S	B^Z	M^A	M^Z	
1	1	2	1	-1	-3	2	1	2	1	-28	-5	2	2	2	1	-19	-9	2	2	2	2	-13
1	1	1	1	1	1	1	2	2	1	-28	-4	1	2	1	1	-14	-16	2	2	2	1	-9
2	1	1	1	2	1	2	2	2	1	-26	0	1	1	1	1	-12	-15	2	2	1	2	-5
1	2	1	1	4	2	1	1	2	1	-26	-7	1	2	2	1	-11	-14	1	2	2	2	-5
2	1	2	1	10	3	1	1	2	2	-24	-7	1	1	2	1	-11	-14	1	1	2	1	-4
1	1	1	2	13	7	2	1	2	2	-24	-4	2	1	2	1	-6	12	1	1	2	2	-3
2	1	1	2	17	11	1	2	2	2	-19	-2	2	2	1	1	-5	36	2	2	1	1	-3
1	1	2	2	20	11	2	1	1	1	-3	-2	2	1	1	1	-4	34	2	1	2	1	-2
2	2	1	1	24	15	2	2	1	1	-3	-1	1	1	2	2	17	143	1	2	2	1	-2
1	2	1	2	29	17	2	1	1	2	-2	-2	1	2	2	2	26	160	2	1	1	2	-1
1	2	2	1	35	19	1	1	1	1	-2	-1	1	1	1	2	45	235	2	1	2	2	-1
2	1	2	2	72	38	1	2	1	1	-2	0	1	2	1	2	61	254	2	1	1	1	-1
2	2	1	2	146	91	1	1	1	2	-2	-1	2	1	2	2	181	682	1	2	1	2	-1
2	2	2	1	196	151	1	2	1	2	0	0	2	2	2	2	185	621	1	1	1	2	-1
1	2	2	2	260	135	2	2	1	2	1	-1	2	1	1	2	287	966	1	1	1	1	0
2	2	2	2	964	584	2	2	2	2	3	0	2	2	1	2	339	990	1	2	1	1	0
1	M	M	M	45		1	M	M	M			1	M	M	M	13	92	1	M	M	M	-2
2	M	M	M	179		2	M	M	M			2	M	M	M	120	417	2	M	M	M	-4
M	1	M	M	17	9	M	1	M	M	-14	-4	M	1	M	M			M	1	M	M	-2
M	2	M	M	207	127	M	2	M	M	-9	-1	M	2	M	M			M	2	M	M	-5
M	M	1	M	29	18	M	M	1	M	-2	-1	M	M	1	M	87	311	M	M	1	M	-1
M	M	2	M	194	117	M	M	2	M	-22	-4	M	M	2	M	45	198	M	M	2	M	-5
M	M	M	1	34	24	M	M	M	1	-15		M	M	M	1	-10	.2	M	M	M	1	
M	M	M	2	190	112	M	M	M	2	-9		M	M	M	2	142	507	M	M	M	2	
M	M	M	M	112	68	M	M	M	M	-12	-2	M	M	M	M	66	254	M	M	M	M	-3

*1 = Low, 2 = High, M = Average

3.3, we speculated that $\bar{y}^{(2)}_{jA}$ may outperform $\bar{y}^{(1)}_{jA}$ for small sample sizes. This does not happen in Table 7, but $\bar{y}^{(2)}_{jA}$ had slightly lower (1%) average rmse in supplementary simulations at a smaller sample size, n=80.

3) Comparisons of $\bar{y}_S$ with $\bar{y}_A$ and $\bar{y}_{jS}$ with $\bar{y}^{(2)}_{jA}$ indicate that poststratification nearly always reduces rmse for the problems simulated. Gains are greatest when a) $M^A=2$, b) $B^Z=1$ and c) the overall mean rather than crossclass means are considered. These findings agree with those of Holt and Smith (1979) for complete data.

4) For crossclass means, imputation ($\bar{y}^{(3)}_{jA}$) outperforms weighting ($\bar{y}^{(2)}_{jA}$) when $M^Z=1$, that is, crossclass means within adjustment cells are nearly equal. This should be the case with PM stratification. However $\bar{y}^{(3)}_{jA}$ does very poorly when $M^Z=2$, where bias dominates its rmse.

5) The model estimator $\bar{y}^{(4)}_{jA}$ should dominate all estimators if crossclass means are equal within adjustment cells. Nevertheless, when this condition is nearly satisfied ($M^Z=1$), $\bar{y}^{(3)}_{jA}$ still dominates $\bar{y}^{(4)}_{jA}$ in our simulations, indicating sensitivity of $\bar{y}^{(4)}_{jA}$ to even mild departures from the modeling assumption. $\bar{y}^{(4)}_{jA}$ is particularly bad in large samples (S=2), and has very high rmse when $M^Z=2$, when it is seriously biased.

5. NEW APPROACHES

5.1 Introduction

It would be quite unjustified to draw general conclusions about the relative merit of the estimators in Tables 4 and 5 from the simulations in section 4, since the results are highly dependent on the parameter levels chosen in the study. Nevertheless the simulations do illustrate how the relative performance of the estimators changes as a result of changes in the sample size and in the population structure. If adjustment cells are chosen by PM stratification, then mean imputation within the adjustment cells works well, and weighting yields the same estimator for domain means and somewhat less efficient estimates for crossclass means. If adjustment cells are chosen by RP stratification, then weighting successfully controls nonresponse bias but may have large variance, and imputation controls variance but may lead to serious bias.

Weighting class estimators based on RP stratification have useful bias reduction properties, and are particularly economical for data sets containing a large set of y variables with the same missing data pattern. However, weighting needlessly increases the variance when the outcome variable y is not related to the propensity to respond. In this concluding section, we propose modified weighting class estimators that seek to limit variance whilst retaining the ability to adjust for nonresponse bias. Section 5.2 considers the case of domain means, and section 5.3 considers crossclass means.

5.2 Modified Weighting Class Estimators for Domain Means

A straightforward approach to limiting the variance of estimates of domain means is to regress y on the estimated response propensity $\hat{p}(x)$ using the respondent sample (David, Little, Samuhel and Triest, 1983). If the coefficient of $\hat{p}(x)$ is significantly different from zero, then adjustment is in order. A regression more closely related to RP stratification is obtained by replacing the regressor $\hat{p}(x)$ by dummy indicators for the adjustment cells. If an F-test for the adjustment cell coefficients reveals significant effects, then $\bar{y}_A$ is chosen; otherwise $\bar{y}_R$ is chosen.

An elaboration of the above approach that leads to a compromise between $\bar{y}_A$ and $\bar{y}_R$ is obtained by fitting separate means for each adjustment cell, but treating the means as random variables from a common distribution. The simplest version of the model assumes that

$$\begin{aligned}(\bar{y}_{cR}|\bar{Y}_{cR}, S^2_{cR},\mu,\tau^2) &\sim N(\bar{Y}_{cR}, S^2_{cR}/n_{cR}) \\ (\bar{Y}_{cR}|S^2_{cR},\mu,\tau^2) &\sim N(\mu,\tau^2) ,\end{aligned} \qquad (10)$$

where N(a,b) denotes the normal distribution with mean a, variance b. Given μ, S^2_{cR} and τ^2, $\bar{Y}_{cR}$ can be estimated by its posterior mean

$$\begin{aligned}&E(\bar{Y}_{cR}|\mu,S^2_{cR},\tau^2) \\ &= \tau^2(\tau^2+S^2_{cR}n^{-1}_{cR})^{-1}\ \bar{y}_{cR}+n^{-1}_{cR}(\tau^2+S^2_{cR}n^{-1}_{cR})\mu \qquad (11) \\ &= n_{cR}(n_{cR}+\lambda)^{-1}\ \bar{y}_{cR}+ \lambda(n_{cR}+\lambda)^{-1}\mu ,\end{aligned}$$

if $S^2_{cR}=S^2_R$ is constant across cells, and $\lambda = S^2_R/\tau^2$, the ratio of within cell to between cell variances. In practice, μ and λ need to be estimated from the data.

If adjustment cells are chosen so that n_{cR} is constant across adjustment cells, then an efficient estimator of μ is $\bar{y}_R$; λ can be estimated by equating observed and expected means squares from analysis of variance, or by the more refined procedures of Hill (1980). If $\{n_{cR}\}$ vary across cells, then the iterative procedures of Carter and Rolph (1974) can be used to estimate μ and λ.

Replacing μ and λ in (11) by estimates $\bar{y}_R$ and $\hat{\lambda}$, and substituting the resulting estimate of $\bar{Y}_{cR}$ for $\bar{y}_{cR}$ in (2) yields an empirical Bayes (EB) estimator of $\bar{Y}$:

$$\bar{y}_{EB}=n^{-1}_{+} \sum_{c=1}^{C} n_{cR}\ \bar{y}_{cR}(\lambda_c b^{-1}_c) , \qquad (12)$$

where

$$\lambda_c = \frac{n_{cR}(n_c + k_1)}{n_c(n_{cR} + k_2)} \qquad (13)$$

is a modification of the weight b^{-1}_c in $\bar{y}_A$, and

$$k_1 = \hat{\lambda} \quad \frac{\sum_{c=1}^{C} n_c(n_{cR}+\hat{\lambda})^{-1}}{\sum_{c=1}^{C} n_{cR}(n_{cR}+\hat{\lambda})^{-1}}; \; k_2 = \hat{\lambda} .$$

Note that $\bar{y}_{EB}$ is close to $\bar{y}_A$ when adjustment is beneficial (large samples, large ratio of between to within variances) and otherwise is close to $\bar{y}_R$. Thus it is an attractive compromise between $\bar{y}_R$ and $\bar{y}_A$. However, the factors λ_c for multiplying the raw weights b^{-1}_c require some computational effort, and are different for each of a set of y variables. Thus some of the simplicity of the propensity weighting scheme is lost.

The key assumption of (10) is that the underlying cell means $\bar{Y}_{cR}$ are exchangeable. If on the contrary the means are systematically related to the cell response rates b_c, then a random effects model that shrinks towards the regression line $\mu_0 + \beta(b_c - b_+)$ rather than towards the point $\bar{y}_R$, would be preferable. Such a model would combine elements of (10) and the fixed effects regression model that opened this section. A plot of $\bar{y}_{cR}$ against b_c should serve as a useful diagnostic tool for determining whether this more elaborate model is needed.

5.3 Modified Estimators for Crossclass Means

The empirical Bayes estimator (12) is obtained by weighting respondents by $\lambda_c b_c^{-1}$, where λ_c is given by (13). If crossclass means are estimated using this weighting scheme, the result is an estimator that behaves like $\bar{y}_{jA}^{(2)}$ when sample sizes are large, and like $\bar{y}_{jR}$ when sample sizes are small. The regression of y on $\hat{p}(x)$ in section 5.2 also provides guidance as to whether adjustment of $\bar{y}_{jR}$ for nonresponse is needed, although more specific information may be obtained by restricting the regression to respondents in the crossclass.

If regression prediction is used to impute for nonrespondents, and elaborate regression modeling of y is too time consuming, then one might wish to restrict the regressors to the propensity score $\hat{p}(x)$ and dummy indicator variables for the crossclasses of interest. The inclusion of the latter variables avoids bias from imputing means that average over crossclasses that are heterogeneous with respect to y. The inclusion of response propensity as a regressor protects against nonresponse bias.

In summary, a number of alternative approaches can be envisioned for improving the estimators of Table 5, short of full modeling of the relationship between y and the crossclass variable and other regressors.

REFERENCES

Carter, G.M. and Rolph, J.E. (1974). "Empirical Bayes methods applied to estimating fire alarm probabilities," Journal of the American Statistical Association, 69, 880-5.

David, M., Little, R., Samuhel, M., and Triest, R. (1983). "Nonrandom nonresponse models based on the propensity to respond," Proceedings of the Survey Research Methods Section, American Statistical Association 1983.

Dawid, A.P. (1979). "Conditional independence in statistical theory," (with discussion), Journal of the Royal Statistical Society, Series B, 41, 1-31.

Ernst, L.R. (1980). "Variance of the estimated mean for several imputation procedures," Proceedings of the Survey Research Methods Section, American Statistical Association 1980, 716-720.

Gonzalez, M.E. and Hoza, C. (1978). "Small area estimation with application to unemployment and housing estimates," Journal of the American Statistical Association, 73, 7-15.

Hanson, R.H. (1978). "The current population survey, design and methodology," Technical Paper 40, Bureau of the Census, Washington, DC 20233.

Hill, B.M. (1980). "Robust analysis of the random model and weighted least squares regression," in Evaluation of Econometric Models, New York: Academic Press.

Holt, D. and Smith, T.M.F. (1979). "Poststratification," Journal of the Royal Statistical Society, Series A, 142, 33-66.

IMSL (1980). IMSL Library, Edition 8. Reference Manual. IMSL, Inc., 7500 Bellaire Boulevard, Houston, TX 77036.

Kalton, G. and Kasprzyk, D. (1982). "Imputing for missing survey responses," Proceedings of the Survey Research Methods Section, American Statistical Association 1982, 22-31.

Kalton, G. and Kish, L. (1981). "Two efficient random imputation procedures," Proceedings of the Survey Research Methods Section, American Statistical Association 1981, 146-151.

Little, R.J.A. (1982). "Models for nonresponse in sample surveys," Journal of the American Statistical Association, 77, 327-350.

Little, R.J.A. and Samuhel, M.E. (1983). "Alternative models for CPS income imputation," Proceedings of the Survey Research Methods Section, American Statistical Association 1983.

Oh, H.L. and Scheuren, F. (1980). "Estimating the variance impact on missing CPS income data," Proceedings of the Survey Research Methods Section, American Statistical Association 1980, 408-415.

Oh, H.L. and Scheuren, F.S. (1983). "Weighting adjustments for unit nonresponse," in Incomplete Data in Sample Surveys, Vol. 2, W.G. Madow, I. Olkin and D.B. Rubin, eds. New York: Academic Press.

Rosenbaum, P.R. and Rubin, D.B. (1983). "The central role of the propensity score in observational studies for causal effects," Biometrika, 70, 41-55.

Rubin, D.B. (1978). "Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse," Proceedings of the Survey Research Methods Section, American Statistical Association 1978, 20-34.

Thomsen, I. (1973). "A note on the efficiency of weighting subclass means to reduce the effects of nonresponse when analyzing survey data," Statistik Tidskrift, 11, 278-283.

Thomsen, I. (1978). "A second note on the efficiency of weighting subclass means to reduce the effects of nonresponse when analyzing survey data," Statistik Tidskrift, 16, 191-196.

Welniak, E.G. and Coder, J.F. (1980). "A measure of the bias in the March CPS earnings imputation scheme," Proceedings of the Survey Research Methods Section, American Statistical Association 1980, 421-425.