Survey Methodology, June 1997
Vol. 23, No. 1, pp. 33-40
Statistics Canada

Diagnostics for Formation of Nonresponse Adjustment Cells, With an Application to Income Nonresponse in the U.S. Consumer Expenditure Survey

JOHN L. ELTINGE and IBRAHIM S. YANSANEH[1]

ABSTRACT

This paper discusses the use of some simple diagnostics to guide the formation of nonresponse adjustment cells. Following Little (1986), we consider construction of adjustment cells by grouping sample units according to their estimated response probabilities or estimated survey items. Four issues receive principal attention: assessment of the sensitivity of adjusted mean estimates to changes in k, the number of cells used; identification of specific cells that require additional refinement; comparison of adjusted and unadjusted mean estimates; and comparison of estimation results from estimated-probability and estimated-item based cells. The proposed methods are motivated and illustrated with an application involving estimation of mean consumer unit income from the U.S. Consumer Expenditure Survey.

KEY WORDS: Incomplete data; Missing data; Quasi-randomization; Response propensity; Sensitivity analysis; Weighting adjustment.

1. INTRODUCTION

1.1 Problem Statement

Survey analysts often use adjustment cell methods to account for nonresponse. The main idea is to define groups, or "cells", of sample units which are believed to have approximately equal response probabilities, or approximately equal values of a specific survey item, *e.g.*, income. Weighting adjustment or simple hot-deck imputation then is carried out separately within each adjustment cell. The resulting adjusted estimator of a population mean or total will have a nonresponse bias approximately equal to zero, provided the within-cell covariances between survey items and response probabilities are approximately equal to zero.

Some previous nonresponse-adjustment work formed adjustment cells through combinations of simple demographic or geographical classificatory variables. However, Little (1986) and others considered formation of cells by direct grouping of sample units according to their estimated response probabilities or estimated item values. The present paper discusses some simple diagnostics that are useful in implementing these cell-formation ideas. Principal attention is directed to the sensitivity of results to the number of cells used; identification of specific cells that require additional refinement; comparison of adjusted and unadjusted mean estimates; and comparison of estimation results from estimated-probability and estimated-item based cells. These diagnostics are illustrated with income data collected in the U.S. Consumer Expenditure Survey.

1.2 Notation, Nonresponse Bias, and Adjustment Cells

Let U be a fixed population of size N with survey items Y_i, $i \in U$; and consider estimation of the population mean $\bar{Y} = N^{-1}\sum_{i\in U} Y_i$. A sample s of size n is selected from U, and π_i is the probability that unit i is included in the sample.

Nonresponse is assumed to satisfy the following quasi-randomization model (Oh and Scheuren 1983). Let R_i be an indicator variable equal to 1 if the selected sample unit i is a respondent and equal to 0 otherwise. Assume that the R_i are mutually independent Bernoulli (η_i) random variables, where the fixed response probabilities η_i are allowed to differ across units. In addition, define the survey weights $\lambda_i = \pi_i^{-1}$ and the unadjusted survey-weighted mean response

$$\hat{\bar{Y}}_1 \overset{\text{def}}{=} \left(\sum_{i\in s} \lambda_i R_i\right)^{-1} \sum_{i\in s} \lambda_i R_i Y_i. \tag{1.1}$$

Because of differences among the η_i, the unadjusted estimator $\hat{\bar{Y}}_1$ has a nonresponse bias approximately equal to $N^{-1}\bar{\eta}^{-1}\sum_{i\in U}\eta_i(Y_i - \bar{Y})$, where $\bar{\eta} = N^{-1}\sum_{i\in U}\eta_i$ and expectations are taken over both the original sample design and the quasi-randomization model. To reduce this bias, one often partitions the population into k "adjustment cells" U_h, partitions the sample s into corresponding groups s_h, and then uses the adjusted estimator

$$\hat{\bar{Y}}_k \overset{\text{def}}{=} \sum_{h=1}^{k} w_h \bar{Y}_{hR}, \tag{1.2}$$

where $w_h = (\sum_{i\in s}\lambda_i)^{-1}\sum_{i\in s_h}\lambda_i$ and $\bar{Y}_{hR} = (\sum_{i\in s_h}\lambda_i R_i)^{-1} \sum_{i\in s_h}\lambda_i R_i Y_i$. Note that if $k = 1$, then estimators (1.1) and (1.2) are identical. For some general discussion of adjustment cell methods see, *e.g.*, Cassel, Särndal and Wretman (1983), Oh and Scheuren (1983), and Kalton and Maligalig (1991).

The adjusted estimator $\hat{\bar{Y}}_k$ has remaining nonresponse bias approximately equal to

$$N^{-1}\sum_{h=1}^{k} \bar{\eta}_h^{-1} \sum_{i\in U_h} (\eta_i - \bar{\eta}_h)(Y_i - \bar{Y}_h), \tag{1.3}$$

[1] John L. Eltinge, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, U.S.A.; Ibrahim S. Yansaneh, Westat, 1650 Research Blvd., Rockville, MD 20850-3195, U.S.A.

where N_h is the number of units in U_h and $(\bar{\eta}_h, \bar{Y}_h) = N_h^{-1} \sum_{i \in U_h} (\eta_i, Y_i)$. Consequently, one prefers to construct cells such that the population covariance between η_i and Y_i is approximately equal to zero within each cell. In practice, one attempts to accomplish this by constructing cells that are approximately homogeneous in the response probabilities η_i or in the items Y_i, or both. In some cases, "natural" sets of cells are defined *a priori* through combinations of classificatory variables that are available for both respondents and nonrespondents. For example, Ezzati and Khare (1992) used 72 cells defined by age, race, region, urbanization status, and household size to perform nonresponse adjustments for part of the National Health and Nutrition Examination Survey. In many practical cases, however, the list of reasonable candidate variables for cell formation is fairly large, and may produce a substantial number of cells that contain few, if any, respondents. Consequently, several authors have developed methods to screen out the less important classificatory variables and to collapse sparse adjustment cells in a way that preserves a reasonable degree of homogeneity within each of the remaining cells. See, *e.g.*, Tremblay (1986); Lepkowski, Kalton and Kasprzyk (1989); Kalton and Maligalig (1991); Göskel, Judkins and Mosher (1991); and the related discussion of pooling of poststrata in Little (1993). In addition, adjustment cell methods are related to other methods like regression-based adjustments (*e.g.*, Rao 1996, Section 2.4 and references cited therein) and generalized raking (Deville, Särndal and Sautory 1993).

1.3 Adjustment Cells Based on Estimated Response Propensities or Predicted Items

Adjustment cells are expected to be approximately homogeneous, so one may argue that such cells implicitly define a model for either the η_i or Y_i values, or both. More explicit modeling leads to two related cell formation methods. First, let X_i be a vector of auxiliary variables observed for both responding and nonresponding sample units *i*, and use the sample (R_i, X_i) values to fit a model for $\eta_i = \eta(X_i)$ through linear, logistic, or probit regression. The sample cells s_h are then formed by grouping the sample units according to their estimated response probabilities $\hat{\eta}_i$. As a second alternative, consider regression of responses Y_i on an auxiliary vector X_i to produce estimated items $\hat{Y}_i$ for both responding and nonresponding sample units. The sample cells s_h are then formed by grouping units according to the values $\hat{Y}_i$.

These two methods were suggested by Little (1986), extending the observational-data propensity-score work of Rosenbaum and Rubin (1983, 1984). See also David, Little, Samuhel and Triest (1983). These ideas were developed originally in a model- based context, but extend directly to the current framework. Little (1986) argued that use of cells based on either the $\hat{\eta}_i$ or $\hat{Y}_i$ values could reduce nonresponse bias, and that the $\hat{Y}_i$-based cells could also control variance. Also, in some cases the $\hat{\eta}_i$ and $\hat{Y}_i$-based cells can be more flexible than cells defined *a priori*. In addition, the $\hat{Y}_i$-based adjustment cells are conceptually related to optimum stratification ideas (*e.g.*, Cochran 1977, Sections 5A.7-5A.8).

Little (1986) did not propose a specific rule to determine cell divisions. However, in keeping with related observational-data work by Cochran (1968) and by Rosenbaum and Rubin (1984), one may consider cell divisions defined by the estimated $k^{-1}j$ quantiles of the $\hat{\eta}_i$ or $\hat{Y}_i$ populations, $j = 1, 2, ..., k-1$. This equal-quantile method gives some control over the expected number of respondents in each cell. In addition, review of the preceding two references suggests that, for a given set of predictors X_i, most of the feasible bias reduction may be achieved with a relatively small number of cells, say $k = 5$. A case study by Czajka, Hirabayashi, Little and Rubin (1992) used $k = 6$ $\hat{\eta}_i$-based adjustment cells within each of several strata, using cell-formation rules that were somewhat more complex than the equal-quantile rule considered here. However, the potential adequacy of a small number of cells should not be over-interpreted. For example, if an important regressor is omitted, then the resulting cell-based adjusted estimators may retain a substantial amount of bias, regardless of the specific number of estimated-probability or estimated-item based cells used.

Finally, an important alternative to weighting adjustment is imputation. For example, simple hot-deck imputation replaces a missing value within a given adjustment cell by randomly selecting respondent donors from the same cell. In parallel with (1.1) and (1.2), the resulting mean estimator is $\hat{\bar{Y}}_{\text{imp}} = (\sum_{i \in s} \lambda_i)^{-1} \sum_{i \in s} \lambda_i Y_i^*$, where Y_i^* is either an observed or imputed value, as appropriate. Practical applications often use weighting adjustment for unit nonresponse and imputation for item nonresponse. However, for a given set of cells, both the weighting adjustment point estimator (1.2) and the imputation estimator $\hat{\bar{Y}}_{\text{imp}}$ have the same approximate bias (1.3). For simplicity, the remainder of this paper will focus on weighting adjustment, but one should bear in mind that for a given set of cells, the same bias-reduction issues arise regardless of whether those cells are used for weighting adjustment or simple hot deck imputation.

1.4 Outline of the Present Paper

This paper discusses some implementation details of the estimated-probability and estimated-item methods of cell formation. We devote special attention to diagnostics to identify problems in a specific set of cells, and motivate and illustrate these diagnostics with an extended example involving income nonresponse in the U.S. Consumer Expenditure Survey. Section 2 gives some general background on this income nonresponse problem. Section 3 describes and applies several diagnostics, including comparison of $\hat{\bar{Y}}_k$ estimates and standard errors for several values of k (Section 3.1); partial assessment of within-cell bias (Section 3.2.1); assessment of cell widths relative to the precision of $\hat{\eta}_i$ estimates (Section 3.2.2); and comparison of the adjusted and unadjusted mean estimates $\hat{\bar{Y}}_k$ and $\hat{\bar{Y}}_1$ (Section 3.3). Section 4 shows that similar diagnostics can be applied to adjustment cells based on predicted incomes $\hat{Y}_i$, and also compares the mean income estimates computed from estimated-probability and estimated-income based cells. Section 5 summarizes the main ideas used in this paper, and notes some areas for future research.

2. INCOME NONRESPONSE IN THE U.S. CONSUMER EXPENDITURE SURVEY

2.1 The Consumer Expenditure Survey, Weighting Methods and Variance Estimation

The U.S. Consumer Expenditure Survey (CE) is a stratified multistage rotation sample survey conducted by the Census Bureau for the Bureau of Labor Statistics. Sample elements are "consumer units", roughly equivalent to households. In the interview component of this survey, each selected sample unit is asked to participate in five interviews. The current CE weighting procedure accounts for initial selection probabilities, a noninterview adjustment, post-stratification based on several demographic variables, and additional refinements; see Zieschang (1990) and United States Bureau of Labor Statistics (1992). The complexity of the CE weighting work has led the BLS to use variance estimators based on pseudo-replication methods with 44 replicates. This pseudo-replication is approximately equivalent to standard balanced repeated replication (Wolter 1985, Ch. 3). All standard errors reported here are based on this pseudo-replication method, with all additional parameter estimation and weighting adjustment steps performed separately within each replicate.

2.2 Income Nonresponse

The noninterview adjustment in the current CE weighting procedure is generally considered to account adequately for unit nonresponse, *e.g.*, noncontact or refusal to participate in a specific interview. Thus, unit nonresponse in the CE will not be considered further here. However, the BLS has had concerns about possible bias in mean income estimates due to item nonresponse that occurs with income questions in the CE; some background is as follows.

Detailed income data are collected in the second and fifth interviews of the CE, and are used to produce estimates of mean consumer unit income (U.S. Bureau of Labor Statistics 1991) and other parameters. CE income data are collected through a complex set of questions, and nonresponse rates for these questions are relatively high. To provide a summary indication of response or nonresponse to the full set of income questions, the BLS classifies each second- or fifth-interview consumer unit as a complete or incomplete reporter of income. The formal definition of "complete income reporter" status is fairly complex; Garner and Blanciforti (1994) give a detailed discussion. Current BLS procedure estimates mean income with the unadjusted mean response $\hat{\bar{Y}}_1$ defined by (1.1), with the R_i equal to indicators of complete income reporting, Y_i equal to income, and weights λ_i as described in Section 2.1. The weighted mean $\hat{\bar{Y}}_1$ uses both second- and fifth-interview data from a specified time period, but does not make direct use of the CE panel-data structure. In parallel with this, the present paper will distinguish between second- and fifth-interview data only in the construction of $\hat{\eta}_i$ and $\hat{Y}_i$ models.

Here, we used data from the second and fifth interview reports from all consumer units that had a second interview scheduled during 1990. The second-interview data involved 5,125 interviewed units and the fifth-interview data involved 5,093 interviewed units. For each interviewed unit (both the complete and the incomplete income reporters), BLS records provided a large number of demographic and expenditure variables; these were used as auxiliary variables in the modeling work described in Sections 3 and 4 below. For both the second and the fifth interviews, approximately 14 percent of the interviewed consumer units were incomplete income reporters.

3. CELLS BASED ON ESTIMATED RESPONSE PROBABILITIES

We first considered construction of adjustment cells based on estimated response probabilities. Logistic regression models for the complete-income-reporter probabilities $\eta_i = \eta(X_i)$ were fit separately for the second and fifth interview data described in Section 2. Model fitting details, including model parameter estimates and standard errors, are reported in Yansaneh and Eltinge (1993). All variance estimates were computed by the pseudo-replication method described in Section 2. The final model fits were used to estimate complete-reporter probabilities $\hat{\eta}_i$ for each second- and fifth-interview unit. Following the strategy in Section 1.3, units were grouped according to their $\hat{\eta}_i$ values into a total of k cells, with cell boundaries defined by the equal-quantile method.

3.1 Initial Sensitivity Analysis for the Number of Cells Used

The first three columns of Table 1 report the adjusted point estimates $\hat{\bar{Y}}_k$ of mean income, and associated standard errors, for several values of k. Comparisons of these point estimates indicate the extent to which the adjusted estimates are sensitive to a specific choice of k. For $k \geq 5$, the reported point estimates are relatively stable, varying between \$32,630 and \$32,664. This is consistent with the suggestion in Section 1.3 that $k = 5$ cells may provide most of the effective bias reduction to be obtained from a given cell-formation method; see Rosenbaum and Rubin (1984, Section 1 and Appendix A) for some related mathematical background.

In addition, note that for $k \geq 3$, the standard errors of $\hat{\bar{Y}}_k$ are also relatively stable, ranging from \$508 to \$530. This is in partial contrast with the general idea that selection of an appropriate number of cells hinges on a bias-variance trade-off. For the present dataset, it appears that the effective bias reduction occurs fairly quickly (at $k = 5$, say), while substantial variance inflation does not occur until some point beyond $k = 20$. This is not unreasonable, since even for $k = 20$, the number of income responses per cell remained fairly large (ranging from 461 to 569), and thus avoided the general unstable-estimator problem associated with increasing numbers of sparse cells. Conversely, bias-variance tradeoff problems may be more severe for moderate k in applications involving smaller effective sample sizes, *e.g.*, estimation for small subpopulations.

Table 1
Adjusted Estimates of Mean Income with Cell Boundaries Determined by Estimated Response Probability Quantiles

Number of Cells	Point Estimate	Standard Error	$\mathrm{SE}(\hat{\bar{Y}}_k - \hat{\bar{Y}}_1)$	MSE Ratio $(\hat{\gamma}_k)$
Unadjusted ($k = 1$)	32,967	569	N/A	N/A
$k = 3$ cells	32,736	530	112	1.30
$k = 4$ cells	32,779	518	122	1.28
$k = 5$ cells	32,630	523	138	1.53
$k = 6$ cells	32,664	515	122	1.51
$k = 10$ cells	32,640	514	116	1.58
$k = 15$ cells	32,638	515	118	1.58
$k = 20$ cells	32,634	508	118	1.63

3.2 Two Simple Cell Diagnostics

To complement the preceding sensitivity analysis, it is useful to study some sets of adjustment cells in additional detail. Let $C_1 = \{s_1, ..., s_k\}$ be a given candidate set of adjustment cells, *e.g.*, the $k = 3$ or $k = 5$ equal-quantile- division cells in Section 3.1. The cells in C_1 can be refined by using equal-quantile divisions with a larger value of k; or by directly splitting one or more of the cells in C_1. This refinement may be worthwhile if there are empirical indications: (1) that the within-cell mean estimator $\bar{Y}_{hR}$ may be substantially biased; or (2) that a cell is wide relative to the precision with which the η_i values are estimated. Subsections 3.2.1 and 3.2.2 use two simple diagnostic methods to address issues (1) and (2), respectively. In each subsection, the proposed diagnostics lead to identification of potential "problem cells", and to construction of a refined set of adjustment cells, C_2, say. Comparisons of estimates of $\bar{Y}$ based on C_1 and C_2 then lead to some conclusions regarding the preferred set of $\hat{\eta}_i$-based adjustment cells.

3.2.1 Assessment of Within-Cell Bias

As noted in Section 1.2, a given adjusted estimator $\hat{\bar{Y}}_k$ reduces, but may not entirely eliminate, nonresponse bias; and the residual bias of $\hat{\bar{Y}}_k$ depends on the biases of the within-cell mean estimates $\bar{Y}_{hR}$. Consider the alternative within-cell mean estimator

$$\bar{Y}_{h\eta} = \left(\sum_{i \in s_h} \hat{\eta}_i^{-1} \lambda_i R_i \right)^{-1} \sum_{i \in s_h} \hat{\eta}_i^{-1} \lambda_i R_i Y_i. \qquad (3.1)$$

If the $\hat{\eta}_i$ estimates were equal to the true response probabilities η_i, then (3.1) would be an approximately unbiased estimator of the true subpopulation mean $\bar{Y}_h$. In that case, an estimator of the within-cell bias $E(\bar{Y}_{hR} - \bar{Y}_h)$ would be $\hat{B}_h = \bar{Y}_{hR} - \bar{Y}_{h\eta}$, and the corresponding estimator of the overall bias $E(\hat{\bar{Y}}_k - \bar{Y})$ would be $\hat{B} = (\sum_{h=1}^{k} \sum_{j \in s_h} \lambda_j)^{-1} \sum_{h=1}^{k} (\sum_{j \in s_h} \lambda_j) \hat{B}_h$.

Because the $\hat{\eta}_i$ values are subject to estimation error, the terms $\hat{B}_h$ and $\hat{B}$ give only a partial indication of potential bias problems. For example, a large value of $\hat{B}_h$ may reflect a substantial bias in $\hat{\bar{Y}}_{hR}$, or may reflect biases in the alternative estimator $\bar{Y}_{h\eta}$ due to the errors $\hat{\eta}_i - \eta_i$; *cf.* the cautionary remarks in Little (1986, p. 146) regarding direct use of the weights $\hat{\eta}_i^{-1}$ in adjusted estimation of $\bar{Y}$. Thus, if one observes a large value of $\hat{B}_h$, it is worthwhile to consider refinement of cell h; but the final decision of whether to use the resulting refined set of cells will depend on whether the refined set produces a substantially different estimate of the overall mean $\bar{Y}$.

Tables 2 and 3 present $\hat{B}_h$ values and associated standard errors and t statistics for equal-quantile-division cells with $k = 3$ and $k = 5$, respectively. Note that for the case $k = 3$, the $\hat{B}_h$ diagnostics indicate a possible bias contribution from the lowest cell. This is consistent with the suggestion from Section 3.1 that $k = 3$ cells may not provide a satisfactory nonresponse adjustment. In addition, the corresponding value of $\hat{B}$ was 111, with a standard error of 75; this value of $\hat{B}$ is very close to the difference $\hat{\bar{Y}}_3 - \hat{\bar{Y}}_5 = 106$ of the estimates $\hat{\bar{Y}}_3$ and $\hat{\bar{Y}}_5$ from Table 1.

Table 2
Within-Cell $\hat{B}_h$ Statistics for Probability-Based Cells, $k = 3$

h	$\hat{B}_h$	$\mathrm{se}(\hat{B}_h)$	$t = \hat{B}_h/\mathrm{se}(\hat{B}_h)$
1	269	136	1.98
2	-19	43	-0.44
3	84	45	1.87

Table 3
Within-Cell $\hat{B}_h$ Statistics for Probability-Based Cells, $k = 5$

h	$\hat{B}_h$	$\mathrm{se}(\hat{B}_h)$	$t = \hat{B}_h/\mathrm{se}(\hat{B}_h)$
1	96	217	0.44
2	-72	116	-0.62
3	-52	56	-0.93
4	-16	27	-0.59
5	98	50	1.96

In light of the preceding results, the low-$\hat{\eta}_i$ cell from the $k = 3$ case was split in half. The upper bounds for the two new cells ($h = 1'$ and $h = 1''$, say) were determined by the

0.167 and 0.333 estimated quantiles of the $\hat{\eta}_i$ population. The resulting $\hat{B}_h$ values and standard errors were 90 and 197 for cell 1′, and −42 and 79 for cell 1″. In addition, the refined set of four cells had $\hat{B} = 30$, with a standard error of 75; and the adjusted estimate of $\bar{Y}$ equal to \$32,652 and standard error of \$518 were close to those obtained from the equal-quantile-division method with $k = 5$.

In contrast with the results for $k = 3$, the $\hat{B}_h$ results for $k = 5$ indicated relatively little basis for concern, with the possible exception of cell $h = 5$, which had a t statistic of 1.96. For $k = 5$, the value of $\hat{B}$ was 11, with a standard error of 93. Additional splitting of cell $h = 5$ did not lead to notable changes in either the estimate of $\bar{Y}$ or the associated standard errors. The $\hat{B}_h$ results, for equal-quantile-division cells with larger values of k showed even fewer indications of within-cell bias. For example, for $k = 6$ all six cells had $\hat{B}_h$ values with t statistics less than or equal to 1.65; and for $k = 10$, all cells had $\hat{B}_h$ values with t statistics less than or equal to 1.54.

3.2.2 Relation of Cell Widths to Precision of η_i Estimates

The relationship between the widths of adjustment cells and the widths of confidence intervals for the response probabilities η_i leads to another diagnostic for identification of potential problem cells. First, define $a_h = (\sum_{i\in s_h}\lambda_i R_i)^{-1} \sum_{i\in s_h}\lambda_i$, the nonresponse-adjustment factor used for responding units in cell h. Second, following standard results for logistic regression, note that an approximate 95% confidence interval for η_i is

$$(LB_i, UB_i) = ([1 + \exp\{-X_i'\hat{\theta} + 1.96D_i^{1/2}\}]^{-1}$$

$$[1 + \exp\{-X_i'\hat{\theta} - 1.96D_i^{1/2}\}]^{-1}),$$

where $\hat{\theta}$ is the vector of logistic regression parameter estimates, $D_i = X_i' \hat{V}_\theta X_i$, and $\hat{V}_\theta$ is the pseudo-replicate-based estimated covariance matrix for $\hat{\theta}$. Let $\bar{d}_h$ be the λ_i-weighted sample mean of the confidence interval widths $UB_i - LB_i$ for units i in cell h, and consider a comparison of $\bar{d}_h$ to the width of cell h. If cell h is relatively wide, both on an absolute scale and relative to $\bar{d}_h$, then division of this cell may produce two new cells with two substantially different weight factors a_h. Conversely, if $\bar{d}_h$ is substantially larger than the width of cell h, then differences among $\hat{\eta}_i$ in that cell may result more from estimation error than from differences in the true η_i. In that case, additional division of cell h is unlikely to produce much useful change in weight factors a_h; and thus there will be relatively little change in the resulting nonresponse-adjusted estimator of $\bar{Y}$.

Tables 4 and 5 report cell boundaries, cell widths, $\bar{d}_h$, and a_h values for $k = 5$ and $k = 10$, respectively. For $k = 5$, the widths of cells 2 through 5 were not large relative to the $\bar{d}_h$ values. Each of these cells is essentially split in half to produce the $k = 10$ cell case. The resulting pairs of a_h for $k = 10$ are relatively close to the corresponding a_h values in cells 2 through 5 with $k = 5$.

By contrast, with $k = 5$, cell 1 is over twice as wide as $\bar{d}_1$. When $k = 10$, this cell is divided into cells with somewhat different nonresponse adjustment weight factors a_h: 1.45 and 1.27, respectively. However, the corresponding cell-mean estimates are relatively close: $\bar{Y}_{1R} = \$24{,}045$ and $\bar{Y}_{2R} = \$24{,}582$ for $k = 10$. Thus, in this example, the nonresponse-adjusted estimates $\hat{\bar{Y}}_5$ and $\hat{\bar{Y}}_{10}$ are relatively close because four of the five cell divisions produced relatively small changes in weights, and because the other cell division produced two cells with similar cell means.

Table 4
Estimated-Probability Cell Boundaries, Cell Widths, Mean Confidence Interval Widths and Nonresponse Adjustment Factors, $k = 5$

h	Lower Bound	Upper bound	Cell Width	$\bar{d}_h$	a_h
1	0.384	0.810	0.426	0.197	1.35
2	0.810	0.861	0.051	0.139	1.20
3	0.861	0.894	0.033	0.110	1.13
4	0.894	0.924	0.030	0.088	1.08
5	0.924	0.994	0.070	0.067	1.07

Finally, the a_h factors in Table 5 indicate that mean response rates in the $k = 10$ cells fall in a moderate range, from $(1.45)^{-1} = 0.69$ to $(1.06)^{-1} = 0.94$. Some other nonresponse datasets involve a wider range, and thus are more likely to produce more pronounced cell-splitting results. Conversely, other nonresponse datasets may display a tighter distribution of response probabilities, and thus are less likely to display notable cell-splitting effects.

Table 5
Estimated-Probability Cell Boundaries, Cell Widths, Mean Confidence Interval Widths and Nonresponse Adjustment Factors, $k = 10$

h	Lower Bound	Upper Bound	Cell Width	$\bar{d}_h$	a_h
1	0.384	0.762	0.378	0.220	1.45
2	0.762	0.810	0.048	0.174	1.27
3	0.810	0.840	0.030	0.146	1.21
4	0.840	0.861	0.021	0.132	1.19
5	0.861	0.878	0.017	0.111	1.14
6	0.878	0.894	0.016	0.108	1.11
7	0.894	0.908	0.014	0.093	1.09
8	0.908	0.924	0.016	0.083	1.08
9	0.924	0.944	0.020	0.072	1.08
10	0.944	0.994	0.050	0.062	1.06

3.3 Comparison of Cell-Based Estimates to the Unadjusted Estimate

To conclude the assessment of $\hat{\eta}_i$-based cells, we compared the adjusted estimates $\hat{\bar{Y}}_k$ with the unadjusted estimate $\hat{\bar{Y}}_1$. First, Table 1 indicates that for the reported values of $k \geq 5$, the differences $\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k$ are greater than or equal to \$303. Second, for $k \geq 5$, the estimated standard errors of the differences $\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k$ are all less than or equal to \$138, and the corresponding t statistics are all greater than 2.44. Thus, for $k = 5$, say, a formal test of the hypothesis H_0: $E(\hat{\bar{Y}}_1 - \hat{\bar{Y}}_5) = 0$ would be rejected at standard significance levels; *i.e.*, the adjustment-cell method has produced a significant change in the mean income estimate.

In addition, a rough comparison of the efficiencies of $\hat{\bar{Y}}_1$ and $\hat{\bar{Y}}_k$ follows from the estimated mean squared error ratio

$$\hat{\gamma}_k = \{\hat{V}(\hat{\bar{Y}}_k)\}^{-1}[\hat{V}(\hat{\bar{Y}}_1) + \max\{0, (\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k)^2 - \hat{V}(\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k)\}]$$

where $\hat{V}(\hat{\bar{Y}}_1)$, $\hat{V}(\hat{\bar{Y}}_k)$, and $\hat{V}(\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k)$ are the pseudo-replicate-based variance estimates for the indicated means. To interpret this ratio, assume for the moment that $\hat{\bar{Y}}_k$ is an approximately unbiased estimator of $\bar{Y}$. Then $\hat{\gamma}_k$ is an estimator of the mean squared error of the unadjusted estimator $\hat{\bar{Y}}_1$, relative to the mean squared error of $\hat{\bar{Y}}_k$. Consequently, $\hat{\gamma}_k$ reflects the loss of efficiency incurred by using the biased, unadjusted estimator $\hat{Y}_1$ instead of the adjusted, unbiased estimator $\hat{\bar{Y}}_k$. However, this interpretation should be viewed with some caution, since it depends on the assumption that $\hat{\bar{Y}}_k$ is approximately unbiased for $\bar{Y}$, and since the $\hat{\gamma}_k$ are functions of the random terms $\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k$, $\hat{V}(\hat{\bar{Y}}_1)$, $\hat{V}(\hat{\bar{Y}}_k)$, and $\hat{V}(\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k)$.

As suggested by a referee, one could also consider a mean squared error ratio

$$\{\hat{V}(\hat{\bar{Y}}_\eta)\}^{-1}[\hat{V}(\hat{\bar{Y}}_k) + \max\{0, (\hat{\bar{Y}}_k - \hat{\bar{Y}}_\eta)^2 - \hat{V}(\hat{\bar{Y}}_k - \hat{\bar{Y}}_\eta)\}]$$

where $\hat{\bar{Y}}_\eta$ equals expression (1.1) with λ_i replaced by $(\hat{\eta}_i)^{-1}\lambda_i$. This would amount to comparing each cell-based estimate $\hat{\bar{Y}}_k$ to $\hat{\bar{Y}}_\eta$. This is appropriate if $\hat{\bar{Y}}_\eta$ is approximately unbiased, but this unbiasedness may be problematic in some cases; *cf.* Little (1986, p. 146).

The final column of Table 1 reports the estimated ratios $\hat{\gamma}_k$ for specified values of k. For $k \geq 5$, each reported $\hat{\gamma}_k$ is greater than 1.5. Finally, note that each adjusted estimate $\hat{\bar{Y}}_k$ fell *below* the unadjusted estimate $\hat{\bar{Y}}_1$. This occurred because, for a given k, cells associated with larger response probabilities tended to have larger mean estimates $\bar{Y}_{hR}$. For example, for $k = 5$, the $\bar{Y}_{hR}$ values were \$24,333, \$33,729, \$33,398, \$34,620, and \$37,057 for $h = 1$ (the low $\hat{\eta}_i$ cell) through $h = 5$ (the high $\hat{\eta}_i$ cell), respectively.

4. CELLS BASED ON ESTIMATED INCOME VALUES

The general diagnostic ideas of Section 3 also apply to $\hat{Y}_i$ based cells. To illustrate this idea, we fit separate weighted regressions of Y_i = reported income for second- and fifth-interview respondents. Yansaneh and Eltinge (1993) report details of the work, including parameter estimates and standard errors. The resulting regression models were used to compute estimated incomes $\hat{Y}_i$ for both complete and incomplete income reporters. Units were then grouped into cells according to their $\hat{Y}_i$ values, with cell boundaries determined by the equal-quantile method.

Table 6 reports the basic sensitivity-analysis and efficiency results for the $\hat{Y}_i$ based cells; the organization of this table is the same as in Table 1. The sensitivity-analysis results are qualitatively similar, but not identical, to those reported for the $\hat{\eta}_i$-based cells. In additional work not detailed here, we considered splitting individual equal-quantile $\hat{Y}_i$-based cells. For $k \geq 4$, the resulting mean estimates and associated standard errors did not differ notably from those reported in Table 6.

Table 6

Adjusted Estimates of Mean Income with Cell Boundaries Determined by Estimated Income Quantiles

Adjustment Method	Point Estimate	Standard Error	SE($\hat{\bar{Y}}_k - \hat{\bar{Y}}_1$)	MSE Ratio
Unadjusted ($k = 1$)	32,967	569	N/A	N/A
$k = 3$ cells	32,512	509	106	2.01
$k = 4$ cells	32,468	512	108	2.14
$k = 5$ cells	32,473	511	115	2.12
$k = 6$ cells	32,492	508	117	2.08
$k = 10$ cells	32,488	510	119	2.07
$k = 15$ cells	32,478	504	124	2.16
$k = 20$ cells	32,495	513	124	2.02

The final two columns of Table 6 permit comparison of $\hat{\bar{Y}}_k$ to the unadjusted estimate $\hat{\bar{Y}}_1$. For $k \geq 4$, the differences $\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k$ are greater than or equal to \$472, with estimated standard errors less than or equal to \$124. The associated t statistics are all greater than 3.80. In addition, the estimated mean squared error ratios $\hat{\gamma}_k$ are all greater than 2.0.

Also, the $\hat{\eta}_i$ and $\hat{Y}_i$-based cells produced somewhat different adjusted estimates of mean income, but the observed differences were not statistically significant at customary α levels. For example, with $k = 5$, the difference between the $\hat{\eta}_i$- and $\hat{Y}_i$-based cell estimates is \$32,630 - \$32,473 = \$157, with a standard error of \$122 and a t statistic of 1.29. Similarly, for $k = 10$, the difference between the $\hat{\eta}_i$- and $\hat{Y}_i$-based estimates is \$152, with a standard error of \$104. Thus, the data provide relatively little power to distinguish between results of the two general cell-formation methods.

Finally, note that a given set of $\hat{Y}_i$-based cells are fundamentally linked with a particular Y variable, *e.g.*, consumer unit income. Consequently, that set of cells will not necessarily work well for estimation of the mean of a different Y variable.

5. DISCUSSION

5.1 Summary of Methods

This paper has discussed some simple diagnostics for formation of nonresponse adjustment cells. The methodology may be summarized as follows.

1. Based on preliminary modeling work and observed auxiliary variables X_i, compute an estimated response probability $\hat{\eta}_i$ for each sample unit (respondents and nonrespondents).
2. Construct k adjustment cells with boundaries determined by the estimated $k^{-1}j$ quantiles of the $\hat{\eta}_i$ population, $j = 1, 2, ..., k - 1$. Compute the resulting adjusted mean estimate, $\hat{\bar{Y}}_k$.
3. Repeat (2) for several integers $k > 1$. As k increases, identify the point at which the $\hat{\bar{Y}}_k$ become approximately constant. In keeping with Rosenbaum and Rubin (1984) and the empirical results discussed here, values of k near 5 may be of special interest.
4. Use simple screening diagnostics (*e.g.*, $\hat{B}_h$ and $\bar{d}_h$ in Section 3.2) to check for potential problems in the equal-quantile-division adjustment cells. If the diagnostics identify potential "problem cells," then try additional refinements of these cells. Compute estimates of $\bar{Y}$ based on these refined sets of cells, and compare these new estimates to the $\hat{\bar{Y}}_k$ from (3).
5. Assess the overall effect of adjustment by comparing the differences $\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k$ to the standard errors $se(\hat{\bar{Y}}_1 - \hat{\bar{Y}}_k)$; and by computing the estimated mean squared error ratios $\hat{\gamma}_k$.
6. Repeat steps (1) through (5), as appropriate, for $\hat{Y}_i$-based adjustment cells. Compare the final estimates of $\bar{Y}$ obtained from the $\hat{\eta}_i$ and $\hat{Y}_i$-based cell methods.

5.2 Areas for Future Research

The results of this work suggest two potentially useful areas for future research. First, the CE income nonresponse problem is similar to nonresponse problems in some other large-scale surveys, but as with any case study one should not over-generalize the empirical results reported here. It would be useful to apply these diagnostics to problems involving different estimands (*e.g.*, cross-class means) or involving nonresponse datasets with somewhat different characteristics, *e.g.*, larger or smaller effective sample sizes; or wider or narrower distributions of $\hat{\eta}_i$ estimates. This in turn would offer additional insight into the operating characteristics of $\hat{\eta}_i$ and $\hat{Y}_i$-based adjustment cell methods in practical applications. Second, extensions to multivariate problems (*e.g.*, relationships involving second-interview and fifth-interview CE income data) also would be of interest.

ACKNOWLEDGEMENTS

The authors thank Richard Dietz, Thesia Garner, Paul Hsen, Eva Jacobs, Geoffrey Paulin, Stuart Scott, and Stephanie Shipp for many helpful discussions of the Consumer Expenditure Survey; and Wayne Fuller, Steve Miller, Geoff Paulin, Stuart Scott, three referees and the editor for helpful comments on earlier versions of this paper. This work was carried out while the authors were visiting the Bureau of Labor Statistics through the ASA/NSF/BLS Research Fellow Program, and was supported by a grant from the National Science Foundation (SES-9022443). Eltinge's research was also supported in part by a grant from the National Institutes of Health (CA 57030-04). The views expressed in this paper are those of the authors and do not necessarily represent the policies of the Bureau of Labor Statistics.

REFERENCES

CASSEL, C.-M., SÄRNDAL, C.-E., and WRETMAN, J.H. (1983). Some uses of statistical models in connection with the nonresponse problem. In *Incomplete Data in Sample Surveys*, (Vol. 3), (Eds. W.G. Madow, I. Olkin, and D. Rubin). New York: Academic Press, 143-160.

COCHRAN, W.G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, 24, 205-213.

COCHRAN, W.G. (1977). *Sampling Techniques*. New York: Wiley.

CZAJKA, J.L., HIRABAYASHI, S.M., LITTLE, R.J.A., and RUBIN, D.B. (1992). Projecting from advance data using propensity modeling: An application to income and tax statistics. *Journal of Business and Economic Statistics*, 10, 117-131.

DAVID, M.H., LITTLE, R.J.A., SAMUHEL, M., and TRIEST, R. (1983). Imputation models based on the propensity to respond. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 168-173.

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.

EZZATI, T., and KHARE, M. (1992). Nonresponse adjustments in a national health survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 339-344.

GARNER, T.I., and BLANCIFORTI, L.A. (1994). Household income reporting: An analysis of U.S. Consumer Expenditure Survey data. *Journal of Official Statistics* 10, 69-91.

GÖKSEL, H., JUDKINS, D.R., and MOSHER, W.D. (1991). Nonresponse adjustments for a telephone follow-up to a national in-person survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 581-586.

KALTON, G., and MALIGALIG, D.S. (1991). A comparison of methods of weighting adjustment for nonresponse. *Proceedings of the 1991 Annual Research Conference*, U.S. Bureau of the Census, 409-428.

LEPKOWSKI, J., KALTON, G., and KASPRZYK, D. (1989). Weighting adjustments for partial nonresponse in the 1984 SIPP panel. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 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.

OH, H.L., and SCHEUREN, F.J. (1983). Weighting adjustment for unit nonresponse. In *Incomplete Data in Sample Surveys*, (Vol. 2), (Eds. W.G. Madow, I. Olkin and D.B. Rubin). New York: Academic Press, 143-184.

RAO, J.N.K. (1996). On variance estimation with imputed survey data. *Journal of the American Statistical Association*, 91, 499-506.

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.

ROSENBAUM, P.R., and RUBIN, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79, 516-524.

TREMBLAY, V. (1986). Practical criteria for definition of weighting classes. *Survey Methodology*, 12, 85-97.

UNITED STATES BUREAU OF LABOR STATISTICS (1991). News: Consumer Expenditures in 1990. Publication USDL 91-607, United States Department of Labor, Washington, DC.

UNITED STATES BUREAU OF LABOR STATISTICS (1992). BLS Handbook of Methods. Bulletin 2414, United States Department of Labor, Washington, DC.

WOLTER, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

YANSANEH, I.S., and ELTINGE, J.L. (1993). Construction of adjustment cells based on surrogate items or estimated response propensities. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 538-543.

ZIESCHANG, K.D. (1990). Sample weighting methods and estimation of totals in the Consumer Expenditure Survey. *Journal of the American Statistical Association*, 85, 986-1001.