Survey Methodology, June 1986
Vol. 12, No. 1, pp. 1–16
Statistics Canada

The Treatment of Missing Survey Data

GRAHAM KALTON and DANIEL KASPRZYK[1]

ABSTRACT

Missing survey data occur because of total nonresponse and item nonresponse. The standard way to attempt to compensate for total nonresponse is by some form of weighting adjustment, whereas item nonresponses are handled by some form of imputation. This paper reviews methods of weighting adjustment and imputation and discusses their properties.

KEY WORDS: Nonresponse; Item nonresponse; Weighting adjustments; Imputation.

1. INTRODUCTION

Surveys typically collect responses to a large number of items for each sampled element. The problem of missing data occurs when some or all of the responses are not collected for a sampled element or when some responses are deleted because they fail to satisfy edit constraints. It is common practice to distinguish between total (or unit) nonresponse, when none of the survey responses are available for a sampled element, and item nonresponse, when some but not all of the responses are available. Total nonresponse arises because of refusals, inability to participate, not-at-homes, and untraced elements. Item nonresponse arises because of item refusals, "don't knows", omissions and answers deleted in editing.

This paper reviews the general-purpose methods available for handling missing survey data. The distinction between total and item nonresponse is useful here since different adjustment methods are used for these two cases. In general the only information available about total nonrespondents is that on the sampling frame from which the sample was selected (e.g., the strata and PSUs in which they are located). The important aspects of this information can usually be readily incorporated into weighting adjustments that attempt to compensate for the missing data. Hence as a rule weighting adjustments are used for total nonresponse. Methods for making weighting adjustments are reviewed in Section 2.

In the case of item nonresponse, however, a great deal of additional information is available for the elements involved: not only the information from the sampling frame, but also their responses for other survey items. In order to retain all survey responses for elements with some item nonresponses, the usual adjustment procedure produces analysis records that incorporate the actual responses to items for which the answers were acceptable and imputed responses for other items. Imputation methods for assigning answers for missing responses are reviewed in Section 3.

In general the choice between weighting adjustments and imputation for handling missing survey data is fairly clearcut; there are cases, however, when the choice is not so clear. These are cases of what may be termed partial nonresponse, when some data are collected for a sampled element but a substantial amount of data is missing. Partial nonresponse can arise, for instance, when a respondent terminates an interview prematurely, when data are not obtained for one or more members of an otherwise cooperating household (for household level analysis), or when a sampled individual provides data for some but not all waves of a panel survey. Discussions of the choice between weighting and imputation to compensate for wave nonresponse in a panel survey are given by Cox and Cohen (1985) and Kalton (1986).

[1] Graham Kalton, Survey Research Center, University of Michigan, Ann Arbor, Michigan, 48106-1248 and Daniel Kasprzyk, Population Division, U.S. Bureau of the Census, Washington, D.C., 20233. The authors would like to thank the referees for their helpful comments.

Although weighting adjustments and imputation are treated as separate approaches in the discussion below, they are in fact closely related. The relationship and differences between the two approaches are briefly discussed in Section 4, which also mentions some alternative ways of handling missing survey data.

2. WEIGHTING ADJUSTMENTS

Weighting adjustments are primarily used to compensate for total nonresponse. The essence of all weighting adjustment procedures is to increase the weights of specified respondents so that they represent the nonrespondents. The procedures require auxiliary information on either the nonrespondents or the total population. The following four types of weighting adjustments are briefly reviewed below: population weighting adjustments, sample weighting adjustments, raking ratio adjustments, and weights based on response probabilities. More details are provided in Kalton (1983).

2.1 Population Weighting Adjustments

The auxiliary information used in making population weighting adjustments is the distribution of the population over one or more variables, such as the population distribution by age, sex and race available from standard population estimates. The sample of respondents is divided into a set of classes, termed here weighting classes, defined by the available auxiliary information (e.g., White males aged 15-24, non-White females aged 25-34, etc.). The weights of all respondents within a weighting class are then adjusted by the same multiplying factor, with different factors in different classes. The adjustment is carried out in such a way that the weighted respondent distribution across the weighting classes conforms to the population distribution.

This type of adjustment is often termed poststratification. That term is avoided here, however, because although population weighting resembles poststratification, there is an important difference between the two. Like population weighting, poststratification weights the sample to make the sample distribution conform to the population distribution across a set of classes (or strata). However, the standard textbook theory of poststratification is concerned only with the sampling fluctuations that cause the sample distribution to deviate from the population distribution, not with the more major deviations that can arise from varying response rates across the classes. Poststratification adjustments are more like a fine tuning of the sample, resulting generally in only small variations in the weights across strata. In consequence, provided that the strata are not small, poststratification leads to lower standard errors for the survey estimates. In contrast, population weighting adjustments may involve more major adjustments and result in higher standard errors.

Population weighting adjustments attempt to reduce the bias created by nonresponse and coverage errors. Consider the estimation of a population mean $\bar{Y}$ from a sample in which the elements are selected with equal probability. Suppose that the population is divided into a set of weighting classes, with a proportion W_h of elements in class h. Assume that respondents always respond and that nonrespondents never do. Let R_h and M_h be the proportions of respondents and nonrespondents respectively in class h, and let $\bar{R} = \Sigma W_h R_h$ be the overall response rate. Then, following Thomsen (1973), the bias of the unadjusted respondent mean $(\bar{y})$ can be expressed as

$$B(\bar{y}) = \bar{R}^{-1}\sum W_h(\bar{Y}_{rh} - \bar{Y}_r)(R_h - \bar{R}) + \sum W_h M_h(\bar{Y}_{rh} - \bar{Y}_{mh}) = A + B \qquad (1)$$

where $\bar{Y}_{rh}$ and $\bar{Y}_{mh}$ are the means for respondents and nonrespondents in class h respectively, and $\bar{Y}_r$ is the population mean for the respondents. The use of the population weighting adjustment leads to the weighted sample mean, $\bar{y}_p = \Sigma W_h \bar{y}_{rh}$, where $\bar{y}_{rh}$ is the respondent sample mean in class h. The bias of $\bar{y}_p$ is simply the second term in $B(\bar{y})$, that is, $B(\bar{y}_p) = B$.

If A and B are of the same sign, the population weighting adjustment reduces the absolute bias in the estimate of $\bar{Y}$ by $|A|$. If $\bar{Y}_{rh} = \bar{Y}_{mh}$, as occurs in expectation when the nonrespondents are missing at random within the weighting classes, then $B = 0$. In this case, the population weighting adjustment eliminates the bias. The term A is a covariance-type term between the class response rates and the class respondent means. It is zero if either the response rates or the respondent means do not vary between classes. In either of these cases, the population weighting adjustment has no effect on the bias of the estimator. It may be noted that population weighting adjustments may increase the absolute bias of the estimate of $\bar{Y}$. This will occur when A and B are of opposite signs and $|A| < 2|B|$.

Population weighting adjustments require external data on the population distributions for the variables to be used. Care is needed to ensure that the data on which the population distributions are based are exactly comparable with the survey data; otherwise, inappropriate weights will result. Since the procedure weights up to population distributions, it does more than just attempt to compensate for nonresponse. It also compensates for coverage errors and makes a poststratification adjustment.

2.2 Sample Weighting Adjustments

As with population weighting adjustments, with sample weighting adjustments the sample is divided into weighting classes; varying weights are then assigned to these classes in an attempt to reduce the nonresponse bias. The essential difference between the two procedures lies in the auxiliary information used. As described above, population weighting adjustments are based on externally obtained population distributions. No data are needed for the sample nonrespondents. In contrast, sample weighting adjustments employ only data internal to the sample and require information about the nonrespondents.

With sample weighting adjustments, the nonresponse adjustment weights for the weighting classes are made proportional to the inverses of the response rates in the classes. In order to compute these response rates, the numbers of respondents and nonrespondents in the classes must be determined. It is therefore necessary to know to which class each respondent and nonrespondent belongs. Since typically very little information about the nonrespondents is available, the choice of weighting class is usually severely restricted. It is often limited to general sample design variables (e.g., PSUs and strata), characteristics of those variables (e.g., urban/rural, geographical region), and sometimes some additional variables available on the sampling frame. On occasion it may also be possible to collect information on one or two variables for the nonrespondents, for instance by interviewer observation.

As population weighting adjustments resemble poststratification, so sample weighting adjustments resemble two-phase sampling. The first phase sample is the total sample of respondents and nonrespondents; the second phase sample is the subsample of respondents, selected with different sampling fractions (response rates) in different strata (weighting classes). The sample weighted mean can be represented by $\bar{y}_s = \Sigma w_h \bar{y}_{rh}$, where w_h is the proportion of the total sample in weighting class h. Assuming no coverage errors, $E(w_h) \doteq W_h$, the population proportion in class h, as used in the population weighted estimator

$\bar{y}_p = \Sigma W_h \bar{y}_{rh}$. The bias of $\bar{y}_s$ is the same as that of $\bar{y}_p$, namely $B(\bar{y}_s) = B$ as given in equation (1); hence the effect of the sample weighting adjustment on the bias of the survey estimate is the same as that of the population weighting adjustment. Since sample weighting adjustments use only data for the sample, they do not compensate for coverage errors (unlike population weighting adjustments).

Population and sample weighting adjustments have different data requirements, and hence address different potential sources of bias. In practice the two forms of adjustment are used in combination. Generally sample weighting adjustments are applied first, and then population weighting adjustments are applied afterwards. A common approach is initially to determine the sample weights needed to compensate for unequal selection probabilities, next to revise these weights to compensate for unequal response rates in different sample weighting classes (e.g., urban/rural classes within geographical regions), and finally to revise the weights again to make the weighted sample distribution for certain characteristics (e.g., age/sex) conform to the known population distribution for those characteristics. The use of this approach in the U.S. Current Population Survey is described by Bailar *et al.* (1978).

As with population weighting adjustments, the aim of sample weighting adjustments is to reduce the bias that nonresponse may cause in survey estimates. An effect of sample weighting adjustments is, however, to increase the variances of the survey estimates. There is therefore a trade-off to be made between bias reduction and variance increase.

An indication of the amount of increase in variance from weighting can be obtained by considering the situation where the element variances within the weighting classes are all the same and the variances between the class means are negligible compared to the within-class variances. In this situation, the loss of precision from weighting is approximately the same as that arising from the use of disproportionate stratified sampling when proportionate stratified sampling is optimum; Kish (1965, Section 11.7C; 1976) discusses this latter case.

Under the above conditions, weighting increases the variance of a sample mean by approximately $L = (\Sigma W_h k_h)(\Sigma W_h / k_h)$, where W_h is the proportion of the population and k_h is the weight for class h. An alternative expression for L is $(\Sigma n_h)(\Sigma n_h k_h^2) / (\Sigma n_h k_h)^2$, where n_h is the sample size in class h. The factor L becomes large when the variance of the weights is large.

A large variance in the weights can arise from segmenting the sample into many weighting classes with only a few sampled elements in each. When the weighting classes are small, their response rates are unstable, and this gives rise to a large variation in the weights. To avoid this effect, it is common practice to limit the extent to which the sample is segmented. Even so, there may still be some weighting classes that require large weights. Sometimes these weighting classes are handled by collapsing them with adjacent ones and sometimes their weights are cut back to some acceptable maximum value (see Bailar *et al.* 1978 and Chapman *et al.* 1986, for examples). These procedures avoid the increase in variance associated with the use of extreme weights, but they may lead to increased bias; their effect on the bias is, however, unknown.

In some cases it seems desirable to use several auxiliary variables in forming the weighting classes for population or sample weighting adjustments. However, if the classes are formed by taking the full crossclassification of the variables, there will be a large number of weighting classes. Unless the sample is very large, the sample sizes in the resultant weighting classes will be small, and the instability in the response rates will lead to a large variance in the weights and loss of precision in the survey estimates. One way to deal with this problem is to cut down on the number of classes by collapsing cells, for instance by discarding some of the auxiliary variables or using coarser classifications. Another way is to base the weights on a model, as is done in raking ratio weighting discussed below.

2.3 Raking Ratio Adjustments

When weighting classes are taken to be the cells in the crossclassification of the auxiliary variables, population weighting adjustments make the joint distribution of the auxiliary variables in the sample conform to that in the population. Similarly, sample weighting adjustments make the joint distribution of the auxiliary variables in the respondent sample conform to that in the total sample. As noted above, however, this crossclassification approach may have the undesirable effect of creating many small, and hence unstable, weighting classes. Also, it is not always possible to employ this approach with population weighting adjustments: in many cases the population marginal distributions, and perhaps some bivariate distributions, of the auxiliary variables are available, but the full joint distribution is unknown.

An alternative approach is to develop weights that make the marginal distributions of the auxiliary variables in the sample conform to marginal population distributions (with population weighting) or marginal total sample distributions (with sample weighting), without ensuring that the full joint distribution conforms. The method of raking ratio estimation, or raking, may be used to obtain weights that satisfy these conditions. Raking corresponds to iterative proportional fitting in contingency table analysis (see, for instance, Bishop *et al.*, 1975).

Consider the use of raking in the simple case of two auxiliary variables. Let W_{hk} be the proportion of the population in the (h, k)-th cell of the crossclassification, and let $\tilde{w}_{hk}$ be the proportion assigned to that cell by the raking algorithm. Conditional on the total and respondent sample sizes in the cells (and assuming all cells have at least one respondent), the bias of the raking ratio adjusted sample mean $\bar{y}_q = \Sigma\Sigma\tilde{w}_{hk}\bar{y}_{hk}$ is

$$B(\bar{y}_q) = \sum\sum W_{hk}M_{hk}(\bar{Y}_{rhk} - \bar{Y}_{mhk}) + \sum\sum(\tilde{W}_{hk} - W_{hk})(\bar{Y}_{rhk} - \bar{Y}_{rh.} - \bar{Y}_{r.k} + \bar{Y}_r)$$

where $\tilde{W}_{hk} = E(\tilde{w}_{hk})$. The first term in this bias corresponds to the bias term B in equation (1) for the population and sample weighting adjustments. It is zero in expectation if the cell nonrespondents are random subsets of the cell populations. The second term is zero if either $\tilde{W}_{hk} = W_{hk}$ or there is no interaction in the $\bar{Y}_{rhk}$ for this classification.

Underlying the raking ratio weighting procedure is a logit model for the cell response rates. With the model $\ln[R_{hk}/(1 - R_{hk})] = \alpha_h + \beta_k$ for the response rates in a two-way classification, $\tilde{W}_{hk} = W_{hk}$. Thus, under this model, the second term in $B(\bar{y}_q)$ is zero.

Further discussion of raking ratio weighting is given by Oh and Scheuren (1978a,1978b, 1983). Oh and Scheuren (1978a) also provide a bibliography on raking.

2.4 Weighting with Response Probabilities

Although a number of methods for weighting with response probabilities have been proposed, this approach has not been widely adopted as an adjustment procedure. The basis of the approach is to assume that all population elements have probabilities (usually required to be non-zero) of responding to the survey. Some method is used to estimate the response probabilities for responding elements. These elements are then given nonresponse adjustment weights that are in inverse proportion to their estimated response probabilities.

An early application of this approach is the well-known procedure of Politz and Simmons (1949, 1950). A single (evening) call is made to each selected household, and during the course of the interview respondents are asked on how many of the previous five evenings they were at home at about the same time. Their response probabilities are then taken to be the fraction of the six evenings (including the one of the interview) that they were at home, and the inverses of these probabilities are used in the analysis. Note that the procedure does not deal with those who were out on all six evenings and those who refused.

Another approach for estimating response probabilities is to regress response status (1 for respondents, 0 for nonrespondents) on a set of variables available for both respondents and nonrespondents, using a logistic or probit regression. The predicted values from the regression for the respondents are then taken to be their response probabilities, and weights in inverse proportion to these predicted values are used in the analysis. A special case is when the predictor variables are dummy variables that identify a set of classes. The predicted response probabilities are then the class response rates, and the method reduces to a sample weighting adjustment. The method is most appropriate for situations where a good deal of information is available for the nonrespondents, as for instance when the nonrespondents are losses after the first wave of a panel survey. Little and David (1983) discuss the application of the method for panel nonresponse. It should be noted that if the regression is highly predictive of response status, the resultant weights will vary markedly, leading to a substantial loss in the precision of the survey estimates.

Drew and Fuller (1980, 1981) describe an approach for estimating response probabilities from the number of respondents secured at successive calls. In their model, the population is divided into classes. Within each class, every element is assumed to have the same response probability which remains the same at each call. The model also allows for a proportion of hard-core nonrespondents that is assumed constant across classes. Under these assumptions, the response probabilities for each class and the proportion of hard-core nonrespondents can be estimated, and hence weighting adjustments can be made. Thomsen and Siring (1983) adopt a similar approach using a more complex model.

Finally, mention should be made of a related approach that compensates for nonresponse by weighting up difficult-to-interview respondents. Bartholomew (1961), for instance, proposed making only two calls in a survey, and weighting up the respondents at the second call to represent the nonrespondents. The assumption behind this approach is that the nonrespondents are like the late respondents. This assumption seems questionable, however, and empirical evidence from an intensive follow-up study of nonrespondents in the U.S. Current Population Survey does not support it (Palmer and Jones 1966; Palmer 1967).

3. IMPUTATION

A wide variety of imputation methods has been developed for assigning values for missing item responses. The aim here is to provide a brief overview of the methods, the basic differences between them, and some of the issues involved in imputation. A fuller treatment is provided by Kalton and Kasprzyk (1982).

Imputation methods can range from simple *ad hoc* procedures used to ensure complete records in data entry to sophisticated hot-deck and regression techniques. The following are some common imputation procedures:

(a) *Deductive imputation*. Sometimes the missing answer to an item can be deduced with certainty from the pattern of responses to other items. Edit checks should check for consistency between responses to related items. When the edit checks constrain a missing response to only one possible value, deductive imputation can be employed. Deductive imputation is the ideal form of imputation.

(b) *Overall mean imputation*. This method assigns the overall respondent mean to all missing responses.

(c) *Class mean imputation*. The total sample is divided into classes according to values of the auxiliary variables being used for the imputation (comparable to weighting classes). Within each imputation class the respondent class mean is assigned to all missing responses.

(d) *Random overall imputation.* A respondent is chosen at random from the total respondent sample, and the selected respondent's value is assigned to the nonrespondent. This method is the simplest form of hot-deck imputation, that is an imputation procedure in which the value assigned for a missing response is taken from a respondent to the current survey.

(e) *Random imputation within classes.* In this hot-deck method, a respondent is chosen at random within an imputation class, and the selected respondent's value is assigned to the nonrespondent.

(f) *Sequential hot-deck imputation.* The term sequential hot-deck imputation is used here to describe the procedure used with the labor force items in the U.S. Current Population Survey (Brooks and Bailar 1978). The procedure starts with a set of imputation classes. A single value for the item subject to imputation is assigned for each class (perhaps taken from a previous survey). The records in the survey's data file are then considered in turn. If a record has a response for the item in question, its response replaces the value stored for the imputation class in which it falls. If the record has a missing response, it is assigned the value stored for its imputation class.

The hot-deck method is similar to random imputation within classes. If the order of the records in the data file were random, the two methods would be equivalent, apart from the start-up process. The non-random order of the list generally acts to the benefit of the hot-deck method since it gives a closer match of donors and recipients provided that the file order creates positive autocorrelation. The benefit is, however, unlikely to be substantial.

The sequential hot-deck suffers the disadvantage that it may easily make multiple uses of donors, a feature that leads to a loss of precision in survey estimates. Multiple use of a donor occurs when, within an imputation class, a record with a missing response is followed by one or more other records with missing responses. The number of imputation classes that can be used with the method also has to be limited in order to ensure that donors are available within each class.

Useful discussions of the sequential hot-deck method are provided by Bailar *et al.* (1978), Bailar and Bailar (1978, 1983), Ford (1983), Oh and Scheuren (1980), Oh *et al.* (1980), and Sande (1983).

(g) *Hierarchical hot-deck imputation.* The above disadvantages of the sequential hot-deck are avoided in the hierarchical hot-deck method, a form of hot-deck imputation developed for the items in the March Income Supplement of the Current Population Survey. The procedure sorts respondents and nonrespondents into a large number of imputation classes from a detailed categorization of a sizeable set of auxiliary variables. Nonrespondents are then matched with respondents on a hierarchical basis, in the sense that if a match cannot be made in the initial imputation class, classes are collapsed and the match is made at a lower level of detail. Coder (1978) and Welniak and Coder (1980) provide further details on the hierarchical hot-deck procedure.

(h) *Regression imputation.* This method uses respondent data to regress the variable for which imputations are required on a set of auxiliary variables. The regression equation is then used to predict the values for the missing responses. The imputed value may either be the predicted value, or the predicted value plus some residual. There are several ways in which the residual may be obtained, as discussed later.

(i) *Distance function matching.* This hot-deck method assigns a nonrespondent the value of the "nearest" respondent, where "nearest" is defined in terms of a distance function for the auxiliary variables. Various forms of distance function have been proposed (e.g., Sande 1979; Vacek and Ashikago 1980), and the function can be constructed to reduce the multiple use of donors by incorporating a penalty for each use (Colledge *et al.* 1978).

Although at first sight these may appear a diverse set of procedures, they can nearly all be fitted within a single unifying framework. The methods can all be described, at least approximately, as special cases of the general regression model

$$\hat{y}_{mi} = b_{ro} + \sum b_{rj} z_{mij} + \hat{e}_{mi} \tag{2}$$

where $\hat{y}_{mi}$ is the imputed value for the ith record with a missing y value, z_{mij} are values reflecting the auxiliary variables for that record, b_{ro} and b_{rj} are the regression coefficients for the regression of y on x for the respondents, and $\hat{e}_{mi}$ is a residual chosen according to a specified scheme for the particular imputation method.

Equation (2) represents the regression imputation method in an obvious way. If the $\hat{e}_{mi}$'s are set at zero, then the imputed value is the predicted value from the regression; otherwise a residual of some form may be added. The equation also represents class mean imputation by defining the z_j's to be dummy variables that represent the classes, and setting $\hat{e}_{mi} = 0$. The regression equation then reduces to $\hat{y}_{mi} = \bar{y}_{rh}$, the class mean. Random imputation within classes is obtained by adding a residual to the class mean, where the residual is the deviation from the class mean for one of the respondents. Then $\hat{y}_{mi} = \bar{y}_{rh} + e_{rhk}$, where e_{rhk} is the deviation for respondent k in class h; this reduces to $\hat{y}_{mi} = y_{rhk}$, the value for that respondent. The sequential and hierarchical hot-deck methods resemble the random within class method. The overall mean and random overall imputation methods are degenerate cases of the class mean and random within class methods that use no auxiliary information.

An important consideration in the choice of imputation method is the type of variable being imputed. All the above methods can be applied routinely with continuous variables, but some of them are not suitable for use with categorical or discrete variables (such as being a member of the labor force (1) or not (0), and the number of completed years of education). Overall mean, class mean, and regression imputations impute values like 0.7 for being a member of the labor force (i.e., a 70% chance) and 10.7 for the number of completed years of education. These values are not feasible for individual respondents, and rounding them to whole numbers leads to bias. For this reason, these imputation methods do not work well for categorical and discrete variables. A notable advantage of all hot-deck methods is that they always give feasible values since the values are taken from respondents.

There are two major distinguishing features of the above imputation methods that deserve elaboration: whether or not a residual is added and, if one is, the form of the residual; and whether the auxiliary information is used in dummy variable form to represent classes or whether it is used straightforwardly in the regression. These features are discussed in the next two subsections. Other issues arising with the use of imputation are then discussed in subsequent subsections.

3.1 Choice of Residuals

Imputation methods may be classified as deterministic or stochastic according to whether the $\hat{e}_{mi}$'s are set at zero or not. For each deterministic imputation method, there is a stochastic counterpart. Let $\hat{y}_{mid}$ be the value imputed by the deterministic method and $\hat{y}_{mis} = \hat{y}_{mid} + \hat{e}_{mi}$ be that imputed by the corresponding stochastic method. Then $E_2(\hat{y}_{mis}) = \hat{y}_{mid}$, where E_2 denotes expectation over the sampling of residuals given the initial sample, provided that $E_2(\hat{e}_{mi}) = 0$ (as generally applies).

The choice between a deterministic and the corresponding stochastic imputation method depends on the form of survey analysis to be conducted. Consider first the estimation of the population mean of the y-variable using the sample mean of the respondents' values and the nonrespondents' imputed values. As Kalton and Kasprzyk (1982) show, given that $E_2(\hat{y}_{mis}) = \hat{y}_{mid}$, it follows that the expectation of the sample mean is the same whether the deterministic method or the corresponding stochastic method is used. Thus both methods have the same effect on the bias of the estimate. However, the addition of random residuals in the stochastic method causes a loss of precision in the sample mean. Although this loss can be controlled by the choice of a suitable method of sampling residuals (Kalton and Kish 1984), nevertheless some loss in precision occurs. For this reason a deterministic scheme is preferable for the purpose of estimating the population mean.

Consider now the estimation of the element standard deviation and distribution of the y-variable. Deterministic imputation methods fare badly for these purposes, since they cause an attenuation in the standard deviation and they distort the shape of the distribution. This may be simply illustrated in terms of the class mean imputation method. By assigning the class mean to all the missing values in a class, the shape of the distribution is clearly distorted with a series of spikes at the class means. The standard deviation of the distribution is attenuated because the imputed values reflect only the between-class and not the within-class variance. The appeal of the stochastic imputation methods is that the residual term captures the within-class (or residual) variance, and hence avoids the attenuation of the element standard deviation and the distortion of the distribution.

Since some survey analyses are likely to involve the distributions of the variables, stochastic imputation methods like the hot-deck methods are generally preferred. Once a decision is made to use a stochastic method, the question of how to choose the residuals arises. If the standard regression assumptions are accepted, the residuals could be chosen from a normal distribution with a mean of zero and a variance equal to the residual variance from the respondent regression. However, this places complete reliance on the model. An alternative that avoids the normality assumption is to choose the residuals randomly from the empirical distribution of the respondents' residuals. Another alternative is to select a residual from a respondent who is a "close" match to the nonrespondent, measuring "close" in terms of similar values on the auxiliary variables. This attractive alternative avoids the assumption of homoscedasticity and guards against misspecification of the distribution of the residual term. In the limit, the closest respondent is one who has the same values of all the auxiliary variables as the nonrespondent. In this case, the nonrespondent is given one of the matched respondents' values. This case arises with hot-deck methods, where nonrespondents and respondents are matched in terms of the auxiliary variables, and nonrespondents are assigned values from matched respondents.

A further consideration in the choice of residuals is to make the imputed values feasible ones. As noted above, deterministic methods may impute values for categorical and discrete variables that are not feasible. Some stochastic methods solve this problem through the allocation of the residuals. In particular, the use of respondents' residuals with the random within class and the sequential and hierarchical hot-deck methods ensures that the imputed values are feasible ones.

3.2 Imputation Class or Regression Imputation

As noted earlier, both imputation class and regression imputation methods fall within the imputation model given by equation (2). The difference between them lies in the ways in which they employ the auxiliary variables.

Imputation class methods divide the sample into a set of classes. For this purpose, continuous auxiliary variables have to be categorized. There is complete flexibility in the way the classes are formed, and the symmetrical use of the auxiliary variables in different parts of the sample is not required. Thus, for instance, in imputing for hourly rate of pay in a sample of employees, the sample might first be divided into two parts, union members and nonmembers; then the imputation classes for the members might be formed in terms of age and occupation whereas those for nonmembers might be formed in terms of sex and industry. As a rule, the aim is to construct classes of adequate size that explain as much of the variance in the variable to be imputed as possible. When the classes are formed by a complete crossclassification of the auxiliary variables, the underlying model contains all main effects and all interactions for the crossclassification. The limitation of imputation class methods is that the number of classes formed has to be constructed to ensure that there is some minimum number of respondents in each class. The hierarchical hot-deck method attempts to extend the amount of auxiliary data used, but even with this method matches of respondents and nonrespondents often cannot be made at the finer levels of detail. Coupled with the use of a random respondent residual within a class, imputation class methods have the valuable property that imputed values are feasible ones: that is, the imputed values are actual respondents' values.

Regression imputation methods have an advantage over imputation class methods in the number and in the level of detail of the auxiliary variables they can employ. Age can, for instance, be taken as a continuous variable rather than being categorized into a few classes. The regression model allows more main effects to be included in the model, but at the price of fewer interactions. Regression models can, of course, include some interactions, but they need to be specified. The models can also include polynomial terms and employ transformations, but again they need to be specified. The regression model has the potential of providing better predictions for the imputed values, but to achieve this careful modelling is required. Careful imputation modelling is unrealistic for all the variables in a survey, but it may be feasible for one or two major ones (and especially so for continuous surveys). Without careful modelling, there is a serious risk of poor imputations, although as noted earlier, this risk can be reduced by the allocation of random residuals from "close" respondents.

If a regression imputation assigns the residual from a respondent with exactly the same values of the auxiliary variables, the imputed value is necessarily a feasible one. If, however, there is even a small difference between the respondent's and nonrespondent's values on the auxiliary variables, the imputed value may not be feasible. A variant of regression imputation that avoids this problem, termed predictive mean matching, is described by Little (1986b) (Little attributes the method to Rubin). With predictive mean matching, the nonrespondent is matched to the respondent with the closest predicted value. Then, instead of adding the respondent's residual to the nonrespondent's predicted value, the nonrespondent is assigned the respondent's value. The method is thus a hot-deck method, and is similar to distance function matching.

The choice between imputation class and regression imputation methods should in part depend on the efforts made to develop the regression model. Unless adequate resources are devoted to the development of a regression model, the imputation class methods may be safer. The choice should also in part depend on the sample size. With large samples, hot-deck methods are likely to be able to use enough classes to take advantage of all the major predictor variables; however, with small samples this may not hold, and regression methods may have greater potential. David *et al.* (1986) describe an interesting study that compares regression models for imputing wages and salary in the U.S. Current Population Survey with hierarchical hot-deck imputations. Despite the extensive efforts made to develop the regression models, the hot-deck imputations were not found to be inferior in this large sample.

3.3 Effect of Imputation on Relationships

Although most of the literature on imputation deals with its effect on univariate statistics such as means and distributions, a large part of survey analysis is concerned with bivariate and multivariate relationships. Here the analysis of relationships can be considered in broad terms to include crosstabulation, correlation or regression analysis, comparisons of subclass means or proportions, and any other analysis involving two or more variables. As will be illustrated below, imputation can have harmful effects on all analyses of relationships, often attenuating the associations between variables. Discussions of the effects of imputations on relationships are provided by Santos (1981), Kalton and Kaspryzk (1982) and Little (1986a).

The general nature of the effect of imputation on relationships can be seen by considering its effect on the estimate of the sample covariance in the simple situation where the y-variable has missing responses that are missing at random over the population and the x-variable has no missing data. The sample covariance, s_{xy}, is calculated in the standard way, based on the actual values for respondents and the imputed values for nonrespondents, as an estimate of the population covariance S_{xy}. Using the fact that $E_2(\hat{y}_{mis}) = \hat{y}_{mid}$ as above, it can be readily shown that the expected value of s_{xy} under a deterministic imputation method is the same as that under the corresponding stochastic method.

As Santos (1981) shows, the relative bias of s_{xy} when the mean overall or random overall imputation methods are used is approximately $-\bar{M}$, where $\bar{M}$ is the nonresponse rate. This occurs because the imputed y-values are unrelated to their x-values, and hence the cases with imputed values attenuate the covariance towards zero. This attenuation is decreased in magnitude by imputation methods that use auxiliary variables. With class mean imputation or random imputation within classes, the relative bias is approximately $-\bar{M}(S_{xy.z}/S_{xy})$, where $S_{xy.z} = \Sigma W_h S_{xyh}$ is the average within-class covariance for classes formed by the auxiliary variables z, S_{xyh} is the covariance within class h, and W_h is the proportion of the population in class h. With predicted regression imputation or regression imputation with a random residual, both with a single auxiliary variable z, the relative bias is approximately $-\bar{M}[1 - (\rho_{xz}\rho_{yz}/\rho_{xy})]$, where ρ_{uv} is the correlation between u and v.

The disturbing feature of these results is that, unless $\bar{M}$ is small, s_{xy} calculated with imputed values under any of these imputation methods may be subject to substantial bias even under the missing at random model. The estimates s_{xy} computed with imputed values obtained under the imputation class and regression methods are unbiased only if the partial covariance $S_{xy.z}$ is zero. In general, there is no reason to assume uncritically that $S_{xy.z}$ is zero. However, there is an important case when $S_{xy.z} = 0$. This occurs when $x = z$, that is when x is used as an auxiliary variable in the imputation procedure. In this case, the sample covariance is unbiased under the missing at random model. This result suggests that if the relationship between x and y is to form an important part of the survey analysis, x should be used as an auxiliary variable in imputing for missing y-values.

The above theory assumes that only the y-variable was subject to missing data. In practice the x-variable will often also be incomplete. If so, the sample covariance may be attenuated because of the imputations for both variables. A special feature occurs when x and y are both missing for a record. If the two values are imputed separately, the covariance is attenuated, but if they are imputed jointly, using the same respondent as the donor of both values, the covariance structure is retained. This suggests that when a record has several missing related values, they should be taken from the same donor. Coder (1978) describes the use of joint imputation from the same donor in the March Income Supplement of the Current Population Survey.

As an illustration of how the above arguments about the attenuation of covariances apply to other forms of relationships, we will give a simple numerical example of the effect of imputation on the difference between two proportions. Let the variable of interest be whether an individual has a particular attribute or not, and suppose that one half of the respondents fail to answer this question. The missing responses are imputed by a random within class imputation method using two classes, A and B. The objective is now to compare the

Table 1
Number of Respondents with the Attribute, and Number of Sampled Persons by Class, Sex and Response Status

	Class A			Class B		
	M	F	Total	M	F	Total
Respondents with the attribute	80	40	120	60	20	80
Total respondents	100	100	200	100	100	200
Nonrespondents	100	100	200	100	100	200
Total sample	200	200	400	200	200	400

percentages of men and women with the attribute. The data are displayed in Table 1. Since 60% of the total respondents in class A have the attribute, 60 of the 100 male and 60 of the 100 female nonrespondents in that class will be imputed to have the attribute. Similarly, in class B 40% of the total respondents have the attribute, and so 40 male and 40 female nonrespondents will be imputed to have the attribute. The proportion of actual and imputed males with the attribute is thus $(80 + 60 + 60 + 40)/400 = 0.6$ or 60%. For females the corresponding proportion is $(40 + 60 + 20 + 40)/400 = 0.4$, or 40%. The difference between these two percentages is 20%.

Had sex also been taken into account in forming the imputation classes, the percentages of males and females with the attribute would have been 70% and 30%, differing by 40%. The failure to include sex as an auxiliary variable in the imputation has thus caused a substantial attenuation in the measurement of the relationship between sex and having the attribute.

3.4 Multiple Imputations

Ideally the analyst using a data set with imputed values should be able to obtain valid results for any analyses by applying standard techniques for complete data. However, as noted in the last section, imputation can distort measures of the relationships between variables. It also distorts standard error estimation.

All imputation methods except deductive imputation fabricate data to some extent. The extent of fabrication depends on how well the imputation model predicts the missing values. If the imputation model explains only a small proportion of the variance in the variable among the respondents, the amount of fabrication in each imputed value is likely to be substantial. If the imputation model explains a high proportion of the respondent variance, the amount of fabrication is likely to be less serious. However, it needs to be recognized that the fit of the imputation model for the respondents is not necessarily a good measure of the fit for the nonrespondents.

Standard errors computed in the standard way from a data set with imputed values will generally be underestimates because of the fabrication involved in the imputed values. Rubin (1978, 1979) has advocated the method of multiple imputations to provide valid inferences from data sets with imputed values (see also Herzog and Rubin 1983; Rubin and Schenker 1986). When multiple imputations are used for the purpose of standard error estimation, the construction of the complete data set by imputing for the missing responses is carried out several (say m) times using the same imputation procedure. The sample estimates z_i $(i = 1, 2, \ldots, m)$ of the population parameter of interest Z are computed from each of the replicate data sets, and their average $\bar{z}$ is calculated. A variance estimator for $\bar{z}$ is then given by $\hat{V} = \hat{W} + [(m + 1)/m]\hat{B}$, where $\hat{W}$ is the average of the within-replicate variance of $\bar{z}$ and $\hat{B} = \Sigma(z_i - \bar{z})^2/(m - 1)$ is the between-replicate variance. Even with the inclusion of the between-replicate variance component, however, the coverages of confidence intervals for Z based on $\hat{V}$ are still overstated, with the amount of overstatement increasing with the level of nonresponse.

This overstatement of the confidence levels can be addressed by modifying the imputation procedure, as described by Rubin and Schenker (1986). Their treatment considers the random overall imputation method, and one of their modifications allows for uncertainty about the population mean and variance in the following way. With the standard random overall imputation method, the conditional expected mean and variance of the imputed values are the sample respondents' mean and variance. With the modification, the expected mean and variance of the imputed values for a replicate are drawn at random from appropriate distributions. The imputed values are then a random selection of respondents' values, modified for the randomly-chosen mean and variance. When estimating the population mean, the effect of the changing expected mean and variance between replicates is to increase the between-replicate variance component in $\hat{V}$. This increase gives improved coverage for the resultant confidence intervals.

A major problem with the use of multiple imputations is the additional computer analysis needed, which increases as the number of replicates, m, increases. For this reason, a small value of m, such as $m = 2$, may be preferred. A small value of m may, however, result in a low level of precision for the variance estimator. Even with small m, it is questionable whether the multiple imputation approach is feasible for routine analyses. It may be best reserved for special studies, such as that described by Herzog and Rubin (1983).

In addition to providing appropriate standard errors, another advantage of multiple imputations from the same imputation procedure is that it reduces the loss of precision in survey estimates arising from the random selection of respondents to act as donors of imputed values (see Section 3.1). This loss is reduced with multiple imputations by averaging over the replicates. A small number of replicates serves well for this purpose. As noted earlier, Kalton and Kish (1984) describe alternative ways of selecting the sample of respondents to achieve this end.

A second major potential application of multiple imputations is to generate the imputations for the several replicates by different imputation procedures, making different assumptions about the nonrespondents. Suppose, for instance, that hourly rates of pay are to be imputed for some earners in the sample. One procedure that might be used is the random within class imputation method, which is based on an assumption that nonrespondents are missing at random within the classes. If it is thought that the nonrespondents might in fact come more heavily from those with higher rates of pay in each class, a simple modification to the random within class method might be to impute values that are, say, 50 cents above the donors' values. Other imputation procedures - for instance, using different imputation classes - could also be tried. Comparison of the survey estimates obtained from the data sets in which the different imputation procedures are applied then provides a valuable indication of the sensitivity of the estimates to the values imputed. If the estimates turn out to be very similar, they can be accepted with greater confidence; if they differ markedly, the estimates need to be treated with considerable caution.

4. CONCLUDING REMARKS

Weighting and imputation have been presented as two distinct methods for handling missing survey data, but in fact there is a close relationship between them. This may be illustrated by considering any imputation method that assigns respondents' values to the nonrespondents. For univariate analyses, this process is equivalent to dropping the nonrespondents' records and adding the nonrespondents' weights to those of the donor respondents (Kalton 1986).

The differences between weighting and imputation emerge when one considers the multivariate nature of survey data. It is possible to impute for the responses of a total nonrespondent by taking all the responses from a single donor; however, weighting is generally simpler in this case and it avoids the loss of precision arising from the sampling of respondents to serve as donors. It is not practicable to use weighting to handle item nonresponse since it would result in different sets of weights for each item; this would cause serious difficulties for crosstabulations and other analyses of the relationships between variables.

Weighting is a single global adjustment that attempts to compensate for the missing responses to all the items simultaneously. Imputation, on the other hand, is item-specific. This difference has consequences for the way that the auxiliary data are used. In forming weighting classes, the focus is on determining classes that differ in their response rates. The choice of auxiliary variables to use in imputation, however, is primarily made in terms of their abilities to predict the missing responses.

An assumption underlying all the procedures reviewed in this paper is that once the auxiliary variables have been taken into account the missing values are missing at random. Thus, for instance, the nonrespondents are assumed to be like the respondents within weighting and imputation classes. This assumption can be avoided by using stochastic censoring models, as has been done by Greenlees *et al.* (1982) in imputing wages and salaries in the Current Population Survey. However, as Little (1986b) observes, these models are highly sensitive to the distributional assumptions made.

An alternative approach for handling missing survey data is to leave the values missing in the data set and let the analyst incorporate appropriate missing data models into the analysis (Little 1982). This approach has much to commend it, but the labor and computing time needed to implement it effectively preclude its use as a general purpose strategy. Rather, the approach seems best suited for a small range of special analyses. In order to permit the analyst to adopt this approach, it is essential that all imputed values be flagged to indicate they are not actual responses, so that they can then be dropped from the analysis.

Finally, we should note that all methods of handling missing survey data must depend upon untestable assumptions. If the assumptions are seriously in error, the analyses may give misleading conclusions. The only secure safeguard against serious nonresponse bias in survey estimates is to keep the amount of missing data small.

REFERENCES

BAILAR III, J.C., and BAILAR, B.A. (1978). Comparison of two procedures for imputing missing survey values. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 462-467.

BAILAR, B.A., and BAILAR III, J.C. (1983). Comparison of the biases of the hot-deck imputation procedure with an "equal-weights" imputation procedure. In *Incomplete Data in Sample Surveys, Volume 3, Proceedings of the Symposium*, (Eds. W.G. Madow and I. Olkin), New York: Academic Press, 299-311.

BAILAR, B.A., BAILEY, L., and CORBY, C.A. (1978). A comparison of some adjustment and weighting procedures for survey data. In *Survey Sampling and Measurement*, (Ed. N.K. Namboodiri), New York: Academic Press, 175-198.

BARTHOLOMEW, D.J. (1961). A method of allowing for 'not at home' bias in sample surveys. *Applied Statistics*, 10, 52-59.

BISHOP, Y.M.M., FIENBERG, S.E., and HOLLAND, P.W. (1975). *Discrete Multivariate Analyses*. Cambridge, Mass: The MIT Press.

BROOKS, C.A., and BAILAR, B.A. (1978). *An Error Profile: Employment as Measured by the Current Population Survey*. Statistical Policy Working Paper 3. U.S. Department of Commerce. Washington, D.C.: U.S. Government Printing Office.

CHAPMAN, D.W., BAILEY, L., and KASPRZYK, D. (1986). Nonresponse adjustment procedures at the U.S. Census Bureau. *Survey Methodology*, forthcoming.

CODER, J. (1978). Income data collection and processing from the March Income Supplement to the Current Population Survey. *The Survey of Income and Program Participation Proceedings of the Workshop on Data Processing*, February 23-24, 1978, (Ed. D. Kasprzyk), Chapter II. Washington, D.C.: U.S. Department of Health, Education and Welfare.

COLLEDGE, M.J., JOHNSON, J.H., PARE, R., and SANDE, I.G. (1978). Large scale imputation of survey data. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 431-436.

COX, B.G., and COHEN, S.B. (1985). *Methodological Issues for Health Care Surveys*. New York: Marcel Dekker.

DAVID, M., LITTLE, R.J.A., SAMUHEL, M.E., and TRIEST, R.K. (1986). Alternative methods for CPS income imputation. *Journal of the American Statistical Association*, 81, 29-41.

DREW, J.H., and FULLER, W.A. (1980). Modelling nonresponse in surveys with callbacks. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 639-642.

DREW, J.H., and FULLER, W.A. (1981). Nonresponse in complex multiphase surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 623-628.

FORD, B.L. (1983). An overview of hot-deck procedures. In *Incomplete data in Sample Surveys, Volume 2, Theory and Bibliographies*, (Eds. W.G. Madow, I. Olkin and D.B. Rubin), New York: Academic Press, 185-207.

GREENLEES, W.S., REECE, J.S., and ZIESCHANG, K.D. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association*, 77, 251-261.

HERZOG, T.N., and RUBIN, D.B. (1983). Using multiple imputation to handle nonresponse in sample surveys. In *Incomplete data in Sample Surveys, Volume 2, Theory and Bibliographies*, (Eds. W.G. Madow, I. Olkin and D.B. Rubin), New York: Academic Press, 209-245.

KALTON, G. (1983). *Compensating for Missing Survey Data*. Ann Arbor: Survey Research Center, University of Michigan.

KALTON, G. (1986). Handling wave nonresponse in panel surveys. *Journal of Official Statistics*, 2, forthcoming.

KALTON, G., and KASPRZYK, D. (1982). Imputing for missing survey responses. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 22-31.

KALTON, G., and KISH, L. (1984). Some efficient random imputation methods. *Communications in Statistics - Theory and Methods*, 13(16), 1919-1939.

KISH, L. (1965). *Survey Sampling*. New York: Wiley.

KISH, L. (1976). Optima and proxima in linear sample designs. *Journal of the Royal Statistical Society*, Ser. A, 139, 80-95.

LITTLE, R.J.A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, 77, 237-250.

LITTLE, R.J.A. (1986a). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 139-157.

LITTLE, R.J.A. (1986b). Missing data in Census Bureau surveys. *Proceedings of the Second Annual Census Bureau Research Conference*, 442-454.

LITTLE, R.J.A., and DAVID, M.H. (1983). Weighting adjustments for non-response in panel surveys. Working Paper, Washington, D.C.: U.S. Bureau of the Census.

OH, H.L., and SCHEUREN, F. (1978a). Multivariate raking ratio estimation in the 1973 Exact Match Study. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 716-722.

OH, H.L., and SCHEUREN, F. (1978b). Some unresolved application issues in raking ratio estimation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 723-728.

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

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

OH, H.L., SCHEUREN, F., and NISSELSON, H. (1980). Differential bias impacts of alternative Census Bureau hot deck procedures for imputing missing CPS income data. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 416-420.

PALMER, S. (1967). On the character and influence of nonresponse in the Current Population Survey. *Proceedings of the Social Statistics Section, American Statistical Association*, 73-80.

PALMER, S., and JONES, C. (1966). A look at alternate imputation procedures for CPS noninterviews. Washington, D.C.: U.S. Bureau of the Census memorandum.

POLITZ, A., and SIMMONS, W. (1949). I. An attempt to get the 'not at homes' into the sample without callbacks. II. Further theoretical considerations regarding the plan for eliminating callbacks. *Journal of the American Statistical Association*, 44, 9-31.

POLITZ, A., and SIMMONS, W. (1950). Note on an attempt to get the 'not at homes' into the sample without callbacks. *Journal of the American Statistical Association*, 45, 136-137.

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

RUBIN, D.B. (1979). Illustrating the use of multiple imputations to handle nonresponse in sample surveys. *Bulletin of the International Statistical Institute*, 48(2), 517-532.

RUBIN, D.B., and SCHENKER, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, 81, 366-374.

SANDE, G. (1979). Numerical edit and imputation. Paper presented to the International Association for Statistical Computing, 42nd Session of the International Statistical Institute.

SANDE, I.G. (1983). Hot-deck imputation procedures. In *Incomplete Data in Sample Surveys, Volume 3, Proceedings of the Symposium*, (Eds. W.G. Madow and I. Olkin), New York: Academic Press, 339-349.

SANTOS, R.L. (1981). Effects of imputation on regression coefficients. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 140-145.

THOMSEN, I. (1973). A note on the efficiency of weighting subclass means to reduce the effects of nonresponse when analyzing survey data. *Statistisk Tidskrift*, 4, 278-283.

THOMSEN, I., and SIRING, E. (1983). On the causes and effects of nonresponse: Norwegian experiences. In *Incomplete Data in Sample Surveys, Volume 3, Proceedings of the Symposium*, (Eds. W.G. Madow and I. Olkin), New York: Academic Press, 25-29.

VACEK, P.M., and ASHIKAGA, T. (1980). An examination of the nearest neighbor rule for imputing missing values. *Proceedings of the Statistical Computing Section, American Statistical Association*, 326-331.

WELNIAK, E.J., and CODER, J.F. (1980). A measure of the bias in the March CPS earnings imputation system. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 421-425.