Journal of Official Statistics
Vol. 4. No. 3. 1988, pp. 251–260

Reduction of Nonresponse Bias Through Regression Estimation

Jelke G. Bethlehem[1]

Abstract: To investigate the properties of estimators of population characteristics, nonresponse is incorporated in the sampling theory by the introduction of response probabilities. Within this framework the characteristics of the Horvitz-Thompson estimator, generalized regression estimator and post-stratification estimator are studied. It is shown that proper use of auxiliary information can reduce the nonresponse bias.

Key words: Nonresponse; regression estimator; post-stratification.

1. Introduction

The results of a sample survey are affected by many kinds of errors, one of the most important sources being nonresponse. The main problem caused by nonresponse is that estimators of population characteristics must be assumed to be biased unless convincing evidence to the contrary is provided.

Hansen and Hurwitz addressed the problem of nonresponse in mail surveys in 1946. Nonresponse remains a subject for ongoing concern. An important symposium on missing data (see Madow and Olkin (1983)) concluded that even with intensive efforts in the data collection stage, nonresponse will inevitably occur. Furthermore, no statistical method will fully compensate for missing data, and biases will almost certainly remain.

Generally, two different approaches can be distinguished in tackling the nonresponse. The first approach is to make new attempts to collect the missing data, and the second approach is to make do with the data, but to apply an adjustment technique. An example of the first approach is the proposal by Hansen and Hurwitz (1946) to let a subsample of mail survey nonrespondents be visited by interviewers. This type of reinterview may also work for face to face interviews, for instance if a subsample of nonrespondents is visited a second time, but now by especially trained interviewers. Another implementation of the first approach is the Basic Question Procedure, proposed by Bethlehem and Kersten (1985), in which refusers are asked to answer only one or two important questions.

An example of the second approach is adjustment of the available data using a weighting or imputation procedure. Imputation is often carried out in the case of item nonresponse. Imputation techniques are discussed by, e.g., Platek and Gray (1983).

[1] Netherlands Central Bureau of Statistics, Voorburg, The Netherlands.

In the case of unit nonresponse weighting is generally preferred, see, e.g., Bailar, Bailey, and Corby (1978), Lindström, Wretman, Forsman, and Cassel (1979), and also Platek and Gray (1983).

A vital part of the adjustment approach is the availability of auxiliary information. Auxiliary information and the sampling data are used to construct models that describe nonrespondent behaviour with respect to the target variables of the survey. In this paper, modified versions of the Horvitz-Thompson estimator and the generalized regression estimator are proposed. It will become clear that use of auxiliary variables can reduce the bias. To study the effect of nonresponse on estimators, a general framework is proposed. In this framework, nonresponse is incorporated in the sampling theory by introducing the concept of individual response probability. The assumption that each sample element has a certain (unknown) response probability is, however, also a vital part of many other theoretical contributions on nonresponse.

In the literature, adjustment is applied in various ways. Platek and Gray (1983) use a framework similar to ours in the context of missing value imputation. Little (1982) presents a model-based theory, in which values in the population are treated as realizations of random variables that are distributed according to a superpopulation model. The theoretical concept of ignorability is used to obtain insight into the role of probability sampling and the mechanism causing nonresponse. Greenlees, Reece, and Zieschang (1982) develop a method for imputing missing values when the response probability depends on the variable that has been imputed by a logistic function. Under an assumed linear regression superpopulation model, the theory of stochastic censoring is applied. Särndal and Swensson (1985) point at the resemblance between two phase sampling and the nonresponse situation. They develop a general theory for two phase sampling (in the case of full response), and apply the results to nonresponse, where nonresponse is conceived as the second phase with unknown, but estimated, sampling probabilities.

In Section 2 the theoretical framework is introduced. Section 3 discusses the Horvitz-Thompson estimator and Section 4 discusses the generalized regression estimator. It can be shown that post-stratification is a special case of the generalized regression estimator. This case is treated in Section 5.

2. The Theoretical Framework

Let the target population of a sample survey consist of N identifiable elements, which are labeled 1,2,..., N. Associated with each element k is a value Y_k of the target variable. The N-vector of all values of the target variable in the population is denoted by Y. The aim of the sample survey is the estimation of the population mean

$$\bar{Y} = \frac{1}{N} \sum_{k=1}^{N} Y_k$$

of the target variable. Of course, inference on population totals can easily be derived from the results on population means.

A sample selected without replacement from the population can be represented by an N-vector $t=(t_1,t_2,\ldots,t_N)'$ of indicators. The k-th indicator t_k assumes the value 1 if element k is selected, and otherwise it assumes the value 0. The expected value of t is equal to

$$E(t) = \pi,$$

where $\pi=(\pi_1, \pi_2,\ldots,\pi_N)'$ is the N-vector of first order inclusion probabilities of the elements. The second order inclusion probability of elements k and l $(k\neq l)$ can be written as

$$\pi_{kl} = E(t_k t_l) \text{ and } \pi_{kk} = \pi_k.$$

There are two ways to incorporate nonresponse in the theory, (1) the fixed response approach and (2) the random response approach. Both are discussed by, e.g., Kalsbeek (1980) and Cassel, Särndal, and Wretman (1983).

In the fixed response approach the population is divided into two strata, one consists of (potential) respondents and the other of (potential) nonrespondents. A sample is selected from the population, thereby disregarding the stratification. Sampled elements which belong to the nonresponse stratum will not respond.

In this paper the more general random response approach is adopted. Each element k in the population is assumed to have an (unknown) response probability ϱ_k. Only responding elements can be observed and the response is represented by an N-vector.

$$r = (r_1, r_2, \ldots, r_N)'$$

of indicators. The kth indicator r_k assumes the value 1 if element k is selected ($t_k=1$) and responds; otherwise it assumes the value 0. The expected value of r_k is equal to

$$E(r_k) = \pi_k \varrho_k,$$

for $k=1,2,\ldots,N$, where the expectation is taken over all possible samples and all possible response patterns. In this theory, the fundamental assumption is that the response behaviour of one element is not influenced by the response behaviour of another element. Consequently, for two elements k and l ($k \neq l$)

$E(r_k r_l)$ is equal to $\pi_{kl}\, \varrho_k\, \varrho_l$.

This result is used in the computation of variances of estimators. It will be assumed that, for respondents, the values of the target variable are observed without measurement error, i.e., there is no response error.

3. The Modified Horvitz-Thompson Estimator

An unbiased estimator of the population mean $\bar{Y}$, given sampling with unequal probabilities without replacement and under full response, is given by Horvitz and Thompson (1952). Using the notation of the previous section this estimator can be written as

$$\bar{y}_{HT} = \frac{1}{N} \sum_{k=1}^{N} \frac{Y_k t_k}{\pi_k} \quad . \tag{3.1}$$

The variance of this estimator is equal to

$$V(\bar{y}_{HT}) = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{l=1}^{N} (\pi_{kl} - \pi_k \pi_l) \frac{Y_k Y_l}{\pi_k \pi_l} \quad .$$

In the case of nonresponse, estimates must be based on responding elements rather than sampled elements. The obvious modification for the Horvitz-Thompson estimator in this case would be to replace t_k by r_k and π_k by $\pi_k \varrho_k$ in (3.1). In practice, however, this will not work, since the response probabilities $\varrho_1, \varrho_2, \ldots, \varrho_N$ are unknown. Furthermore, if the r_k and π_k are used and if the nonresponse is ignored, the estimator will be biased. For example, if all Y_k are positive, the estimator will always be biased downwards. There are two ways to solve this problem, but both approaches are based on the use of auxiliary information. In the first approach the estimate is improved by using a generalized regression estimator instead of the Horvitz-Thompson estimator. Examples of this approach can be found in Greenlees et al. (1982) and Bethlehem and Keller (1987). In the second approach the available auxiliary information is used to make estimates of the response probabilities. These estimates will then be substituted in the Horvitz-Thompson estimator. This approach is suggested in Särndal (1981).

Both approaches may reduce the bias of the resulting estimator. The success of the regression approach relies on the extent to which the auxiliary information can describe the behaviour of the values of the target variable. In the second approach the auxiliary variables must be able to describe the behaviour of the values of the response probabilities. Here, we concentrate on the regression approach.

In both approaches something must be done about the unknown response probabilities. A reasonable, but probably not ideal, solution is to replace each ϱ_k by

$$\bar{r}_{\mathrm{HT}} = \frac{1}{N} \sum_{k=1}^{N} \frac{r_k}{\pi_k} ,$$

which is an unbiased estimator of the mean response probability

$$\bar{\varrho} = \frac{1}{N} \sum_{k=1}^{N} \varrho_k .$$

Thus, in the case of nonresponse the modified Horvitz-Thompson estimator is defined as

$$\bar{y}^*_{\mathrm{HT}} = \frac{1}{N} \sum_{k=1}^{N} \frac{Y_k r_k}{\pi_k \bar{r}_{\mathrm{HT}}} =$$

$$\left(\sum_{k=1}^{N} \frac{Y_k r_k}{\pi_k} \right) / \left(\sum_{k=1}^{N} \frac{r_k}{\pi_k} \right) . \qquad (3.2)$$

Note that in the case of full response, estimator (3.2) is not reduced to the original Horvitz-Thompson estimator. Instead the population size N is replaced by its estimator based on the sample. This estimator is also discussed by Särndal (1980), who claims that this estimator has properties which make it preferable to the unbiased Horvitz-Thompson estimator.

The modified Horvitz-Thompson estimator (3.2) has the form of a ratio estimator. Therefore, its properties can be derived using the standard first order Taylor series expansion. The Taylor expansion technique is discussed in, e.g., Wolter (1985). If this technique is applied, it turns out that the expected value of estimator (3.2) can be approximated by

$$E(\bar{y}^*_{\mathrm{HT}}) \doteq \bar{Y}^* , \qquad (3.3)$$

where

$$\bar{Y}^* = \frac{1}{N} \sum_{k=1}^{N} \frac{\varrho_k Y_k}{\bar{\varrho}} . \qquad (3.4)$$

An approximation of the bias of the modified Horvitz-Thompson estimator can now be obtained by comparing (3.4) with the population mean $\bar{Y}$. This results in

$$B(\bar{y}^*_{\mathrm{HT}}) = E(\bar{y}^*_{\mathrm{HT}}) - \bar{Y} \doteq \bar{Y}^* - \bar{Y} = C_{\varrho Y} / \bar{\varrho}, \qquad (3.5)$$

in which

$$C_{\varrho Y} = \frac{1}{N} \sum_{k=1}^{N} (\varrho_k - \bar{\varrho}) (Y_k - \bar{Y})$$

is the population covariance between response probabilities and the values of the target variable. Hence, the modified Horvitz-Thompson estimator is unbiased if there is no correlation between the target variable and the response behaviour. The stronger the relationship between the target variable and the response behaviour, the larger the bias.

The variance of the modified Horvitz-Thompson estimator can be approximated by

$$V(\bar{y}^*_{\mathrm{HT}}) \doteq \frac{1}{(N\bar{\varrho})^2} \times \sum_{k=1}^{N} \sum_{l=1}^{N} (\varrho_{kl}\pi_{kl} - \varrho_k \varrho_l \pi_k \pi_l) \frac{(Y_k - \bar{Y}^*)(Y_l - \bar{Y}^*)}{\pi_k \pi_l} , \qquad (3.6)$$

where $\varrho_{kl}=\varrho_k\varrho_l$ for $k\neq l$, and $\varrho_{kk}=\varrho_k$.

In the case of full response, the variance of the Horvitz-Thompson estimator vanishes, provided the first order inclusion probabilities are proportional to the values of the target variable (if the sample size is fixed). Thus, in practice one tries to establish inclusion probabilities using an auxiliary variable which is as proportional as possible to the target variable. This practice does not work in the case of nonresponse. The extra quantity $\bar{Y}^*$ in (3.6) disturbs the nice variance property of the Horvitz-Thompson estimator.

4. The Generalized Regression Estimator

In the full response case, the precision of the Horvitz-Thompson estimator can be improved if suitable auxiliary information is available. Suppose there are p auxiliary variables. Each element k in the population is associated with a p-vector $X_k=(X_{k1},X_{k2},\ldots,X_{kp})'$ of variable values. The population mean of the vectors is denoted by

$$\bar{X}=\frac{1}{N}\sum_{k=1}^{N}X_k \ .$$

The $N\times p$ matrix of all values of the auxiliary variables is denoted by X. If the auxiliary variables are correlated with the target variable, then for a suitably chosen vector $\beta=(\beta_1,\beta_2,\ldots,\beta_p)'$ of regression coefficients for a best fit of Y on X, the residuals in the vector $\varepsilon=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_N)'$, defined by

$$\varepsilon=Y-X\beta \quad ,$$

vary less than the values of the target variable itself. Application of ordinary least squares results in

$$\beta=(X'X)^{-1}X'Y=\Big(\sum_{k=1}^{N}X_kX_k'\Big)^{-1}\Big(\sum_{k=1}^{N}X_kY_k\Big).$$

In the case of full response this quantity can be estimated by

$$\hat{\beta}=\left(\sum_{k=1}^{N}\frac{t_kX_kX_k'}{\pi_k}\right)^{-1}\left(\sum_{k=1}^{N}\frac{t_kX_kY_k}{\pi_k}\right). \tag{4.1}$$

The estimator $\hat{\beta}$ is an asymptotically design unbiased (ADU) estimator of β. This means that $\hat{\beta}$ is asymptotically unbiased for large samples. Using (4.1) the *generalized regression estimator* for the case of full response is defined as

$$\bar{y}_{GR}=\bar{y}_{HT}+(\bar{X}-\bar{x}_{HT})'\,\hat{\beta} \quad , \tag{4.2}$$

in which $\bar{x}_{HT}$ is the analogue of $\bar{y}_{HT}$, as defined in (3.1). The generalized regression estimator is an ADU estimator of the population mean $\bar{Y}$. The estimator and its properties are also discussed in Robinson and Särndal (1983), Isaki and Fuller (1982), and Bethlehem (1985). Bethlehem and Keller (1987) investigate generalized regression estimation in the context of weighting sample survey data. If there exists a p-vector c of fixed numbers such that $Xc=\iota$, where ι is a vector consisting of 1's, estimator (4.2) can also be written as

$$\bar{y}_{GR}=\bar{X}'\hat{\beta}. \tag{4.3}$$

This condition is fulfilled if the regression model contains a constant term, or if post-stratification is used (see Section 5). In the remainder of this paper it is assumed that $Xc=\iota$. Another consequence of this assumption is that $\bar{Y}-X\beta=\bar{\varepsilon}=0$.

Given simple random sampling and only one continuous auxiliary variable, the generalized regression estimator reduces to the well-known simple regression estimator as, e.g., discussed by Cochran (1977). If can be

shown that the variance of estimator (4.5) can be approximated by

$$V(\bar{y}_{GR}) \doteq \frac{1}{N^2} \sum_{k=1}^{N} \sum_{l=1}^{N} (\pi_{kl} - \pi_k \pi_l) \frac{\varepsilon_k \varepsilon_l}{\pi_k \pi_l} \quad ,$$

where $\pi_{kk} = \pi_k$, (see e.g. Särndal (1982) or Bethlehem and Keller (1987)). This variance will be small if the residual values $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ are small. Hence, the use of auxiliary variables which can explain the behaviour of the target variable will result in a precise estimator.

In the case of nonresponse, the Horvitz-Thompson estimators $\bar{y}_{HT}$ and $\bar{x}_{HT}$ cannot be used. Furthermore, also estimation of β will have to be based on available observations only. The modified generalized regression estimator is now defined by

$$\bar{y}^*_{GR} = \bar{y}^*_{HT} + (\bar{X} - \bar{x}^*_{HT})' \hat{\beta}^*, \qquad (4.4)$$

in which $\bar{y}^*_{HT}$ is defined by (3.2), $\bar{x}^*_{HT}$ is the analogue of $\bar{y}^*_{HT}$, and $\hat{\beta}^*$ is equal to

$$\hat{\beta}^* = \left(\sum_{k=1}^{N} \frac{r_k X_k X_k'}{\pi_k} \right)^{-1} \left(\sum_{k=1}^{N} \frac{r_k X_k Y_k}{\pi_k} \right) .$$

Using the same techniques as Cochran (1977, p.193) in his proof of the consistency of the simple regression estimator, it can be shown that $\hat{\beta}^*$ is an ADU estimator of

$$\beta^* = \left(\sum_{k=1}^{N} \varrho_k X_k X_k' \right)^{-1} \left(\sum_{k=1}^{N} \varrho_k X_k Y_k \right) .$$

The expected value of $\bar{x}^*_{HT}$ can be approximated by

$$E(\bar{x}^*_{HT}) \doteq \bar{X}^* = \frac{1}{N} \sum_{k=1}^{N} \frac{\varrho_k X_k}{\bar{\varrho}} .$$

Therefore, the expected value of the modified regression estimator can be approximated by

$$E(\bar{y}^*_{GR}) \doteq \bar{Y}^* + (\bar{X} - \bar{X}^*)' \beta^*,$$

and the bias of this estimator is approximately equal to

$$\begin{aligned} B(\bar{y}^*_{GR}) &= E(\bar{y}^*_{GR}) - \bar{Y} \\ &= (\bar{Y}^* - \bar{Y}) - (\bar{X}^* - \bar{X})' \beta^* \end{aligned} \qquad (4.5)$$

Since $\bar{Y}^* = \bar{X}^* \beta^*$ (assuming $Xc = \iota$), the bias can also be written as

$$B(\bar{y}^*_{GR}) = (\bar{X} \beta^* - \bar{Y}). \qquad (4.6)$$

From (4.6) it is clear that the bias vanishes if β^* is equal to β. Thus, if nonresponse does not affect the regression coefficients, the resulting regression estimator will not be biased. By writing

$$\beta^* = \beta + \left(\frac{1}{N} \sum_{k=1}^{N} \frac{\varrho_k X_k X_k'}{\bar{\varrho}} \right)^{-1} \bar{\varepsilon}^*,$$

where

$$\bar{\varepsilon}^* = \frac{1}{N} \sum_{k=1}^{N} \frac{\varrho_k \varepsilon_k}{\bar{\varrho}} \quad ,$$

two conclusions can be drawn. First, β^* and β will be equal if $\bar{\varepsilon}^* = 0$, and that will be the case if there is no correlation between the residuals of the regression model and the response behaviour. Second, β^* and β will be approximately equal if $\bar{\varepsilon}^*$ is small. Hence, a good fit (small residuals) will reduce the bias.

5. A Special Case: Post-Stratification

Post-stratification is a well-known and frequently used technique to reduce nonresponse bias. For example, Thomsen (1973, 1978) presents formulae for the bias of estimators after post-stratification, which are based on the fixed response approach. A general approach to post-stratification based on linear models is given by Bethlehem and Keller (1987).

Post-stratification can also be treated from the random response viewpoint. In fact, post-stratification theory is a special case of the theory given in the previous section. Suppose we want to stratify the sample into L strata. To that end L dummy variables are introduced. Associated with each element k is a vector $(X_{k1}, X_{k2},\ldots,X_{kL})'$ of dummy values. The hth dummy X_{kh} assumes the value 1 if element k belongs to stratum h and 0 if it belongs to another stratum. In the case of full response, β turns out to be

$$\beta = (\bar{Y}_1,\bar{Y}_2,\ldots,\bar{Y}_L)',$$

in which $\bar{Y}_h$ is the mean value in stratum L. The estimator for β becomes

$$\hat{\beta} = (\bar{y}_{HT,1},\bar{y}_{HT,2},\cdots,\bar{y}_{HT,L})', \qquad (5.1)$$

where $\bar{y}_{HT,h}$, for $h=1,2,\ldots,L$, is the Horvitz-Thompson estimator in stratum h (the adapted version, in which the population size is replaced by its estimator).

The mean vector $\bar{X}$ of the dummy values in the population is equal to

$$\bar{X} = (W_1,W_2,\ldots,W_L)', \qquad (5.2)$$

in which $W_h=N_h/N$ is the relative size of stratum h. Substitution of (5.1) and (5.2) in (4.3) gives

$$\bar{y}_{GR} = \sum_{h=1}^{L} W_h\,\bar{y}_{HT,h} \ . \qquad (5.3)$$

For simple random sampling, this is the post-stratification estimator as discussed by, e.g., Cochran (1977). Therefore, estimator (5.3) will also be denoted by $\bar{y}_{PS}$. In the case of full response, the post-stratification estimator is approximately unbiased. The estimator is not exactly unbiased, because there is a non-zero, but generally small, probability of empty strata. If there is nonresponse, we have to rely on the modified generalized regression estimator which in the case of post-stratification turns into

$$\bar{y}^*_{GR} = \sum_{h=1}^{L} W_h\,\bar{y}^*_{HT,h} \ . \qquad (5.4)$$

In (5.4) $\bar{y}^*_{HT,h}$ is a modified Horvitz-Thompson estimator for stratum h. We denote the estimator (5.4) also by $\bar{y}^*_{PS}$ and its bias is

$$B(\bar{y}^*_{PS}) = \sum_{h=1}^{L} W_h(\bar{Y}^*_h - \bar{Y}_h) \ .$$

The quantity $\bar{Y}^*_h$ is the analogue of $\bar{Y}^*$, but applied to stratum h. Apparently the bias of the post-stratification estimator is a weighted sum of the stratum biases. By applying (3.5) the bias can be rewritten as

$$B(\bar{y}^*_{PS}) = \sum_{h=1}^{L} W_h\,C_{\varrho Y,h}/\bar{\varrho}_h \ , \qquad (5.5)$$

in which

$$C_{\varrho Y,h} = \frac{1}{N_h}\sum_{k=1}^{N_h} (\varrho_{kh}-\bar{\varrho}_h)\,(Y_{kh}-\bar{Y}_h),$$

and where the double subscript kh denotes the kth element in stratum h. The quantity $\bar{\varrho}_h$ is the mean of the response probabilities in stratum h. Within a stratum the bias will vanish if there is no relationship between the response probabilities and the values of the target variable. This gives us some guidance

on how to stratify: construct strata in which the response behaviour is independent of the target variable. The following is a discussion of rules that are helpful in practice.

1. Construct strata which are homogeneous with respect to the target variable. If the values of this variable differ very little, the covariance $C_{\varrho Y,h}$ will be close to zero.

2. Construct strata which are homogeneous with respect to the response probabilities. Then again, the covariance will be close to zero.

Rule 1 concentrates on the target variable. It is a well-known rule which is also applied in the case of full response. In that situation post-stratification will lead to a small variance. Therefore, it is very important to look for good stratification variables that will reduce both variance and bias. The choice of stratification variables cannot be made solely on the basis of the available observations. Over or underrepresentation of some groups can mislead us about the relationship between the target and the stratification variable. There has to be additional information about the homogeneity of the target variable.

Rule 2 concentrates on the response probabilities. If the strata can be further divided into substrata and the population sizes of the substrata are known, then the mean response probabilities can be estimated in each substratum. Comparing these estimates may give an indication of the homogeneity of strata with respect to the response probabilities. If all probabilities within a stratum are equal, then the population distribution and the observed distribution of the target variable will coincide.

These rules are discussed in the literature. Thomsen (1973) splits the nonresponse bias in two components: one which measures the difference between response probabilities over strata, and another which contains the biases within strata. In Madow, Nisselson and Olkin (1983), Chapter 1, Recommendation 17, one is advised to select strata so that differences between respondents and nonrespondents are relatively small. Lock Oh and Scheuren (1983) stress the importance of a uniform response mechanism within strata. Platek and Gray (1983) indicate that the correlation between response probabilities and the characteristic to be investigated should be minimal.

Preferably, the statistician should apply both rules simultaneously. The stratification should be done in such a way that strata are homogeneous with respect to the target variables (thus decreasing the variance and bias) and with respect to the response probabilities (thus decreasing bias). In practice it will not always be easy to obtain such a stratification. Still, it is important that the statistician is aware of the possible effects if he/she is in a position to choose one of several possible stratifications.

Another way to look at the bias in case of post-stratification is to compare the bias of $\bar{y}^*_{\mathrm{PS}}$ with the bias of the modified Horvitz-Thompson estimator $\bar{y}^*_{\mathrm{HT}}$. It turns out that

$$B(\bar{y}^*_{\mathrm{PS}}) = B(\bar{y}^*_{\mathrm{HT}}) + \sum_{h=1}^{L} W_h \bar{Y}^*_h \left(1 - \frac{\bar{\varrho}_h}{\bar{\varrho}}\right). \tag{5.6}$$

It can be observed that a change in the bias is mainly caused by differences in mean response probabilities between strata and differences between the values of $\bar{Y}^*_h$. It is also clear from (5.6) that post-stratification does not necessarily reduce the bias; a badly chosen stratification may increase the bias.

Since all quantities can be estimated using sample information, we can estimate the bias shift for a given stratification. A large shift in bias is no guarantee that the bias is reduced. Nevertheless, usually a reduction does occur,

especially if the model describing the relationship between the target variable and the auxiliary variables fits well. Thus, the residuals in ε must be small. In the case of post-stratification, the residual ε_{kh}, associated with element k in stratum h, is equal to

$$\varepsilon_{kh} = Y_{kh} - X'_k\beta = Y_{kh} - \bar{Y}_h.$$

Again, we must conclude that strata must be homogeneous with respect to the target variable.

One should be aware of the fact that a reduction of the bias does not necessarily imply a reduction of the corresponding mean square error. A stratification may reduce the bias but increase the variance, especially if the number of observations per stratum is small. Still, one may prefer a bias reduction, thus permitting a more valid inference about population parameters.

6. Conclusion

This paper presents a general framework to study the behaviour of estimators given nonresponse by introducing response probabilities. It is shown that the bias can be reduced if models are built that can explain the behaviour of the target variable. The better the model fits, the smaller the nonresponse bias becomes. In the special case of stratification it becomes clear that good stratifications (in the traditional sense, i.e., strata that are homogeneous with respect to the target variable) also perform well in reducing the bias. Stratifications that are homogeneous with respect to the response probabilities work well too.

7. References

Bailar, B.A., Bailey, L., and Corby, C. (1978): A Comparison of Some Adjustment and Weighting Procedures for Survey Data. In: Namboodiri, N.K. (ed): Survey Sampling and Measurement, Academic Press, Cambridge.

Bethlehem, J.G. (1985): The Non-response Bias of Some Estimators. Internal CBS-report, Netherlands Central Bureau of Statistics, Voorburg.

Bethlehem, J.G. and Keller, W.J. (1987): Linear Weighting of Sample Survey Data. Journal of Official Statistics, 3, pp. 141–154.

Bethlehem, J.G. and Kersten, H.M.P. (1985): On the Treatment of Nonresponse in Sample Surveys. Journal of Official Statistics, 1, pp. 287–300.

Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1983): Some Uses of Statistical Models in Connection with the Nonresponse Problem. In: Madow, W.G. and Olkin, I., (eds.) (1983): Incomplete Data in Sample Surveys. Vol. 3: Proceedings of the Symposium, Academic Press, New York.

Cochran, W.G. (1977): Sampling Techniques. Third Edition, Wiley, New York.

Greenlees, J.S., Reece, W.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, pp. 251–261.

Hansen, M.H. and Hurwitz, W.H. (1946): The Problem of Nonresponse in Sample Surveys. Journal of the American Statistical Association, 41, pp. 517–529.

Horvitz, D.G. and Thompson, D.J. (1952): A Generalization of Sampling Without Replacement from a Finite Universe. Journal of the American Statistical Association, 47, pp. 663–685.

Isaki, C.T. and Fuller, W.A. (1982): Survey Design Under the Regression Superpopulation Model. Journal of the American Statistical Association, 77, pp. 89–96.

Kalsbeek, W.D. (1980): A Conceptual Review of Survey Error Due to Non-response. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 131–136.

Lindström, H., Wretman, J., Forsman, G., and Cassel, C. (1979): Standard Methods for Non-response Treatment in Statistical Estimation. Statistics Sweden, Stockholm.

Little, R.J.A. (1982): Models for Nonresponse in Sample Surveys. Journal of the American Statistical Association, 77, pp. 237–250.

Lock Oh, H. and Scheuren, F.J. (1983): Weighting Adjustment for Unit Nonresponse. In: Madow, W.G., Olkin, I., and Rubin, D.B. (eds.): Incomplete Data in Sample Surveys. Vol. 2: Theory and Bibliography, Academic Press, New York, pp. 249–333.

Madow, W.G., Nisselson, H., and Olkin, I. (eds.): Incomplete Data in Sample Surveys. Vol. 1: Report and Case Studies. Academic Press, New York.

Madow, W.G. and Olkin, I. (eds.) (1983): Incomplete Data in Sample Surveys. Vol. 3: Proceedings of the Symposium, Academic Press, New York.

Platek, R. and Gray, G.B. (1983): Imputation Methodology. In: Madow, W.G., Olkin, I., and Rubin, D.B. (eds.): Incomplete Data in Sample Surveys. Vol. 2: Theory and Bibliography, Academic Press, New York, pp 249–333.

Robinson, P.M. and Särndal, C.E. (1983): Asymptotic Properties of the Generalized Regression Estimator in Probability Sampling. Sankhyā, B, 45, pp. 240–248.

Särndal, C.E. (1980): On Pi-inverse Weighting Versus Best Linear Unbiased Weighting in Probability Sampling. Biometrika, 67, pp. 639–650.

Särndal, C.E. (1981): Frameworks for Inference in Survey Sampling with Application to Small Area Estimation and Adjustment for Non-response. Bulletin of the International Statistical Institute, 49, pp. 494–513.

Särndal, C.E. (1982): Implications for the Survey Design for Generalized Regression Estimation of Linear Functions. Journal of Statistical Planning and Inference, 7, pp. 155–170.

Särndal, C.E. and Swensson, B. (1985): Incorporating Nonresponse Modelling in General Randomization Theory. Bulletin of the International Statistical Institute, 51:3, pp. 15.2-1–15.2-15.

Thomsen, I. (1973): A Note on the Efficiency of Weighting Subclass Means to Reduce the Effects of Non-response When Analyzing Survey Data. Statistisk tidskrift, 11, pp. 278–283.

Thomsen, I. (1978): A Second Note on the Efficiency of Weighting Subclass Means to Reduce the Effects of Non-response When Analyzing Survey Data. Statistisk tidskrift, 16, pp. 191–196.

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

Received May 1986
Revised September 1988